# Diagnosis and Identification of Exceptional Learners

Learners with disabilities, the gifted and the twice-exceptional



**Dr. Omar A. Alsamani**

# Diagnosis and Identification of Exceptional Learners

Learners with disabilities, the gifted and the twice-exceptional

**Dr. Omar A. Alsamani**

**Diagnosis and Identification of Exceptional Learners**
Learners with disabilities, the gifted and the twice-exceptional

This book is published under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence (CC BY-NC-ND 4.0). It may be copied, distributed and shared, in full or in part, in any medium or format, for non-commercial educational and research purposes, provided it is attributed to its author and its text is not altered. Licence text: creativecommons.org/licenses/by-nc-nd/4.0/

Website: alsamani.com

© 2026 Omar A. Alsamani

Trademarks and instrument names mentioned in this book belong to their owners and are cited for descriptive and educational purposes only. The book reproduces no items from any standardized instrument; diagrams depicting forms are schematic.

First edition, 1448 AH / 2026
Legal deposit no. 1448/7008
ISBN 978-603-06-7714-6

# About the author

**Dr. Omar A. Alsamani** is an Associate Professor and Saudi researcher at the University of Hail, specializing in gifted education, talent development, creativity and innovation. He holds a PhD in Special Education from the University of Northern Colorado, USA, with a concentration in giftedness and in developing creativity and innovation among the gifted; during his doctoral studies he completed two doctoral minors, in Applied Statistics and Research Methods and in Technology, Innovation, and Pedagogy. He also holds a master's degree in gifted education and talent development from the University of Exeter, UK.

His research and professional interests extend to twice-exceptionality; diagnosis, identification and assessment; artificial intelligence in education; and innovation and entrepreneurship. His work focuses on connecting research and practice, and on developing learning environments that help discover and develop abilities and turn ideas into value and innovation.

# Preface

Many who work with children find themselves facing much the same questions. A child does not turn when his name is called: does he have autism spectrum disorder? A girl reads very slowly despite her apparent intelligence: does she have a learning disability? A boy is years ahead of his peers in mathematics yet struggles with writing: is he gifted, does he have a problem, or are both true at once? And a report crowded with numbers: what do they actually mean?

In this book we set out to answer these questions systematically. We do not stop at defining categories and listing their criteria; we take up the details that are so often overlooked: what an instrument really looks like, how a testing session is conducted, who is qualified to administer it, how its results are read, and when the evidence is sufficient to reach a judgment and when it falls short. We have also chosen to bring together in a single volume what the literature usually keeps apart: disabilities and disorders on the one hand, and giftedness and twice-exceptionality on the other, since the same student may need to be viewed from both sides at once.

This is why the title pairs “diagnosis” with “identification.” In gifted education, identification is an educational process whose purpose is to determine who needs gifted services; it is not a diagnosis, because giftedness is not a disorder. It begins with the program's definition, draws on multiple criteria and local norms, and ends in a decision to provide services. In the field of disabilities, identification is tied to determining eligibility for services, and it runs alongside a clinical diagnosis grounded in DSM or ICD criteria.

We read the international standards (DSM-5-TR and ICD-11) and the detailed American experience (IDEA) in light of published Saudi regulations. Each category chapter therefore includes a table comparing the three frameworks, and where we found no published provision we say so rather than assume one. Every instrument profile describes the status of the instrument's Arabic version, and each category chapter closes with a summary of the Arabic tools and studies that have been published.

**The principle behind the book**

A diagnostic judgment does not rest on a single instrument but on the convergence of evidence from multiple sources: observation, interviews, standardized measures, developmental history and formal diagnostic criteria. Every score is a range, not a point; ruling out comes before ruling in; and every judgment is provisional and open to revision. These principles recur throughout the book, and the cases test them from different angles, including cases in which the student is not diagnosed.

## How to use this book

We have taken care to make each chapter self-contained: its cases and examples are complete within it, and the core concepts it relies on are briefly explained, with a parenthetical cross-reference to the chapter that treats them in depth for readers who want more. Readers can therefore begin with whichever chapter concerns them. Reading in sequence, however, builds a deeper understanding: Part One lays down a shared language, Part Two presents the tools for gathering evidence, Part Three applies them to the categories, and Part Four addresses the fairness of judgment and what follows from it.

**The teacher's path**

**The assessment team's path**

Chapter 1, then Chapter 4 (observation and curriculum-based measurement), then the chapter on the category that concerns you, then Chapter 15 (reading the report and the plan).

Chapters 2 and 3 for the measurement foundation, then Chapter 6 (identification models), then the category chapters, then Chapter 14 (fairness).

**The giftedness path**

Chapter 2 (floor, ceiling and norms), then Chapter 6 (combination rules), then Chapters 12 and 13, then the section on gifted decisions in Chapter 15.

**The path for families and decision makers**

Chapter 1, then Chapter 15, then the summary of the relevant category chapter and its decision pathway, then Appendix B for terminology.

## How is a chapter structured?

All chapters share a common structure, and the category chapters (7–13) add elements of their own:

**1 Scene and objectives**

A realistic situation opens the chapter, followed by measurable behavioral objectives, each matched by a question or activity.

**2 Concept and criteria**

The definition in DSM-5-TR, ICD-11, IDEA and Saudi regulations, with a box comparing it to the international classification.

**3 Decision pathway**

A flowchart from the first indicators to the final judgment, beginning with hearing and vision screening before any diagnostic tool.

**4 Instrument profiles**

For each tool: its rationale, its format, a scene of its administration and a data

**5 When a judgment can be made**

Cases with numbers across different ages and stages, some in which the

**6 Conclusion and assessment**

A summary, a glossary, questions tied to the objectives, a new unsolved

table (ages, qualifications, scores, Arabic version), followed by its interpretation and limitations.

diagnosis is confirmed and some in which it is not, followed by differential diagnosis and co-occurrence.

"Case for application," activities with ethical safeguards, and references.

## Guide to the boxes

**For the teacher**
The teacher's role and classroom tools.

**Example or local context**
An example that makes an idea concrete, or what Saudi and Arab regulations and studies have to say.

**Key idea**
A rule or conclusion worth remembering.

**Caution**
A common error or professional pitfall.

**Misconception**
A widespread belief that research corrects.

**For the specialist**
A technical detail or scholarly debate that may be skipped on a first reading.

**About the instruments, their versions and the cases**

In this book we describe instruments and explain how they are administered and interpreted, but we do not reproduce their items or copy their record forms, out of respect for publishers' rights and for the validity of the instruments themselves. Illustrations of forms are schematic and show no more than the layout of the page. Each profile reports the status of the instrument's Arabic version as we found it while preparing the book, and readers are advised to verify the licensing and norms of any version before relying on it. The blank forms in Chapter 4 are our own. All cases are hypothetical, written for teaching purposes, and any resemblance to real persons is unintended.

**About regulatory information**

The Saudi regulatory texts cited in this book (the Regulations for Special Education Institutes and Programs, 1422 AH; the Special Education Organizational Guide, 1436–1437 AH; the Law on the Rights of Persons with Disabilities, 1445 AH; and others) are presented as we found them published while preparing the book. Because regulations are subject to revision, the authoritative reference is always the version currently in force.

# Contents


**Part One: Foundations** 1

**1 Concepts of diagnosis and identification** 2

1.1 Why do we assess? The purposes of assessment 5
1.2 Screening, identification, diagnosis and eligibility 6
1.3 Measurement, testing and assessment 12
1.4 Models of disability 13
1.5 Classification systems 15
1.6 Principles of fair assessment 26
1.7 Sources of evidence: the RIOT/ICEL framework 29
1.8 The team and the family 31
1.9 The general pathway and its common errors 33
1.10 Summary, glossary and activities 41

**2 Measurement fundamentals** 51

2.1 Norm-referenced and criterion-referenced measures 53
2.2 The mean and the standard deviation 55
2.3 The normal curve and standard scores 56
2.4 Percentile ranks and their pitfalls 60
2.5 Reliability/precision and the standard error of measurement 62
2.6 Validity 74
2.7 Norms, standardization and test adaptation 80
2.8 Reading a realistic score table 85
2.9 Ethics and professional responsibility 88
2.10 Summary, key terms and activities 95

**Part Two: Methods** 105

**3 Formal standardized methods** 106

3.1 What makes an instrument standardized? 108
3.2 Intelligence tests 110
3.3 Achievement tests 120
3.4 Adaptive behavior scales 122
3.5 Behavior rating scales 127
3.6 Accommodations during testing 131

3.7 Digital assessment and teleassessment 132
3.8 How is an instrument chosen? 137
3.9 Summary, key terms and activities 143

**4 Informal methods 152**
4.1 Why informal methods? 154
4.2 Structured observation 155
4.3 The interview 161
4.4 Curriculum-based measurement 163
4.5 Error analysis 170
4.6 Portfolios and rubrics 174
4.7 Limits of informal methods 177
4.8 Forms for use 179
4.9 Summary, key terms, and activities 187

**5 Alternative methods 197**
5.1 Why do we need alternative methods? 199
5.2 Dynamic assessment 201
5.3 Functional behavioral assessment 207
5.4 Authentic and performance assessment 215
5.5 Play-based assessment and assessment in early childhood 217
5.6 Assessing learners with severe and multiple disabilities 222
5.7 Summary, key terms, and activities 225

**6 Identification models 235**
6.1 Why do we need a model? 237
6.2 The discrepancy model 237
6.3 Response to intervention 241
6.4 Patterns of strengths and weaknesses 246
6.5 Hybrid models 248
6.6 Multiple-criteria models in giftedness 253
6.7 Comparing the models and reading the Saudi regulations 256
6.8 Summary, key terms, and activities 260

**Part Three: Categories 268**

**7 Intellectual disability 269**
7.1 What is intellectual disability? 271
7.2 Diagnostic criteria 273
7.3 Severity levels and support needs 277
7.4 Early warning signs 279
7.5 The decision pathway and genetic evaluation 281

7.6 Instrument profiles 285
7.7 When is the judgment made? Four cases 299
7.8 Differential diagnosis 309
7.9 Co-occurrence 311
7.10 In the Saudi context: Eligibility criteria 314
7.11 Summary, key terms, and activities 320

**8 Learning disabilities 329**
8.1 What are learning disabilities? 331
8.2 Diagnostic criteria 335
8.3 Dyslexia and reading in Arabic 338
8.4 Difficulties in writing and mathematics 342
8.5 Early warning signs 344
8.6 The decision pathway 345
8.7 Instrument profiles 347
8.8 When is the judgment made? Four cases 360
8.9 Differential diagnosis and co-occurrence 370
8.10 In the Saudi and Arab context 373
8.11 Summary, glossary, and activities 380

**9 Autism spectrum disorder 390**
9.1 What is autism spectrum disorder? 392
9.2 Diagnostic criteria in DSM-5-TR 395
9.3 Support levels, specifiers, and comparison with ICD-11 401
9.4 Early red flags 405
9.5 The decision pathway: From observation to judgment 406
9.6 Instrument profiles 413
9.7 How do you read a diagnostic report? 434
9.8 When is the judgment made? Three cases 436
9.9 Differential diagnosis 444
9.10 Co-occurrence and medical evaluation 446
9.11 In the Saudi and Arab context 449
9.12 Summary, key terms and activities 455

**10 Attention-deficit/hyperactivity disorder and behavioral and emotional disorders 466**
10.1 Attention-deficit/hyperactivity disorder 469
10.2 Diagnostic criteria and specifiers 470
10.3 Who is misdiagnosed and who is overlooked? 476
10.4 Behavioral and emotional disorders: clinical pictures 478
10.5 The educational category: emotional disturbance 483
10.6 Signs that warrant attention 486

10.7 The decision pathway 488
10.8 Instrument profiles 491
10.9 When is a judgment made? Four cases 508
10.10 Differential diagnosis and co-occurrence 517
10.11 In the Saudi and Arab context 521
10.12 Summary, key terms and activities 528

**11 Sensory and multiple disabilities and communication disorders 539**
11.1 Why do these categories matter for every category? 542
11.2 Hearing impairment 543
11.3 Visual impairment 551
11.4 Communication disorders 555
11.5 Developmental language disorder 558
11.6 Difference versus disorder 562
11.7 Multiple disabilities, deafblindness, physical and health impairments, and brain injury 563
11.8 Warning signs and decision pathways 572
11.9 Instrument profiles 579
11.10 When is a judgment made? Four cases 589
11.11 Differential diagnosis and co-occurrence 598
11.12 In the Saudi and Arab context 601
11.13 Summary, key terms and activities 609

**12 Giftedness 620**
12.1 What is giftedness? 624
12.2 Principles of identifying gifted learners 633
12.3 Nomination and the opportunity to display 636
12.4 Ability measures and their ceilings 640
12.5 Creativity, achievement, and performance 643
12.6 Instrument profiles 645
12.7 Identification in the Saudi context and the decision pathway 656
12.8 When is the judgment made? Three cases 666
12.9 What can be mistaken for giftedness, and what can hide it 672
12.10 Summary, key terms and activities 674

**13 Twice-exceptional learners 686**
13.1 Who are twice-exceptional learners? 689
13.2 Masking and the three groups 692
13.3 Why do the usual models fail them? 695
13.4 The debate over identification criteria 696
13.5 Assessment: the protocol and the uneven profile 697
13.6 The decision pathway: opening both tracks together 706
13.7 Differential diagnosis: three possibilities 709

13.8 When is a judgment warranted? Four cases 712
13.9 In the Saudi and Arab context 728
13.10 Summary, key terms and activities 731

**Part Four: Applied issues 740**

**14 Fairness in diagnosis 741**
14.1 What is fairness in diagnosis? 743
14.2 Disproportionality and how it is measured 744
14.3 Where does bias enter? 747
14.4 Fairness in gifted identification 750
14.5 Language and culture 756
14.6 Assessor biases 760
14.7 A fairness checklist 762
14.8 Three cases and a school audit 764
14.9 Summary, key terms and activities 771

**15 From diagnosis to decision 782**
15.1 The eligibility decision: what is actually being decided? 784
15.2 Writing the report 786
15.3 Meeting with the family 796
15.4 From the report to the individualized education program 798
15.5 Progress monitoring, reevaluation and exit from services 803
15.6 Decisions for gifted learners: from identification to placement 807
15.7 The regulatory framework in Saudi Arabia compared with US law 812
15.8 Ethics and the student's voice 819
15.9 Three cases 821
15.10 Principles to take with you from this book 829
15.11 Summary, key terms, and activities 831

**Appendices and lists 842**

**A Guide to instruments 843**

**B Glossary 860**

**List of figures 892**

**List of tables 895**

**Index 900**

# List of abbreviations

| Abbreviation | Full name | Note |
|---|---|---|
| **2e** | Twice-exceptional | — |
| **AAC** | Augmentative and alternative communication | — |
| **AAIDD** | American Association on Intellectual and Developmental Disabilities | — |
| **AAMD** | American Association on Mental Deficiency | — |
| **ABAS-3** | Adaptive Behavior Assessment System, Third Edition | — |
| **ABC** | Antecedent-behavior-consequence | — |
| **ABR** | Auditory brainstem response | — |
| **ADHD** | Attention-Deficit/Hyperactivity Disorder | — |
| **ADI-R** | Autism Diagnostic Interview–Revised | — |
| **ADOS-2** | Autism Diagnostic Observation Schedule, Second Edition | — |
| **AERA** | American Educational Research Association | — |
| **ALP** | Advanced Learning Plan | — |
| **APA** | American Psychiatric Association; American Psychological Association | Psychiatric association: publisher of **DSM-5-TR**; psychological association: author of the ethical principles |
| **ASEBA** | Achenbach System of Empirically Based Assessment | — |
| **ASHA** | American Speech-Language-Hearing Association | — |
| **ASQ-3** | Ages & Stages Questionnaires, Third Edition | — |
| **BASC-3** | Behavior Assessment System for Children, Third Edition | — |
| **BESS** | Behavioral and Emotional Screening System | — |
| **BICS** | Basic Interpersonal Communicative Skills | — |
| **CALP** | Cognitive Academic Language Proficiency | — |
| **CARS-2** | Childhood Autism Rating Scale, Second Edition | — |
| **CARS-2-HF** | Childhood Autism Rating Scale, Second Edition – High-Functioning Version | High-functioning form of **CARS-2** |
| **CARS-2-ST** | Childhood Autism Rating Scale, Second Edition – Standard Version | Standard form of **CARS-2** |
| **CAS2** | Cognitive Assessment System, Second Edition | — |
| **CAT** | Computerized adaptive testing | — |

| Abbreviation | Full name | Note |
|---|---|---|
| **CATALISE** | Criteria and Terminology Applied to Language Impairments: Synthesising the Evidence | International consensus study on the terminology of developmental language disorder |
| **CBCL** | Child Behavior Checklist | — |
| **CBM** | Curriculum-Based Measurement | — |
| **CDI-2** | Children's Depression Inventory, Second Edition | — |
| **C-DM** | Concordance-Discordance Model | — |
| **CELF-5** | Clinical Evaluation of Language Fundamentals, Fifth Edition | — |
| **C.F.R.** | Code of Federal Regulations | US federal implementing regulations (34 **C.F.R.**) |
| **CHC** | Cattell–Horn–Carroll | Theory of cognitive abilities |
| **C-LIM** | Culture-Language Interpretive Matrix | — |
| **CogAT** | Cognitive Abilities Test | — |
| **CPI** | Cognitive Proficiency Index | — |
| **CPT** | Continuous Performance Tests | — |
| **CTOPP-2** | Comprehensive Test of Phonological Processing, Second Edition | — |
| **CVI** | Cerebral/Cortical visual impairment | — |
| **DABS** | Diagnostic Adaptive Behavior Scale | — |
| **DAS-II** | Differential Ability Scales, Second Edition | — |
| **DCD** | Developmental coordination disorder | — |
| **D/CM** | Discrepancy/Consistency Model | — |
| **DD/C** | Dual Discrepancy/Consistency | — |
| **DIF** | Differential item functioning | — |
| **DISCOVER** | — | Project DISCOVER |
| **DLD** | Developmental language disorder | — |
| **DMDD** | Disruptive mood dysregulation disorder | — |
| **DMGT** | Differentiated Model of Giftedness and Talent | — |
| **DSM-5-TR** | Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision | — |
| **DSM-IV** | Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition | — |
| **DT→PI** | Diagnostic Testing → Prescriptive Instruction | — |
| **FBA** | Functional Behavioral Assessment | — |
| **FERPA** | Family Educational Rights and Privacy Act | — |
| **FSIQ** | Full Scale IQ | — |
| **FVA** | Functional Vision Assessment | — |
| **GAI** | General Ability Index | — |
| **GARS-3** | Gilliam Autism Rating Scale, Third Edition | — |
| **GRS** | Gifted Rating Scales | — |
| **HiTOP** | Hierarchical Taxonomy of Psychopathology | — |
| **HOTV** | HOTV chart (the letters H, O, T, V) | — |

| Abbreviation | Full name | Note |
|---|---|---|
| **ICD-11** | International Classification of Diseases, 11th Revision | — |
| **ICEL** | Instruction, Curriculum, Environment, Learner | Domains of the **RIOT/ICEL** framework |
| **ICF** | International Classification of Functioning, Disability and Health | — |
| **IDEA** | Individuals with Disabilities Education Act | US federal law |
| **IDEIA** | Individuals with Disabilities Education Improvement Act | — |
| **IEP** | Individualized Education Program | — |
| **IISCA** | Interview-informed synthesized contingency analysis | — |
| **IRT** | Item Response Theory | — |
| **JCIH** | Joint Committee on Infant Hearing | — |
| **KABC-II NU** | Kaufman Assessment Battery for Children, Second Edition, Normative Update | — |
| **KTEA-3** | Kaufman Test of Educational Achievement, Third Edition | — |
| **LMA** | Learning Media Assessment | — |
| **LPAD** | Learning Potential Assessment Device | — |
| **M-CHAT** | Modified Checklist for Autism in Toddlers | — |
| **M-CHAT-R/F** | Modified Checklist for Autism in Toddlers, Revised with Follow-Up | — |
| **MHS** | Multi-Health Systems | Publisher |
| **MMCAT** | — | Mawhiba Multiple Cognitive Abilities Test |
| **MTSS** | Multi-tiered system of supports | — |
| **NAGC** | National Association for Gifted Children | — |
| **NASP** | National Association of School Psychologists | — |
| **NNAT3** | Naglieri Nonverbal Ability Test, Third Edition | — |
| **OAE** | Otoacoustic emissions | — |
| **PKU** | Phenylketonuria | — |
| **PSW** | Patterns of strengths and weaknesses | — |
| **PTA** | Pure-tone audiometry | — |
| **PTSD** | Posttraumatic stress disorder | — |
| **QPC** | Questionnaire for Parents or Caregivers | Parent questionnaire (part of **CARS-2**) |
| **RAN** | Rapid automatized naming | — |
| **RCADS** | Revised Child Anxiety and Depression Scale | — |
| **RIOT** | Review, Interview, Observe, Test | Methods of the **RIOT/ICEL** framework |
| **RTI** | Response to intervention | — |
| **SB5** | Stanford-Binet Intelligence Scales, Fifth Edition | — |
| **SCARED** | Screen for Child Anxiety Related Emotional Disorders | — |
| **SCQ** | Social Communication Questionnaire | — |

| Abbreviation | Full name | Note |
|---|---|---|
| **SD** | Standard deviation | — |
| **SDQ** | Strengths and Difficulties Questionnaire | — |
| **$SE_{diff}$** | Standard error of the difference | — |
| **SEE** | Standard error of estimation | — |
| **SEM** | Standard Error of Measurement | — |
| **SIS-C** | Supports Intensity Scale–Children's Version | — |
| **SMPY** | Study of Mathematically Precocious Youth | — |
| **SRBCSS** | Scales for Rating the Behavioral Characteristics of Superior Students | — |
| **SRS-2** | Social Responsiveness Scale, Second Edition | — |
| **SRSS** | Student Risk Screening Scale | — |
| **SRSS-IE** | Student Risk Screening Scale–Internalizing and Externalizing | — |
| **SSBD** | Systematic Screening for Behavior Disorders | — |
| **SSI-4** | Stuttering Severity Instrument, Fourth Edition | — |
| **TABs** | Traits, Aptitudes, and Behaviors | — |
| **TEMA-3** | Test of Early Mathematics Ability, Third Edition | — |
| **TONI-4** | Test of Nonverbal Intelligence, Fourth Edition | — |
| **TOPS** | Teacher's Observation of Potential in Students | — |
| **TOWL-4** | Test of Written Language, Fourth Edition | — |
| **TOWRE-2** | Test of Word Reading Efficiency, Second Edition | — |
| **TRF** | Teacher's Report Form | — |
| **TTCT** | Torrance Tests of Creative Thinking | — |
| **UNIT2** | Universal Nonverbal Intelligence Test, Second Edition | — |
| **VADRS-A** | Arabic Vanderbilt Children's ADHD Diagnostic Rating Scale | — |
| **WHO** | World Health Organization | — |
| **WIAT-4** | Wechsler Individual Achievement Test, Fourth Edition | — |
| **WISC-R** | Wechsler Intelligence Scale for Children–Revised | — |
| **WISC-V** | Wechsler Intelligence Scale for Children, Fifth Edition | — |
| **WJ IV ACH** | Woodcock-Johnson IV Tests of Achievement | — |
| **WJ V** | Woodcock-Johnson V | — |
| **WPPSI-III** | Wechsler Preschool and Primary Scale of Intelligence, Third Edition | — |
| **WPS** | Western Psychological Services | Publisher |
| **YSR** | Youth Self-Report | — |

Part One

# Foundations

Concepts and a shared language

Chapter 1: Concepts of diagnosis and identification

Chapter 2: Measurement fundamentals

Chapter 1

# 1 Concepts of diagnosis and identification

Screening, identification, diagnosis and eligibility are four terms often used as if they were synonyms, yet they are in fact four distinct steps, each with its own question, its own tools and its own responsible party. In this chapter we sketch the conceptual map on which all the chapters of the book are organized, and we place it within its regulatory framework in Saudi Arabia.

---

**Scene from the teachers' lounge**

At the start of the school year, Munira, a teacher, received the files of the students in her third-grade class. In Abdullah's file was a hospital report stating that he had been "diagnosed with autism spectrum disorder." Sultan's file held the result of a reading screening the school had conducted the previous month, marked "needs follow-up." Reem's file was empty, but last year's teacher had left Munira a note: "Keep an eye on Reem. I think she's gifted."

> Munira wonders: Is Abdullah entitled to special education services because he has a diagnosis? Does Sultan have a learning disability because the screening flagged him? And is one teacher's opinion enough to consider Reem gifted?

The answer to all three questions is the same: "Not necessarily." Why this is so becomes clear only once we draw careful distinctions between steps that are frequently confused in practice. That is the work of this chapter, and we will return to the three students at its close.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Determine the purpose of assessment in a given situation and the type of tool suited to it.
2. Distinguish among screening, identification, diagnosis and eligibility determination by the question each step asks and the kind of evidence it requires.
3. Compare the three models of disability in terms of the diagnostic question toward which each one points.
4. Map the special education categories in Saudi regulations onto the categories of US law and onto medical classification systems.
5. Identify which principle of fair assessment has been violated in a given assessment situation.
6. Build a **RIOT/ICEL** matrix for gathering evidence about a referred student, indicating the source of each piece of information in it.
7. Plan a conversation with a family about a teacher's concern, following a structured sequence of steps.

8 Detect common errors in a given assessment situation and propose ways to avoid them.

9 Judge whether the available evidence is sufficient to move from one step of the assessment pathway to the next, and justify that judgment.

# Introduction

Every school year, teachers and assessment and diagnosis teams make decisions that bear directly on the paths of particular students: who is referred for assessment, who meets the criteria for a given category, and who needs special education services or a gifted program. Many of the errors in these decisions stem less from the limitations of the tools than from confusing distinct steps that go by similar names. A screening result is treated as a diagnosis, a medical diagnosis is taken as an educational eligibility decision, and a single teacher's opinion is deemed sufficient to judge a student gifted, just as Munira finds in the opening scene. For this reason we open the book by pinning down concepts before turning to tools and categories, since it is the question each step asks that determines the kind of evidence required, who is responsible for it, and how much of it is enough to move on to the next step.

We begin the chapter with the purposes of assessment, starting from the premise that assessment is a means of making decisions (Section 1.1). We then distinguish among screening, identification, diagnosis and eligibility determination (Section 1.2), and among measurement, testing and assessment (Section 1.3). Next we present the models of disability and the question toward which each points (Section 1.4), followed by medical and educational classification systems and the regulatory framework in Saudi Arabia (Section 1.5). We then turn to the principles of fair assessment (Section 1.6), the **RIOT/ICEL** framework for organizing sources of evidence (Section 1.7), and the roles of the team and the family (Section 1.8), and we close with the general assessment pathway and its common errors (Section 1.9).

## 1.1 Why do we assess? The purposes of assessment

Assessment is a means of making a decision. The first question an assessor asks, therefore, is "What decision do I need to make?" Only after that comes the question "Which test should I give?" A tool that is excellent for one decision may be useless for another. Specialists in educational measurement identify several purposes of assessment, each with its own question (Salvia et al., 2017):

**Table 1.1: Purposes of assessment, the question each purpose answers, and an example of a tool for it.**

| Purpose | Question it answers | Example of a tool |
|---|---|---|
| **Screening** | Who needs a closer look? | A short screening checklist given to all kindergarten children |
| **Diagnosis and classification** | Does the student have a disability, a disorder or giftedness? Of what kind? | A comprehensive battery: intelligence, achievement, adaptive behavior and observation |
| **Eligibility determination and placement** | Is the student entitled to special services? In what setting should they be provided? | A team decision based on the diagnosis and its impact on learning |
| **Instructional planning** | What should we teach? Where should we start? | Criterion-referenced tests (which compare a student's performance with a defined level of mastery rather than with the performance of peers), and error analysis |
| **Progress monitoring** | Is the student learning? Is the plan working? | Brief, repeated measures such as curriculum-based measurement (**Curriculum-Based Measurement, CBM**) |

| Purpose | Question it answers | Example of a tool |
|---|---|---|
| **Program evaluation** | Is the program meeting its goals for its students as a whole? | Aggregated data on the program's students across years |

In this book we focus on the first two purposes, with some attention to the third. The teacher, however, is present in all of them, because diagnostic information must ultimately feed into instructional planning; otherwise it loses its value.

## 1.2 Screening, identification, diagnosis and eligibility

Picture these four steps as a funnel: at the top are all students, and at the bottom are those who receive special services. The funnel narrows at each layer, because each step asks a more precise question than the one before it and requires stronger evidence.

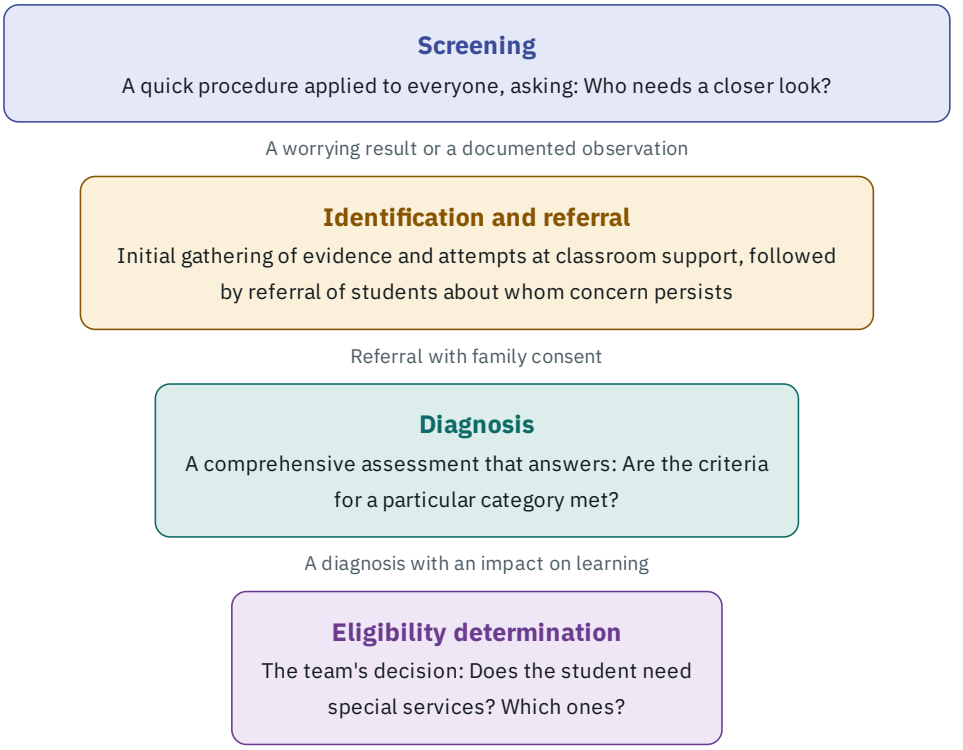



**Figure 1.1:** From screening to eligibility. The number of students shrinks at each layer, while the depth of the questions and the strength of the evidence required increase.

## Screening

Screening is applied to an entire group, such as all first graders, using a short tool that does not require a specialist. Its aim is to ensure that no child who needs a closer look slips past us. Screening tools are therefore preferably high in **sensitivity** (**Sensitivity**), that is, able to capture the largest possible proportion of those who actually have the problem, even if they also flag some who do not, because the cost of missing a child is greater than the cost of examining a child who turns out not to need it.

Sensitivity alone, however, is not enough to judge a screening tool. The tool also needs acceptable **specificity** (**Specificity**), meaning the ability to rule out those who do not have the problem; otherwise it will swamp the team with

unnecessary referrals. A school principal will care about a third question as well: if a student's result is negative, how likely is it that the student truly does not have the problem? This is the **negative predictive value** (**Negative predictive value**), the most important question to ask of a screening, because it measures how confident we can be that we have not missed anyone. Its magnitude, like that of the positive predictive value, depends jointly on the properties of the tool and on the prevalence of the problem in the group screened (Chapter 2 works through these concepts with numbers in Section 2.6).

**Misconception: "The screening flagged him, so the problem is confirmed"**

A positive screening result is an invitation to look more closely; diagnosis is a later stage that requires broader evidence. Screening tools are deliberately designed to catch more than they should, and when a problem is rare in a group, many of those flagged by the screening may in fact be free of it. The reverse holds as well: a negative result is no certificate of health. If the teacher's or the family's concern persists, a documented observation is reason enough to move on to the identification stage.

## Identification and referral

After screening, or after a teacher's observation, the identification stage (**Identification**) begins: the stage of gathering initial evidence. At this stage the teacher and the school try to provide support within the classroom, whether by adjusting the method of instruction or by offering additional small-group teaching, while monitoring the student's response. If the difficulty persists despite this support, the student is referred for a comprehensive assessment. Referral is a significant decision, because it signals that the school has tried what it can and that this has not been enough.

In the field of giftedness, this stage is usually called **nomination**. Teachers and families nominate students in whom they see indicators of giftedness, and those students are then assessed. Relying on nomination alone, however, misses many gifted students, because nominators tend to notice those who display their excellence clearly in achievement and behavior, and overlook those whose giftedness is hidden behind a second language, a disability or a quiet disposition (Chapter 12 addresses nomination in detail).

**The concept of identification: Why is it different in the field of giftedness?**

In this book we restrict "identification" to the stage that precedes referral, a terminological choice that gives each step of the pathway a name of its own. In the literature, however, the term **Identification** usually denotes the entire process of determination, up to and including the decision, and in the Arabic literature it overlaps with "**al-kashf**" (detection), as in the name of the "National Program for Gifted Identification," which encompasses both testing and classification. Yet this broad sense does not carry the same content in the two fields. In the field of giftedness, identification is the complete process of determining who needs gifted services, and it is an educational process; the term "diagnosis" is avoided here because giftedness is neither a disorder nor a deficit whose cause must be sought. The process starts from the program's definition and goals, draws on multiple criteria and on local norms that compare students with peers in their own setting, and links the tool to the type of service offered: a measure of mathematical reasoning suits an enrichment program in mathematics but is not, on its own, appropriate for a program in creative writing. Its outcome is a decision to place a student in a program or to provide a service, one that is tied to the student's educational need and open to revision whenever the evidence or the program changes (National Association for Gifted Children [NAGC], 2019).

In the field of disabilities, by contrast, "identification" usually means determining eligibility for special education services under the categories recognized by the education system, as in the US Individuals with Disabilities Education Act and its implementing regulations published in the Code of Federal Regulations (IDEA, 2004; 34 C.F.R. §300.8). For some categories, such as autism spectrum disorder and attention-deficit/hyperactivity disorder, this educational decision is paralleled or preceded by a clinical diagnosis under the Diagnostic and Statistical

Manual of Mental Disorders (American Psychiatric Association, 2022) or the International Classification of Diseases (ICD-11). The educational decision thus meets the medical one without either replacing the other, as Section 1.5 explains in detail.

The two pathways converge in twice-exceptional learners: the same student may need a clinical diagnosis of a disability and an educational identification of giftedness, and each pathway may obscure what the other reveals if the team pursues one without the other (Chapter 13 examines this overlap in detail). So when you encounter either term in another source, infer its meaning from the context and the field, and do not assume that it refers to the narrow stage we intend here.

## Diagnosis

Diagnosis answers the question: Does the student meet the criteria for a particular category? It requires a comprehensive team assessment using multiple tools and multiple sources. A diagnosis may be medical, issued by a physician or a qualified psychologist according to a diagnostic manual, or educational, issued by a school team according to education regulations.

## Eligibility determination

Here we arrive at the distinction that puzzled Munira. Diagnosis answers the question “What does the student have?”, whereas eligibility answers the question “Does the student need special services in order to learn?” A student may have a medical diagnosis yet learn successfully in the general classroom without needing special services, and so not be eligible for them. Conversely, a student may need services without yet having a medical diagnosis, in which case the student is assessed educationally.

**The equation that governs eligibility**

In most education systems, eligibility for special education services requires two conditions together: the student must meet the criteria for one of the disability categories, and the disability must affect the student's learning to the point that he or she needs specially designed instruction. If either condition is absent, the student is not eligible, even if the other is met. Saudi regulations phrase the second condition in their own words: an impairment "to the extent that it necessitates the provision of special education services" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]).

## 1.3 Measurement, testing and assessment

These words overlap in everyday use, but each has a specific meaning in the literature, and knowing the difference helps in reading reports and sources accurately:

**Table 1.2: The difference among testing, measurement, assessment and evaluation.**

| Term | Meaning | Example |
|---|---|---|
| **Test** | An organized set of tasks or questions administered in a specified way | An intelligence test, or a 20-word spelling test |
| **Measurement** | Expressing a characteristic in numbers according to rules | Scored 14 out of 20, or has a standard score of 85 |
| **Assessment** | The complete process of gathering information from multiple sources in order to make a decision | Tests, interviews, observation and records, ending in a decision |

| Term | Meaning | Example |
|---|---|---|
| **Evaluation** | Making a judgment about the worth or adequacy of something; in some systems also used to mean the comprehensive assessment of a student | Evaluating a program's effectiveness, or the comprehensive evaluation of a student's eligibility |

The key point here is that testing is a small part of assessment. Sound assessment combines tests with observations, interviews and a review of records and work samples, as we will see in the evidence-gathering framework in Section 1.7.

## 1.4 Models of disability

Before we diagnose a disability, we must ask: What is disability in the first place? Where does the problem lie? How we answer these two questions is a stance with consequences, for it determines what we measure and where we look for solutions. There are three main models:

### The medical model

**The problem lies in the individual**

Disability is a defect in the individual's body or functions, to be diagnosed and treated or compensated for. The focus is on the impairment and its causes.

In diagnosis: looking for symptoms, matching

### The social model

**The problem lies in the environment**

Disability arises from the barriers society places in front of those who differ: buildings a wheelchair cannot reach, curricula that ignore individual differences. The focus is on removing barriers (Oliver, 1990).

### The biopsychosocial model

**The problem lies in the interaction between the two**

Disability results from an interaction between the individual's health condition and the environmental and personal factors

| | | |
|---|---|---|
| them against criteria, and determining severity. | In diagnosis: asking about the barriers in the student's environment and what needs to change around the student. | surrounding him or her (Engel, 1977). This model is the foundation of the International Classification of Functioning, Disability and Health issued by the World Health Organization (World Health Organization [WHO], 2001).<br><br>In diagnosis: describing the impairment, what the student is able to do, and the environmental factors that hinder or help. |

It is often said that modern diagnosis follows the third model, but this claim is a generalization that needs qualifying. The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, text revision (DSM-5-TR), remains at its core a categorical classification belonging to the medical model: the criteria are either met or not met, even though the manual adds severity specifiers and levels of support that move it closer to graded description (American Psychiatric Association, 2022). The biopsychosocial model, by contrast, underlies the International Classification of Functioning specifically, a descriptive classification that addresses functioning, participation and environmental factors and leaves the diagnosis of conditions to the medical systems. There is also growing discussion in the literature of dimensional models that describe psychological difficulties along graded continua rather than in discrete categories, one of the best known being the Hierarchical Taxonomy of Psychopathology (**HiTOP**).

Good practice comes closest to combining the two logics: the assessor judges whether criteria are met using a categorical logic when the system requires it, and then describes the impact of the condition and its relationship to the environment using a biopsychosocial logic. When we say that a student with autism spectrum disorder "requires substantial support," we are describing the student's relationship to the environment as well as what the student has. The same student may need less support in a calm, structured classroom and more in a crowded, changeable one.

## 1.5 Classification systems

Diagnosis rests on classification systems that define categories and their criteria. The most important thing for a teacher to know is that there are two kinds of systems, medical and educational, and that each serves a different purpose.

**Table 1.3: Medical and educational classification systems and their purposes.**

| System | Issuing body | Purpose | Who uses it |
|---|---|---|---|
| **Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)** | American Psychiatric Association, 2022 | Diagnosing mental and developmental disorders using detailed criteria | Psychiatrists and psychologists |

| System | Issuing body | Purpose | Who uses it |
|---|---|---|---|
| **International Classification of Diseases, Eleventh Revision (ICD-11)** | World Health Organization; adopted in 2019 and in force since January 1, 2022; its clinical descriptions and diagnostic requirements for mental and developmental disorders were published in 2024 | Classifying all diseases and health conditions, including developmental ones, for diagnostic and statistical purposes | Health systems around the world |
| **International Classification of Functioning, Disability and Health (ICF)** | World Health Organization, 2001 | Describing what a person can do and participate in, and the environmental factors at play, rather than diagnosing the condition | Rehabilitation and service-planning teams |
| **Educational categories** | Each country's education legislation and regulations | Determining who is entitled to special education services | School teams and education departments |

Clinical diagnosis relies on one of the first two manuals (American Psychiatric Association, 2022; WHO, 2019, 2024). The two converge for most developmental categories, though they differ in naming and detail, differences taken up in the category chapters of Part Three.

## An example of an educational system: the US categories

The US Individuals with Disabilities Education Act (**Individuals with Disabilities Education Act, IDEA**) was reauthorized in 2004 as the Individuals with Disabilities Education Improvement Act (**IDEIA**), though it is still known by its original acronym. The law specifies thirteen categories eligible for

services: autism, deaf-blindness, deafness, emotional disturbance, hearing impairment, intellectual disability, multiple disabilities, orthopedic impairment, other health impairment, specific learning disability, speech or language impairment, traumatic brain injury (resulting from a blow or impact), and visual impairment including blindness. The law also permits a developmental delay category for children aged 3 to 9 (IDEA, 2004).

Note three important points about this list:

1. **Attention-deficit/hyperactivity disorder (Attention-Deficit/Hyperactivity Disorder, ADHD) is not a separate category.** A student diagnosed with it usually qualifies for services under the category of "other health impairment." In other words, medical diagnoses and educational categories do not always coincide. Saudi regulations depart from this arrangement, as we will see in the next section.
2. **Giftedness is absent.** The law concerns disabilities, and gifted services in the United States are governed by separate legislation. This explains why a twice-exceptional student, that is, a gifted student who also has a disability or disorder, falls between two systems that are not coordinated with each other (Chapter 13 discusses this group in detail).
3. **The labels differ from those of the medical manual.** "Emotional disturbance" is a broad educational category that encompasses conditions bearing different names in the medical manual.

### 1.5.1 The regulatory framework in Saudi Arabia

In this section we present the Saudi education system in the same detail as the US example, since it is the framework within which the assessment procedures discussed in this book are carried out. Special education in Saudi Arabia is organized at two levels: general laws that establish the rights of persons with disabilities, and regulatory documents issued by the Ministry of Education

that define the categories, eligibility criteria and assessment and diagnosis procedures. The identification of gifted students is organized along a pathway separate from both. Table 1.4 summarizes the principal documents.

**Table 1.4: The principal regulatory documents for diagnosis and identification in Saudi Arabia.**

| Document or body | Issuance | What matters to the assessor |
|---|---|---|
| **Regulations for Special Education Institutes and Programs** | Ministry of Education [Al-Ma'arif] (the former name of the Ministry of Education), Decision No. 1674 of 5/4/1422 AH (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]) | Definitions of the categories and admission requirements for each; a section on assessment and diagnosis (Articles 78–82) setting out its rules, its team and its seven steps; the components of the individualized education program (Articles 83–85); and the obligations of the institute or program toward the family (Article 76) |
| **Special Education Organizational Guide and Special Education Procedural Guide** | Ministry of Education, first edition for the 1436–1437 AH school year, which counts the Regulations among its foundations (Ministry of Education, 1436 AH-a, 1436 AH-b [Arabic]) | The recognized categories and the eligibility criteria for each, the composition of the assessment and diagnosis team, parental consent, timelines for reevaluation, and eligibility for early intervention |

| Document or body | Issuance | What matters to the assessor |
|---|---|---|
| **Law on the Rights of Persons with Disabilities** | Royal Decree No. (M/27) of 11/2/1445 AH (2023), in 33 articles; it replaced the Disability Welfare Law issued by Royal Decree No. (M/37) of 23/9/1421 AH | The right of persons with disabilities to "modern assessment and diagnostic tools" and to early intervention programs (Article 8), and to screening and diagnostic services and the "early detection of disability" (Article 9) |
| **Authority for the Care of Persons with Disabilities** | Its organizational statute was approved by the Council of Ministers in 1439 AH (2018) (Al-Jazirah, 2018 [Arabic]) | A financially and administratively independent body whose board is chaired by the Minister of Human Resources and Social Development; it works to safeguard the rights of persons with disabilities, develop policies, programs and services, and support research (Saudipedia, n.d. [Arabic]) |

| Document or body | Issuance | What matters to the assessor |
|---|---|---|
| **National Program for Gifted Identification, and the regulations for gifted programs at the Ministry of Education** | The program was launched in 2011 as a partnership among Mawhiba, the Ministry of Education and the National Center for Assessment (King Abdulaziz and His Companions Foundation for Giftedness and Creativity [Mawhiba], n.d. [Arabic]); the General Administration for the Gifted organizes gifted classes under a dedicated guide (Ministry of Education, n.d. [Arabic]) | The Mawhiba Multiple Mental Abilities Test, the nomination criteria for gifted classes, and the rule for combining them |

Note that the two education documents use the term ʿawq (impairment) where this book uses iʿāqa (disability). The Regulations use "mental retardation," dividing it into "educable" and "trainable" according to ranges of IQ scores, whereas the Organizational Guide uses "mental impairment" and additionally requires deficits in at least two adaptive domains. In both cases the condition meant is what this book calls "intellectual disability."

### Categories and their criteria: a comparison with the US system

The Organizational Guide recognizes eleven categories. Table 1.5 maps them onto the categories of US law, together with the main eligibility criteria the Guide specifies.

**Table 1.5: The categories of US law (IDEA) and their counterparts in the Special Education Organizational Guide, with the main eligibility criteria.**

| US legal category | Counterpart in the Organizational Guide | Main eligibility criteria in Saudi regulations |
|---|---|---|
| **Intellectual disability** | Mental impairment | IQ no higher than 75 on the Wechsler scale or 69 on the Stanford-Binet Fifth Edition, or the equivalent of either, together with deficits in at least two adaptive skill domains. US law sets no numerical cut-off score |
| **Specific learning disabilities** | Learning disabilities | A marked discrepancy between ability and achievement in one academic area, or a disorder in one of the basic psychological processes; with intellectual and sensory impairments and behavioral disorder ruled out; and evidence that general education services are inappropriate or of limited effectiveness. The text specifies no numerical size for the discrepancy and makes no mention of the response to intervention model. US law, by contrast, prohibits states from requiring a discrepancy and obliges them to permit the response to intervention model (34 C.F.R. §300.307) |
| **Deafness; hearing impairment** | Hearing impairment (deaf and hard of hearing) | A hearing loss of 70 decibels or more for deafness, and between 35 and 69 decibels for hard of hearing, in the better ear with hearing aids. The US definition is functional, with no numerical threshold |

| US legal category | Counterpart in the Organizational Guide | Main eligibility criteria in Saudi regulations |
|---|---|---|
| **Visual impairment including blindness** | Visual impairment (blind and low vision) | Low vision: acuity between 6/24 and 6/60 meters. Blindness in the Regulations: acuity below 6/60 in the better eye after correction, or a visual field of less than 20 degrees |
| **Deaf-blindness** | Deaf-blind | A separate category in both systems |
| **Multiple disabilities** | Multiple impairments | A separate category in both systems |
| **Orthopedic impairment; other health impairment** | Physical and health impairment | A single category in the Guide corresponding to two categories in US law |
| **Autism** | Autism spectrum disorder | The Guide defines it by persistent deficits in social communication and social interaction across multiple contexts, in wording that matches the first criterion of the diagnostic manual |
| **Speech or language impairment** | Language and speech disorders | A separate category in both systems |
| **Emotional disturbance** | Behavioral and emotional disorders | A separate category in both systems; the Saudi label is broader in its wording |
| **Not a separate category (usually served under "other health impairment")** | Attention-deficit and hyperactivity disorders | "Hyperactivity, attention deficit or both, requiring intervention according to an accredited report" |

| US legal category | Counterpart in the Organizational Guide | Main eligibility criteria in Saudi regulations |
|---|---|---|
| **Traumatic brain injury** | No separate category in the Guide | We found no separate category for it in the Organizational Guide |
| **Developmental delay (ages 3–9)** | Early intervention: developmentally delayed children and children at risk (up to age 6) | A delay of two standard deviations below the mean in two or more domains, or of 25%; or exposure to at least three environmental risk factors |
| **Giftedness: outside the law** | Giftedness: outside the Special Education Organizational Guide | Organized along a separate pathway (the National Program for Gifted Identification and the programs of the General Administration for the Gifted) |

Sources: Ministry of Education (1436 AH-a [Arabic]); Ministry of Education [Al-Ma'arif] (1422 AH [Arabic]) for the definition of blindness; and US law and its implementing regulations (IDEA, 2004; 34 C.F.R. §300.8). The Guide also sets general eligibility requirements, among them the statutory age for admission to first grade, with special education students permitted an exception of no more than three years.

### Identification of gifted students

The identification of gifted students in Saudi Arabia rests on two complementary pathways. The first is the "National Program for Gifted Identification," which targets students from third grade through the first year of secondary school at three levels. Students may sit the test once at each level, and the program administers the "Mawhiba Multiple Mental Abilities Test" upon registration by the family (Mawhiba, n.d. [Arabic]). The second consists of the Ministry of Education's gifted classes, which draw their students from those classified as gifted by the national program. If places remain unfilled,

the classes complete their numbers using four criteria: a mental abilities test, personal traits, academic achievement and creative products. A student must satisfy two or more criteria, one of which must be the mental abilities test with a score of at least 120 (Ministry of Education, n.d. [Arabic]). For a recent overview of gifted education in Saudi Arabia, see Alfaiz and colleagues (Alfaiz et al., 2022).

**In the Saudi context: What the comparison reveals**

- **Attention-deficit/hyperactivity disorder is a separate category in Saudi Arabia.** This brings the educational category closer to the medical diagnosis than the US system does, but it also makes the "accredited report" a cornerstone of eligibility, which raises the stakes for the quality of the assessment on which that report was based.
- **The learning disabilities model in the text rests on discrepancy, exclusion and nonresponse to general instruction.** The text specifies neither the size of the discrepancy nor how nonresponse is to be established. Recent studies indicate that practices resembling the response to intervention model exist in Saudi schools, but that they are largely informal and lack systematic progress monitoring, and that the absence of formal adoption of the model is among the barriers to its implementation (Almulla & Mahama, 2026; Altheiabi et al., 2025).
- **Gifted students with disabilities fall between two documents.** The Special Education Organizational Guide does not address gifted students, and the gifted guides we reviewed do not address students with disabilities. We found no published regulatory text that either permits or prohibits combining a gifted program with learning disabilities services. A recent exploratory study estimated that about 5.9% of a sample of 6,875 students in four private schools in Riyadh may be twice-exceptional, and called for policies specific to them (Almulhim et al., 2026).

The documents cited in this section were consulted in September 2026, and we found no published edition of the Organizational and Procedural Guides more recent than their first edition; readers should consult whichever edition is in force at the time of reading. **Question for**

**reflection:** If you were invited to join a committee updating the Organizational Guide, which of these three gaps would you address first, and why?

## 1.6 Principles of fair assessment

An assessment that ends in a diagnosis changes the course of a student's life: the school the student attends, the curriculum the student learns, how others see the student, and how the student sees himself or herself. For this reason, legislation and professional standards have set out principles that protect students from error and injustice. Among the clearest formulations are those in the implementing regulations of US law governing evaluation procedures (34 C.F.R. §300.304). The Standards for Educational and Psychological Testing treat fairness as a third foundation of testing alongside validity and reliability, devoting their third chapter to it (American Educational Research Association [AERA] et al., 2014). Table 1.6 sets these principles against the provisions of Saudi regulations.

**Table 1.6: Principles of fair assessment, what they mean in practice, and their counterparts in Saudi regulations.**

| Principle | What it means in practice | Counterpart in Saudi regulations |
|---|---|---|
| **Multiple tools and strategies** | No single measure is used as the sole basis for any decision; a single score, however precise, is not enough on its own to support a judgment. | "Not relying solely on the results of a single measure when determining a pupil's eligibility," and using formal and informal methods (Rules, Article 80; Organizational Guide) |

| Principle | What it means in practice | Counterpart in Saudi regulations |
|---|---|---|
| **Assessing all areas of suspected need** | If there is doubt about hearing, language or attention, each is assessed, rather than only the area for which the student was referred. | A multidisciplinary team and information gathered from varied sources (Rules, Article 80) |
| **Appropriate language and mode** | The student is assessed in his or her native language or mode of communication; a deaf student is assessed in sign language if that is the student's language. | We found no explicit provision in either document specifying the language or mode of assessment |
| **Freedom from bias** | Tools are chosen so that they do not disadvantage the student because of culture or background. | Standardized tests that are “valid, reliable and suited to the child's environment” (Rules, Article 80) |
| **Measuring what is intended to be measured** | If a student has a motor disability, his or her intelligence is not measured with a test that depends on motor speed, lest the score fall because of the student's hand rather than ability. | “Appropriate measures and clear instructions” (Rules, Article 80), with no detail on adapting administration for students with sensory and motor disabilities |

| Principle | What it means in practice | Counterpart in Saudi regulations |
|---|---|---|
| **Validity of the tool and qualifications of the examiner** | The tool is used for the purpose for which its validity has been established, and it is administered by a trained person according to its instructions. | Using "tools and measures approved by the Ministry of Education," administered and interpreted "by qualified specialists" (Organizational Guide). We found no published text specifying the qualification levels required for each class of standardized psychological test |
| **Informed family consent** | The family knows what will happen and why, and consents in writing before the process begins. | Written consent from the parent or guardian after formal notification; if no reply is received within fourteen days, this is deemed consent (Organizational Guide; the Regulations specify two weeks) |
| **Confidentiality** | Only those entitled to access the information may see it, and it is stored in a way that protects it. | "Maintaining the confidentiality of all information" (Rules, Articles 76 and 80); the Personal Data Protection Law (1443 AH) classifies health data as sensitive data |

Sources: implementing regulations of US law (34 C.F.R. §300.304), Standards for Educational and Psychological Testing (AERA et al., 2014), Ministry of Education [Al-Ma'arif] (1422 AH [Arabic]), and Ministry of Education (1436 AH-a [Arabic]).

The comparison reveals the strength of Saudi regulations on the principles of multiple tools and examiner qualifications, and their silence on the language and mode of assessment, a silence with real consequences for the assessment of deaf students and of students whose native language is not Arabic. The consent rule deserves particular attention. In the professional literature, informed consent is an affirmative act that the family performs after it understands, and silence does not count as consent under the US definition (34

C.F.R. §300.9). Treating a lack of response as consent may streamline procedures, but it does not ensure that the family understood what it agreed to. The team would therefore do well, even when the rule allows it to proceed, to seek explicit consent and to arrange a meeting in which it explains what will take place.

These principles do not take the place of the ethical codes of the professions in which assessors work, such as the principles of the US National Association of School Psychologists (National Association of School Psychologists, 2020). Technical standards set out how a test is built and used; ethical codes set out what the assessor owes the student and the family (Chapter 2 discusses the difference between the two in detail).

## 1.7 Sources of evidence: the RIOT/ICEL framework

How can we be sure that the evidence we have gathered is sufficient and varied? School psychology offers a simple, practical framework for this purpose, known by its acronym (**RIOT/ICEL**; Christ, 2008). The acronym is formed from the initial letters of its words: the first part (**RIOT**) covers the methods of gathering information (**Review, Interview, Observe, Test**), and the second part (**ICEL**) covers the domains (**Instruction, Curriculum, Environment, Learner**). The framework is a two-dimensional matrix: **four methods** of gathering information, applied across **four domains** in which the problem may lie.

- **The four methods:** review, interview, observation and testing.
- **The four domains:** instruction, curriculum, environment and the learner.

The central idea of the framework is that the problem may lie outside the student: in the way he is taught, in a curriculum pitched above his level, or in a noisy environment. An assessment that looks at the student alone may

diagnose a disability that does not exist. Table 1.7 presents the matrix with examples in each cell.

**Table 1.7: The RIOT/ICEL matrix, with examples in each cell. A balanced assessment fills cells in every row and every column rather than confining itself to the testing cell in the learner row.**

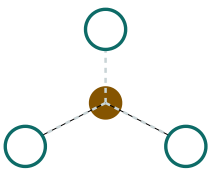
| | | | | |
|---|---|---|---|---|
| **Instruction Instruction** | Lesson plans and the teaching methods used so far | Asking the teacher how the skill was presented | Observing the clarity of explanation and the pace of teaching | Trying a different teaching method and measuring the response to it |
| **Curriculum Curriculum** | Difficulty level of the textbook and the sequence of skills | Asking the teacher about the prerequisites for the skill | Observing whether the task matches the student's level | Measuring text readability, or assessing prerequisite skills |
| **Environment Environment** | Attendance records, and changes of school and teacher | Asking the family about circumstances at home | Observing noise, lighting and where the student sits | Classroom environment scales |
| **Learner Learner** | Medical reports, previous results and samples of the student's work | Asking the student what he finds difficult | Observing his behavior across different situations | Ability and achievement tests and diagnostic scales |

**For the teacher**

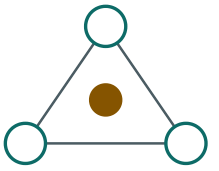
Most of the cells in this matrix are in your hands before they are in the specialist's. You are the one who knows how the student is taught, the level of the curriculum and the classroom environment, and you hold the student's daily work. When you attach this information to the referral, you give the team half the picture before it begins.

## 1.8 The team and the family

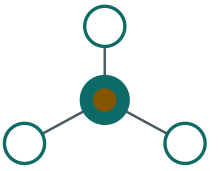
No single specialist has all the knowledge needed to understand a student fully. Diagnosis is therefore the work of a team, which, depending on the case, may include the psychologist, the special education teacher, the classroom teacher, the speech-language pathologist, the physician, the occupational therapist, the audiologist, the social worker and the family. Teams differ in how they work together:



**Multidisciplinary**
Each specialist assesses alone and writes a separate report. Their recommendations may conflict without their ever meeting.



**Interdisciplinary**
Each specialist assesses within his or her own field; they then meet, share their findings and reach a joint decision and plan.



**Transdisciplinary**
The team assesses the child together in a single session, with one member working directly with the child under the guidance of the others. This model is common in early childhood.

**Figure 1.2:** Three models of teamwork. The solid circle at the center represents the child, the hollow circles around it represent the specialists, and the lines represent communication: dashed between each specialist and the child alone, solid between specialists when they work together.

### The assessment team in Saudi regulations

The Regulations assign assessment and diagnosis to a team made up of the relevant special education teacher, the behavioral training teacher (the psychologist), the child's guardian, and anyone whose participation the case requires (Article 81). Among the psychologist's duties they list “conducting

assessment and diagnosis" using formal tools, such as standardized intelligence scales and adaptive behavior scales, and informal tools, such as interviews, observation and checklists (Articles 48 and 49) (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). The Organizational Guide widens the team to include the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the student counselor or social worker, specialists according to the disability category, the physician or nurse where available, and the student's guardian "if possible" (Ministry of Education, 1436 AH-a [Arabic]). The phrase "if possible" is weaker than what the Regulations themselves establish when they make the guardian a member of the team, and weaker than what this book holds in the next section.

## The family as a partner in assessment and decision-making

The family has known the student since birth, and in situations no one else sees. Its role in assessment goes well beyond answering questions:

- **The holder of consent:** Assessment does not begin without the family's informed consent, and the family has the right to ask questions and to object.
- **A source of developmental history:** When the child walked, when he began to talk, and what changed and when.
- **A witness to behavior in a different setting:** What the family sees at home may differ from what the teacher sees, and the difference is itself information.
- **A partner in the decision and the plan:** The family attends the results meeting, understands the findings and takes part in setting goals.

The Regulations grant the family explicit rights in this regard: the guardian's permission must be obtained for diagnosis, for preparing the program and for approving any modification; the family is to be invited to take part in diagnosis, program planning and evaluation; their child's condition is to be

explained to them “in understandable language”; and the student is to be rediagnosed at the family's request if they believe he has not been placed in the appropriate program (Article 76). The Regulations also require the family's participation in preparing, implementing, evaluating and monitoring the individualized education program (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]).

**For the teacher: When you talk to the family about your concern**

1. Begin with what the student does well; the family needs to know that you see their child as a whole person.
2. Describe what you have noticed in terms of specific behaviors and examples, and avoid labels: say “He doesn't turn when I call his name,” not “I think he has autism.”
3. Ask: “Do you notice anything similar at home?” Then listen.
4. Explain the next step and the reason for it: “I suggest that a specialist team assess him, so that we can understand how to help him.”
5. Do not promise an outcome, and do not suggest a diagnosis to the family; that is outside your role, and it may be wrong.

## 1.9 The general pathway and its common errors

Detailed procedures vary across countries and categories, but most systems follow a broadly similar pathway:

1 **Observation and screening**

The teacher or the family notices indicators, or the student is flagged in screening results.

Teacher, family, screening team

2 **Classroom support and documentation**

Instructional adjustments and additional support for a set period, with the response measured and documented.

Teacher, supported by the school team

3 **Referral and consent**

If the response is insufficient, a referral is written with the evidence attached, and the family's consent is requested.

Teacher, school administration, family

4 **Comprehensive assessment**

The team gathers evidence from multiple sources, using tools suited to each area of suspected need.

Multidisciplinary team

5 **Decision meeting**

The team and the family discuss the results and determine whether the criteria are met and whether the student is eligible for services.

Team, family

6 **Plan, monitoring and reevaluation**

The individualized education program (**Individualized Education Program, IEP**) is developed, progress is monitored, and the student is reevaluated periodically or as needed.

Special education teacher, classroom teacher, family

Some systems set binding timelines for these steps. In the US system, for example, the initial evaluation must be completed within sixty days of parental consent, unless the state sets a different timeline (IDEA, 2004). The purpose of such timelines is to ensure that a student does not wait for months without services or a decision. Table 1.8 sets each step against the provisions of the two systems.

**Table 1.8: Steps of the general pathway in the US system and their counterparts in Saudi regulations.**

| Step | In the US system (as an example) | Counterpart in the Saudi system |
|---|---|---|
| **1. Observation and screening** | The state is obliged to locate, identify and evaluate children with disabilities (**Child find**; 34 C.F.R. §300.111) | A right to the "early detection of disability" (Law on the Rights of Persons with Disabilities, Article 9); for giftedness: the National Program for Gifted Identification, upon registration |
| **2. Classroom support and documentation** | For learning disabilities, evidence of appropriate instruction and repeated measurement of progress is required, and the response to intervention model may be used (34 C.F.R. §300.307, §300.309) | For learning disabilities, it must be established that general education services are inappropriate or of limited effectiveness, without specifying how this is to be established; the student may be placed under observation "for a full semester" where needed (Regulations, Article 80) |
| **3. Referral and consent** | Informed written parental consent before the initial evaluation (34 C.F.R. §300.300) | Written consent from the guardian after formal notification; no response within 14 days is treated as consent (Organizational Guide) |
| **4. Comprehensive assessment** | Completed within sixty days of consent, unless the state sets a different timeline | Carried out by the assessment team using tools approved by the Ministry, in seven steps that begin with notifying the guardian and end with the final report (Regulations, Article 82). We found no timeline in either document for completing the initial evaluation that corresponds to the sixty-day limit |

| Step | In the US system (as an example) | Counterpart in the Saudi system |
|---|---|---|
| **5. Decision meeting** | A team that includes the parents determines eligibility, and the family receives a copy of the evaluation report (34 C.F.R. §300.306) | The assessment team issues the final report; the guardian is a member of the team under the Regulations, and "if possible" under the Guide |
| **6. Plan and reevaluation** | An individualized education program with specified components (34 C.F.R. §300.320); reevaluation at least once every three years, and not more than once a year unless the family and the agency agree (34 C.F.R. §300.303) | An individualized education program covering present levels of performance, long-term and short-term goals written in measurable behavioral terms, services, and performance criteria and dates for evaluating the plan (Regulations, Articles 83–85); rediagnosis no sooner than six months and no later than eighteen months after the initial diagnosis, and thereafter as needed (Organizational Guide); the family may request rediagnosis (Regulations, Article 76) |

Sources: Ministry of Education [Al-Ma'arif] (1422 AH [Arabic]), Ministry of Education (1436 AH-a [Arabic]), the Law on the Rights of Persons with Disabilities (1445 AH [Arabic]), and the implementing regulations of US law (34 C.F.R. Part 300).

## Common errors in diagnosis

**Table 1.9: Common errors in diagnosis, how they occur and how to avoid them.**

| Error | How it occurs | How to avoid it |
|---|---|---|
| **Diagnostic overshadowing** | New symptoms are attributed to a disability the student is already known to have, so that another disorder is missed; for example, anxiety in a student with an intellectual disability is explained away as "part of his disability" (Reiss et al., 1982). | Asking every time: Does the known disability really explain this symptom? |
| **Confirmation bias** | The assessor starts from a hypothesis, sees the evidence that supports it and overlooks the evidence against it. A teacher convinced that a very active student has attention-deficit/hyperactivity disorder, for example, records every moment of distraction but fails to notice his sustained attention when a task captures his interest. | Deliberately seeking alternative explanations and attending to discrepancies among sources. |
| **Reliance on a single score** | A decision is made because the score fell above or below the cut-off score (the threshold that the scale or the system sets for the decision), without regard to the confidence interval, the range within which the student's true score is likely to fall, or to other evidence. | The principle of multiple tools, and reading the score within its surrounding confidence interval (Chapter 2 explains both concepts in detail). |
| **Overdiagnosis or underdiagnosis** | A group of students is represented among those diagnosed at a higher or lower rate than expected because of language, sex or social and cultural background. | Fair tools, local norms and periodic review of referral data. |

| Error | How it occurs | How to avoid it |
|---|---|---|
| **Overlooking the effect of labeling** | The team forgets that a diagnosis can change what people expect of the student, so that less is asked of him. | Having the report describe strengths and focus on the support the student needs. |

**Back to Ms. Munira's classroom**

- **Abdullah:** He has a medical diagnosis of autism spectrum disorder, but his eligibility for services is determined by the impact of the disorder on his learning. He may need extensive services, or only simple classroom accommodations, that is, changes in how he learns or is tested that leave the level of the curriculum untouched. The medical report in his file is evidence the assessment team can draw on in reaching the eligibility decision.
- **Sultan:** The screening flagged him, which means he needs a closer look; concluding that he has a learning disability would be premature. The next step is reading support and monitoring of his response, before any referral.
- **Reem:** The teacher's opinion is a useful nomination and the starting point for identification. Reem needs an assessment based on multiple criteria, that is, on evidence from more than one source: an ability test, a creativity measure, teacher ratings and samples of her work. In Saudi Arabia she could start along either of two pathways: registering for the National Program for Gifted Identification when she reaches a grade it targets, or the nomination criteria for the Ministry of Education's gifted classes.

**A running student file: Sultan's file at the outset**

Sultan's file can serve as a running exercise, with each chapter of Parts One and Two adding what is relevant to it: scores and confidence intervals, results from formal tools, repeated measurement data, and then the application of an identification model. Here is what the file contains now, organized by the **RIOT/ICEL** framework:

**Table 1.10: Sultan's file at the outset: what we have and what is missing.**

| Cell | What we have | What is missing |
|---|---|---|
| **Review / Learner** | Reading screening result: "Needs monitoring" | Detailed screening scores, results from previous years, and samples of his writing and reading |
| **Review / Environment** | Nothing | Attendance record, and whether he has moved between schools or teachers |
| **Interview / Environment** | Nothing | Asking the family about reading at home, and about hearing, vision and developmental history |
| **Observation / Instruction and curriculum** | Nothing | Observing the reading lesson: Are the texts at his level? Is decoding taught explicitly? |
| **Testing / Learner** | Nothing yet | Brief, repeated measures of reading fluency during the weeks of support |

The question that remains open is this: Is his weakness in reading a response to instruction that did not suit him, or a difficulty that persists despite appropriate instruction? It can only be answered after weeks of documented support.

## Conclusion

The argument we have developed in this chapter holds that assessment begins with the decision to be made and then selects its tool, and that each step in the pathway has its own question: screening sorts, identification gathers initial evidence and refers, diagnosis matches criteria, and eligibility determines the need for services. It follows that no step should be asked for what it cannot provide. A screening result does not establish a disability, a medical diagnosis does not automatically confer educational eligibility, and a teacher's opinion is the beginning of identification. It follows, too, that evidence should be gathered from multiple tools and sources, that instruction, curriculum and environment should be examined before a shortfall is attributed to the student, and that the family should be a partner in the decision as well as a source of information.

Some of what this chapter has presented, however, is a matter of terminology or of judgment. Using “identification” for the stage before referral is a terminological choice we have adopted for this book; the meaning of the term differs between the field of giftedness and the field of disabilities; medical and educational categories do not always coincide; and combining criterion-based judgment with a description of functioning in context calls for professional judgment that no ready-made rule can supply. Deciding whether the evidence is sufficient to move from one step to the next remains a professional judgment, one that must be reasoned and that weighs the numbers, the documents and all the other evidence.

# 1.10 Summary, glossary and activities

## Chapter summary

- Assessment is a means to a decision, and its first question is: What decision do we want to make?
- Screening sorts, identification gathers initial evidence and refers, diagnosis matches criteria, and eligibility determines the need for services. Using “identification” for the stage before referral is a terminological choice made for this book.
- A screening tool is judged by its sensitivity and its specificity together, and its negative predictive value tells us how confident we can be that we have not missed anyone.
- Medical diagnosis and educational eligibility are two different questions, and medical and educational categories do not always coincide.
- The Diagnostic and Statistical Manual is categorical at its core, the biopsychosocial model underpins the International Classification of Functioning, and good practice combines criterion-based judgment with a description of functioning in context.
- In Saudi Arabia, diagnosis is governed by the Regulations for Special Education Institutes and Programs (1422 AH) and the Special Education Organizational Guide (1436–1437 AH); the Law on the Rights of Persons with Disabilities (1445 AH) establishes the right to diagnosis and early detection; and the identification of gifted students follows a separate pathway.
- Fair assessment draws on multiple tools and sources, is conducted in the student's language, uses tools valid for their purpose, and rests on the family's informed consent.

- The **RIOT/ICEL** framework is a reminder that the problem may lie in instruction, curriculum or the environment, as well as in the student.

## Glossary of terms

| | |
|---|---|
| **Screening** | A quick procedure applied to an entire group to determine who needs a more in-depth assessment. |
| **Identification** | In this book: the stage of gathering initial evidence and attempting support within the classroom, in preparation for a referral decision. In much of the literature, the term denotes the entire process of determination up to the decision; in the field of giftedness, it is a multi-criteria educational process for determining who needs gifted services, not a clinical diagnosis. |
| **Identification; detection** | A term that overlaps in the Arabic literature with "identification" and is generally used for the process of locating and identifying gifted students or students with disabilities, as in the "National Program for Gifted Identification" and the "early detection of disability." |
| **Referral** | A formal request for a comprehensive assessment of the student, accompanied by the evidence and the family's consent. |
| **Nomination** | A proposal by a teacher, the family or the student himself that a student be assessed within a gifted identification program. |

| | |
|---|---|
| **Diagnosis** | A judgment that the criteria for a particular category apply to the student, based on a comprehensive assessment. |
| **Eligibility** | A decision on whether the student is entitled to special services, based on the diagnosis and its impact on his learning. |
| **Sensitivity** | The ability of a tool to capture those who actually have the problem; a property strongly required of screening tools. |
| **Specificity** | The ability of a tool to rule out those who do not have the problem. |
| **Negative predictive value** | The proportion of those with a negative result who are actually free of the problem; it is affected by the prevalence of the problem in the group. |
| **Informed consent** | The family's agreement after learning what will be done, why, and who will have access to the results. |
| **Diagnostic overshadowing** | Attributing new symptoms to a known disability, which leads to another condition being missed. |
| **Confirmation bias** | The assessor's tendency to see the evidence that supports a prior hypothesis and to overlook the evidence against it. |
| **Cut-off score** | The threshold that a scale or a system sets for making a decision. |

| | |
|---|---|
| **Multidisciplinary team** | A group of specialists from different fields who take part in assessing the student, each usually working alone. |
| **Interdisciplinary team** | A team in which each member assesses the student within his or her own field; members then share their findings and reach a joint decision and plan together. |
| **Transdisciplinary team** | A team that assesses the child together in a single session, with one member working directly with the child under the guidance of the others. |
| **Biopsychosocial model** | A model that views disability as the outcome of an interaction between the individual's health condition and environmental and personal factors. |
| **Review, Interview, Observe, Test / Instruction, Curriculum, Environment, Learner** | A matrix for gathering evidence through four methods (review, interview, observation and testing) across four domains (instruction, curriculum, environment and the learner). |

## Review questions

1. Identify the purpose of the assessment and the appropriate type of tool in each situation: (a) a school wants to know which of all its first graders need monitoring in reading; (b) a teacher wants to know whether the plan for a student with a difficulty in arithmetic is working; (c) an education directorate wants to know whether its gifted program is meeting its objectives.
2. What is the difference between diagnosis and eligibility? Give an example of a student who has a diagnosis but is not eligible for services.

3. Why is it preferable for screening tools to be sensitive, even if they sometimes flag students who do not need help? And why is a school principal more concerned with a tool's negative predictive value than with its positive predictive value?
4. This book uses "identification" for the stage before referral, while the "National Program for Gifted Identification" uses the word "al-kashf" (detection). What is the difference between the two meanings? And why does this difference matter to readers of the literature in other languages?
5. How does the diagnostic question in the Diagnostic and Statistical Manual (**DSM-5-TR**) differ from that in the International Classification of Functioning (**ICF**)? Which of the two expresses the biopsychosocial model more clearly?
6. Under which category is a student diagnosed with attention-deficit/hyperactivity disorder served under US law, and under which in the Saudi Organizational Guide? What corresponds in the Guide to the category of "traumatic brain injury"? And where does giftedness fall in each system?
7. A student with low vision obtained a low score on an intelligence test that relies on small pictures. Which principle of fair assessment was violated? Can you find a provision in Saudi regulations that protects the student from this error?
8. The Organizational Guide provides that a guardian's failure to respond within 14 days is treated as consent. Discuss this provision in light of the principle of informed consent, and propose a practice that safeguards the family's rights without stalling the process.
9. Fill in the cells of the **RIOT/ICEL** matrix for a second grader suspected of having a reading difficulty, and indicate the source of each piece of information.
10. What is diagnostic overshadowing? How might it delay the detection of attention-deficit/hyperactivity disorder in a student diagnosed with a learning disability?

11. **Case for application:** Ghada is thirteen and in the first year of intermediate school. Her mother has brought the school a report from a private clinic stating that she has "attention-deficit/hyperactivity disorder, predominantly inattentive presentation"; the report was based on an interview with the mother and a rating scale completed by the mother alone. The mother is asking for Ghada to be admitted to the program for attention-deficit/hyperactivity disorders. Her mathematics teacher says that Ghada forgets her homework and does not finish tests on time, while her science teacher sees no problem. Her grades were excellent throughout elementary school and have dropped this year in mathematics alone. She has received no classroom adjustments, and her hearing and vision have not been checked since she started school. (a) At which step of the general pathway does Ghada actually stand? (b) Is the available evidence sufficient for an eligibility decision under the Organizational Guide? Justify your answer. (c) Identify three cells of the **RIOT/ICEL** matrix that need evidence before any decision, and rank them by priority. (d) Which of the common errors in Table 1.9 might the team fall into here? And what question would you ask the mother about when the difficulties began?

**Table 1.11: Alignment of chapter objectives with review questions and learning activities.**

| Objective | Assessed by |
|---|---|
| **1. Identify the purpose of assessment and the type of tool** | Question 1 |
| **2. Distinguish among the four steps of the pathway** | Questions 2, 3 and 4 |
| **3. Compare models of understanding disability** | Question 5 |
| **4. Map Saudi categories onto US and medical categories** | Question 6 and Activity 1 |
| **5. Identify which principle of fair assessment has been violated** | Questions 7 and 8 |

| Objective | Assessed by |
|---|---|
| **6. Build a RIOT/ICEL matrix** | Question 9 and Activity 2 |
| **7. Frame the conversation with the family** | Activity 3 |
| **8. Detect common errors** | Question 10 and part (d) of Question 11 |
| **9. Judge the sufficiency of evidence** | Question 11 (Case for application) |

## Learning activities

General caution: Any activity involving real students or families is carried out with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a specialist if it involves administering a tool. Nothing gathered for training purposes may serve as the basis for a decision, and its results are communicated only through official channels.

1. **Mapping the local pathway:** Draw a student's pathway from observation to the plan under Saudi regulations, using Tables 1.5 and 1.8. Then identify three issues that the documents reviewed in this chapter either do not address or address only in general terms, and propose an alternative regulatory wording for each. If possible, ask a special education teacher about actual practice in his or her school on these issues, without mentioning any student. *Output:* A one-page diagram and a comparative paragraph. *Approximate time:* Two hours.
2. **A matrix for a case:** Fill in the **RIOT/ICEL** matrix for Ghada's case in Question 11, or for a student you know, with his or her identity concealed and under the conditions set out in the general caution. Which cells do you have information for? Which need someone to fill them in, and who is that? *Output:* A completed matrix with a column for the source of each piece of information. *Approximate time:* 45 minutes.
3. **A conversation with a family (paired activity):** One partner plays the teacher and the other the guardian; the teacher practices communicating a

concern using the five steps in Section 1.8, and then the partners switch roles. *Output:* Written notes from each partner on the other's performance at each step. *Approximate time:* 30 minutes.

## Chapter references

Al-Jazirah. (2018, February 14). [News report on the Council of Ministers' approval of the organizational statute of the Authority for the Care of Persons with Disabilities] [In Arabic]. https://www.al-jazirah.com/2018/20180214/ln36.htm

Saudipedia. (n.d.). *[Authority for the Care of Persons with Disabilities]* [In Arabic]. Retrieved September 29, 2026, from https://saudipedia.com/en/authority-of-people-with-disability

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Personal Data Protection Law, Royal Decree No. (M/19) (9/2/1443 AH), as amended by Royal Decree No. (M/148) (5/9/1444 AH) [In Arabic].

Disability Welfare Law, Royal Decree No. (M/37) (23/9/1421 AH) [In Arabic]. https://www.un.org/development/desa/disabilities/wp-content/uploads/sites/15/2019/11/Saudi-Arabic_Disability-Law.pdf

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH-a). *[Special Education Organizational Guide]* (1st ed., for the 1436–1437 AH school year) [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (1436 AH-b). *[Special Education Procedural Guide]* (1st ed., for the 1436–1437 AH school year) [In Arabic]. https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/اللائحة%20الإجرائية%20للتربية%20الخاصة%201437%20هـ.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alfaiz, F. S., Alfaid, A. A., & Aljughaiman, A. M. (2022). Current status of gifted education in Saudi Arabia. *Cogent Education, 9*(1), Article 2064585. https://doi.org/10.1080/2331186X.2022.2064585

Almulhim, N., Aljughaiman, A., Alnaim, F., Alsayed, A., Alfawzan, S., & Alarfaj, A. (2026). Hidden potential: An exploratory study of the prevalence of twice-exceptional students in Saudi Arabia. *Journal of Intelligence, 14*(5), Article 83. https://doi.org/10.3390/jintelligence14050083

Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications, 13*(1), Article 974. https://doi.org/10.1057/s41599-026-07376-9

Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties, 30*(1), 75–92. https://doi.org/10.1080/19404158.2025.2473425

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Assistance to States for the Education of Children With Disabilities, 34 C.F.R. pt. 300 (2006) (§300.8, §300.9, §300.111, §300.300, §300.301, §300.303, §300.304, §300.306, §300.307, §300.309, §300.320).

Christ, T. J. (2008). Best practices in problem analysis. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 159–176). National Association of School Psychologists.

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), 20 U.S.C. § 1400 et seq. (2004).

National Association for Gifted Children. (2019). *2019 Pre-K–Grade 12 gifted programming standards*.

National Association of School Psychologists. (2020). *The professional standards of the National Association of School Psychologists*.

Oliver, M. (1990). *The politics of disablement*. Macmillan.

Reiss, S., Levitan, G. W., & Szyszko, J. (1982). Emotional disturbance and mental retardation: Diagnostic overshadowing. *American Journal of Mental Deficiency, 86*(6), 567–574.

Salvia, J., Ysseldyke, J. E., & Witmer, S. (2017). *Assessment in special and inclusive education* (13th ed.). Cengage Learning.

World Health Organization. (2001). *International classification of functioning, disability and health (ICF)*.

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

# 2 Chapter 2 Measurement fundamentals

What does a score of 78 mean? How does a percentile rank differ from a percentage? And why does the psychologist write two more numbers on either side of a score? In this chapter we provide the keys for reading assessment reports of every kind, and we show what a number can tell us and where its meaning stops.

---

**Scene from a team meeting**

Umm Majed sits in the school meeting with a six-page report in front of her. She reads: “Full-scale score 72, percentile rank 3, confidence interval 65 to 79.” She looks up and asks the teacher: “Does this mean my son passed with three percent? Does he have a disability or not?”

The teacher hesitates. He knows the number is not a pass rate, but he does not know how to explain that, nor why the report lists two additional numbers around the score.

We return to Umm Majed at the end of the chapter to give her two questions a precise answer. The concepts presented here recur throughout the category chapters, for they are the shared language in which every diagnostic report is written.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Classify a reported assessment result as norm-referenced or criterion-referenced, and justify the classification by the question the result answers.
2. Convert a score among the common standard score metrics, and locate it on the normal curve and within its descriptive classification.
3. Interpret a percentile rank correctly, and correct a statement that confuses it with a percentage or an age equivalent.
4. Calculate the standard error of measurement, the confidence interval and the standard error of the difference, and use them to judge a score close to a cut-off or a difference between two scores.
5. Calculate positive and negative predictive values from sensitivity, specificity and prevalence, and explain why they differ between universal screening and a referred group.
6. Evaluate whether an instrument is appropriate for a particular student in light of the five sources of validity evidence, reliability, norms, language and fairness evidence.
7. Identify the ethical or professional principle at stake in a described assessment situation, and propose a course of action consistent with it.

# Introduction

Assessment results reach teachers and families as numbers: a standard score, a percentile rank, a confidence interval and a descriptive classification. The percentile rank is the percentage of the student's peers in the norming sample (the representative group from which the instrument's norms were derived) who scored lower than the student did; a percentile rank of 84 means the student outperformed 84% of them. Each of these numbers carries a specific meaning and has limits it cannot exceed, and misreading it leads to a

mistaken decision: a percentile rank taken as the proportion of correct answers, as Umm Majed understood it in the opening scene; a score treated as an exact figure free of error; or a firm decision made when the confidence interval straddles the cut-off. In most categories, the diagnostic judgment rests on where the student stands relative to peers, on how precisely that position is estimated, and on whether the instrument suits its purpose and the student to whom it was administered.

We open the chapter by distinguishing norm-referenced from criterion-referenced measures (Section 2.1), then present the mean and standard deviation (Section 2.2), the normal curve, standard scores and their descriptive classifications (Section 2.3), and percentile ranks and their pitfalls (Section 2.4). We then take up reliability/precision and the standard error of measurement, together with the related matters of score differences, regression toward the mean and floor and ceiling effects (Section 2.5), followed by validity, sensitivity, specificity and fairness (Section 2.6), and then norms, standardization and test adaptation (Section 2.7). We apply all this to reading a realistic score table (Section 2.8), and close with ethics, professional responsibility and levels of qualification (Section 2.9).

## 2.1 Norm-referenced and criterion-referenced measures

Before we ask “How much did the student score?” we must ask “Compared with what?” A score on its own has no meaning. Eighty points may be excellent on a hard test and weak on an easy one. Measures are therefore divided into two main types according to what the score is compared with:

**Norm-referenced measures** Norm-referenced

**Where does the student stand relative to peers?**

**Criterion-referenced measures** Criterion-referenced

**Has the student mastered the target skill?**

The student's score is compared with the scores of a large group of same-age or same-grade peers, called the **norming sample**. The result tells us the student's position among them: average, above average or below average.

Examples: intelligence tests, adaptive behavior scales, and rating scales such as the Gilliam Autism Rating Scale, Third Edition (**Gilliam Autism Rating Scale, GARS-3**).

The student's score is compared with a predetermined level of performance. The result tells us what the student can and cannot do, regardless of peers.

Examples: "Reads 40 words per minute with 95% accuracy," or "Adds two two-digit numbers with regrouping on 8 of 10 problems."

The two types complement each other. Diagnosis usually calls for norm-referenced measures, because its question is "Is the student's performance unusually far from that of peers?" Instructional planning, by contrast, calls for criterion-referenced measures, because its question is "What skill should we teach next?"

A single tool may combine both logics. One example is **curriculum-based measurement**, a set of brief, repeated probes used to track a student's reading or mathematics week by week (defined and discussed in detail in Chapter 4). It serves as a general indicator of growth: its scores are interpreted normatively when compared with seasonal grade-level norms, such as fall, winter and spring percentile ranks, and at the same time against a specific growth goal set for the student. It thus draws on both frames of reference at once.

**An example of the difference**

Saad, a fourth grader, obtained a percentile rank of 50 on a standardized reading test, placing him exactly in the middle of his peers. This is norm-referenced information, and it does not point to a problem. On a criterion-referenced test, however, his teacher noticed that he could not distinguish between letters that share the same shape and differ only in their dots, such as ب (bāʾ), ت (tāʾ) and ث (thāʾ). The first test does not reveal this, yet it is precisely what the teacher needs to know in order to decide what to teach tomorrow.

## 2.2 The mean and the standard deviation

Every standard score is built on two numbers: the **mean** (**Mean**), the center of the norming sample's scores, and the **standard deviation** (**Standard deviation, SD**), the degree to which scores are spread around that center. A small standard deviation means people's scores lie close together; a large one means they are widely dispersed.

To make the idea of standard deviation concrete, imagine two classes of twenty students each, both with an average height of 140 centimeters. In the first class, heights range from 135 to 145; in the second, from 120 to 160. The mean is the same, but the spread differs. A student who is 150 centimeters tall would be considered very tall in the first class and fairly ordinary in the second. It is the standard deviation that tells us so.

This gives us the rule on which diagnosis rests: we measure how far a student lies from the mean in standard deviations rather than in points. A distance of two standard deviations from the mean signifies the same thing on any test, however its scores are calculated.

## 2.3 The normal curve and standard scores

When we measure a trait such as intelligence or height in a large number of people, the scores usually fall into a bell shape: most people cluster near the mean, and their numbers dwindle the further we move from it in either direction. This shape is the **normal curve** (**Normal curve**), and among its properties are fixed proportions that do not change:

- About 68% of people fall between one standard deviation below the mean and one standard deviation above it.
- About 95% fall between two standard deviations below the mean and two above it.
- Only about 2% fall more than two standard deviations below the mean, and roughly the same proportion more than two standard deviations above it.

This last point is the origin of thresholds that recur throughout this book. The criterion of "two standard deviations below the mean" is used for intellectual disability, and "two standard deviations above the mean" appears in many definitions of giftedness. By this statistical criterion alone, then, each of the two groups represents about 2% of people at either end of the distribution.

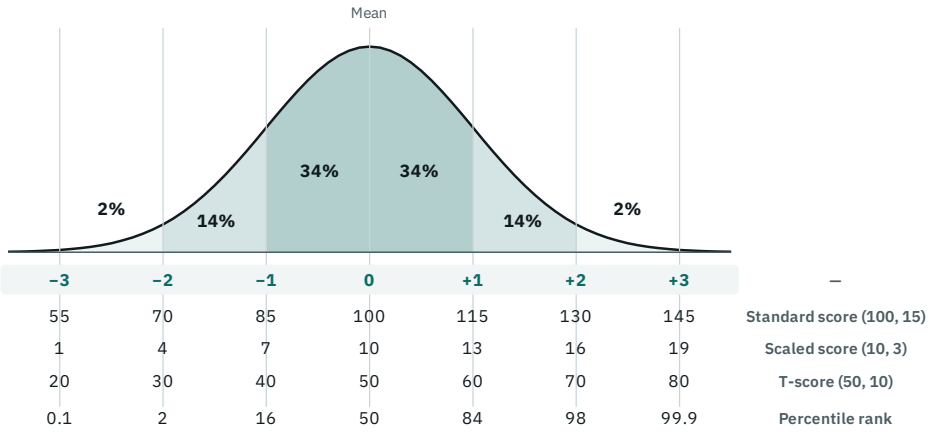



**Figure 2.1:** The normal curve, with the common score types aligned beneath it by position. Each column represents the same position expressed in different metrics: a full-scale score of 70 on an intelligence test is equivalent to a subtest score of 4 and to a T-score of 30, and all three mean two standard deviations below the mean.

## Types of standard scores

Tests convert the raw score, the number of correct answers, into a standard score with a fixed mean and standard deviation. The benefit is that standard scores can be compared across ages and tests. The most widely used types are:

**Table 2.1: Common types of standard scores.**

| Type | Mean | — | Where you will encounter it |
|---|---|---|---|
| **z-score (z)** | 0 | 1 | In calculations and research; it is the basis from which all other types are derived |
| **Standard score** | 100 | 15 | Composite and index scores on intelligence and achievement tests and adaptive behavior scales |
| **Scaled score** | 10 | 3 | Subtests of intelligence scales, and the subscales of the **GARS-3** |
| **T-score (T)** | 50 | 10 | Behavioral and emotional rating scales, including many measures of attention-deficit/hyperactivity disorder (**Attention-Deficit/Hyperactivity Disorder, ADHD**) |

To convert any score to another type, we first convert it to a z-score and then from the z-score to the desired type:

$$z = (X - M) \div SD$$

z-score = (student's score – mean) ÷ standard deviation

**Worked example**

1. Hind obtained a standard score of 82 on an intelligence test with a mean of 100 and a standard deviation of 15.
2. The z-score: (82 – 100) ÷ 15 = –1.2, meaning she is one and one-fifth standard deviations below the mean.
3. Expressed as a T-score, it would be: 50 + (–1.2 × 10) = 38.
4. Her percentile rank is about 12, meaning her performance exceeds that of about 12% of her peers in the norming sample.

⚠ **Watch the direction of the score**

On ability tests, a higher score means better performance. On many behavioral rating scales, however, a higher score means more problems. A T-score of 70 is excellent on a reading test and worrying on a hyperactivity scale. Always check the test manual or the report for what a high score means.

## Descriptive classifications

Tests translate the standard score into a verbal description that is easy to understand. The following example shows the classifications used by the Wechsler Intelligence Scale for Children, Fifth Edition (**Wechsler Intelligence Scale for Children, Fifth Edition, WISC-V**), together with the official English label for each category, since specialists frequently read reports written with these labels (Wechsler, 2014):

**Table 2.2: Descriptive classifications of composite scores on the Wechsler Intelligence Scale for Children, Fifth Edition.**

| Standard score range | Descriptive classification | Approximate percentage of people |
|---|---|---|
| **130 and above** | Extremely high | 2% |
| **120–129** | Very high | 7% |
| **110–119** | High average | 16% |
| **90–109** | Average | 50% |
| **80–89** | Low average | 16% |
| **70–79** | Very low | 7% |
| **69 and below** | Extremely low | 2% |

Three cautions apply to this table. First, the labels differ across tests, and even across editions of the same test. The fourth edition of this very scale called the 70–79 range "borderline" (**Borderline**) and anything above 130 "very superior" (**Very Superior**). So if you read the phrase **Very High** in a foreign report describing a score of 125, bear in mind that a higher category sits above it. Second, the "average" category here (90–109) is a descriptive classification specific to the Wechsler scales; it is narrower than the 85–115 range, which spans one standard deviation on either side of the mean and is the range used in this book's figures. Third, and most important, a descriptive classification is not a diagnosis. The label "extremely low" describes only where the score falls; a judgment of intellectual disability requires other criteria, such as adaptive behavior, that is, the everyday living skills the student actually uses at home and at school (Chapter 7 addresses this in detail).

## 2.4 Percentile ranks and their pitfalls

The percentile rank is, on the surface, the easiest score for families to understand, and it is also the one most prone to misunderstanding. It means **the percentage of peers in the norming sample who scored lower than the student**. A percentile rank of 30 means the student outperformed about 30% of his peers, and that about 70% of them outperformed him.

### The first pitfall: confusing it with percentage correct

The percentile rank has nothing to do with the percentage of correct answers. A student who answered 80% of the questions may have a percentile rank of 20 if the test was easy and most of his peers answered more than that. This is the trap Umm Majed fell into when she thought her son had "passed with three percent."

## The second pitfall: the distances between ranks are unequal

Because most people cluster around the mean, moving a few percentile ranks near the mean takes only a small difference in performance, whereas moving the same number of ranks at the extremes takes a large one. Compare these two moves, each spanning a similar number of ranks:

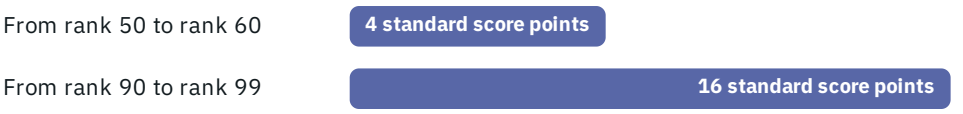



**Figure 2.2:** What each move is worth on a standard score scale with a mean of 100 and a standard deviation of 15. Ten ranks near the mean equal about 4 points, whereas nine ranks at the extreme equal about 16 points.

For this reason, percentile ranks should not be subtracted from one another or averaged, nor used to measure a student's progress. If a student moves from the 5th to the 10th percentile, his progress may be far greater than the two numbers suggest.

## The third pitfall: age and grade equivalents

Age equivalents (**Age equivalent**) and grade equivalents (**Grade equivalent**) express performance in terms of an age or a school grade. A teacher may read in a report that "the student's reading age equivalent is 7 years, 3 months," while the student is actually ten, and conclude that he reads like a seven-year-old. That reading is inaccurate. An age equivalent means only that the student's raw score equals the average raw score of children of that age; it does not mean that his skills or errors resemble theirs. These scores are also unevenly distributed, and they do not show whether the gap is large relative to the student's peers. This is why the Standards for Educational and Psychological Testing urge great caution in their use (AERA et al., 2014), and why standard scores and percentile ranks are preferred.

## 2.5 Reliability/precision and the standard error of measurement

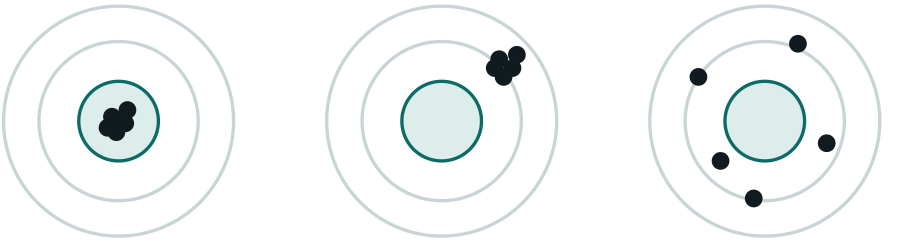
**Reliability** (**Reliability**) is the consistency of a measure's results. If the test were given to the same student twice under similar conditions, would he obtain two similar scores? A scale that shows a different weight each time is useless, however accurate it may be on average. The Standards for Educational and Psychological Testing use the compound term "reliability/precision" (**Reliability/precision**) to cover traditional reliability coefficients as well as every other index of score precision, such as the standard error of measurement and the information function, which we present at the end of this section (AERA et al., 2014).

### Types of reliability

**Table 2.3: Types of reliability and the question each one answers.**

| Type | Question it answers | How it is calculated |
|---|---|---|
| **Test–retest reliability** | Does the score remain stable over time? | The test is given twice with a short interval between administrations, and the correlation between the two is calculated |
| **Internal consistency** | Do the test items measure the same thing? | Calculated from the item intercorrelations in a single administration; the best known index is Cronbach's alpha (**Cronbach's alpha**) |
| **Interrater reliability** | Do two different raters assign the same score? | Two people rate the same student and their ratings are compared; this is especially important for observation and rating scales |
| **Alternate-forms reliability** | Do two forms of the test yield similar results? | Two equivalent forms are administered, and the correlation between them is calculated |

Reliability is expressed as a coefficient ranging from zero to one. Salvia and colleagues (Salvia et al., 2017) proposed widely used benchmarks: at least 0.60 when scores are used to describe groups, at least 0.80 for screening decisions, and at least 0.90 for important decisions about an individual, such as diagnosis and classification.



**Reliable and valid**
The results are consistent and hit what is meant to be measured

**Reliable but not valid**
The results are consistent but measure something else

**Not reliable**
The results are scattered, so the measure cannot be valid

**Figure 2.3:** The relationship between reliability and validity, illustrated with target shooting. Reliability is a prerequisite for validity, but it is not sufficient on its own.

—

No measure is perfectly reliable. A student's score on any given day is affected by fatigue, attention, guessing and the conditions in the room. The score we obtain is therefore an **estimate** of the true score, the score the student would have obtained had he been measured without any error. The standard error of measurement (**Standard Error of Measurement, SEM**) is a number that tells us how much the student's score would be expected to fluctuate if we measured him many times over:

$$SEM = SD \times \sqrt{(1 - r)}$$

standard error of measurement = standard deviation × square root of (1 – reliability coefficient)

From it we build a **confidence interval** (**Confidence interval**): a range of scores within which we can say, with a given degree of confidence, that the true score lies. For a 95% confidence interval, we multiply the standard error by approximately 1.96, then add the result to the score and subtract it from the score.

**Worked example: Majed's score**

1. Majed obtained a full-scale score of 72 on an intelligence test with a standard deviation of 15 and a reliability coefficient of 0.95.
2. The standard error of measurement: $15 \times \sqrt{(1 - 0.95)} = 15 \times 0.224 \approx 3.4$
3. The 95% confidence interval: $72 \pm (1.96 \times 3.4) = 72 \pm 6.6$, that is, from about 65 to about 79.
4. The result: we can be 95% confident that Majed's true score on this test lies between 65 and 79. This range **straddles the cut-off score of 70**.

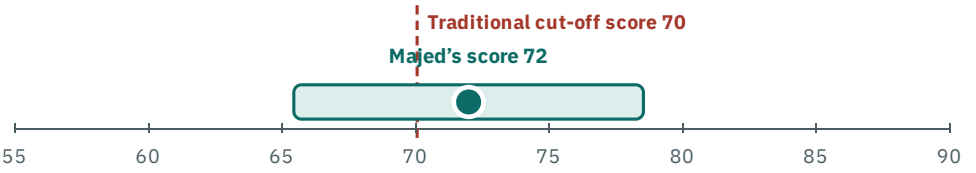



**Figure 2.4:** Majed's score and the confidence interval around it. The point lies above the cut-off score, but the band extends to either side of it, so we cannot say on the strength of the number alone that he "exceeded" it or "fell short" of it.

A **cut-off score** (**Cut-off score**) is the score that separates two categories in a decision, such as "refer" versus "do not refer" on a screening tool, or "meets the criterion" versus "does not meet it" for a diagnostic criterion. We use this term throughout the book.

It is precisely for this reason that the American Association on Intellectual and Developmental Disabilities (**American Association on Intellectual and Developmental Disabilities, AAIDD**) calls for the standard error of measurement to be taken into account when determining significant limitations in intellectual functioning, so that 70 is not treated as a hard dividing line (Schalock et al., 2021). A score of 72 or 74 does not automatically rule out intellectual disability, and a score of 68 does not automatically establish it. What settles the matter is the convergence of evidence, especially evidence of adaptive behavior.

**The rule to take with you**

Every score in a diagnostic report is a point inside a band. The less reliable the measure, the wider the band. And when the band falls on both sides of the cut-off score, the decision becomes a professional judgment that rests on other evidence besides the number.

**For the specialist: Why are confidence intervals in reports not symmetrical around the score?**

The way we calculated Majed's interval is a simplified method that centers the interval on the observed score itself. Many modern tests, however, including the Wechsler Intelligence Scale for Children, center the interval on the **estimated true score** (**Estimated true score**) and use the **standard error of estimate** (**Standard error of estimation, SEE**) to do so (Sattler, 2018):

$$T' = M + r\,(X - M)$$

$$SEE = SD \times \sqrt{r\,(1 - r)}$$

estimated true score = mean + reliability × (observed score − mean); standard error of estimate = standard deviation × square root of [reliability × (1 − reliability)]

Because reliability is less than one, the estimated true score always lies closer to the mean than the observed score does. Applied to Majed, his estimated true score would be $100 + 0.95 \times (72 - 100) \approx 73.4$ and the standard error of estimate about 3.3, so the interval would run from about 67 to about 80. The interval is thus skewed toward the mean rather than symmetrical around 72, yet it straddles the cut-off score of 70 under both methods. This is why readers see intervals in report tables such as "84: 78–92" and take them for a typo, when they are nothing of the kind (see Table 2.8). When you read a report, check the test manual for how its intervals were constructed before comparing them with the intervals of another test.

**In the Saudi context: Cut-off scores in the regulations**

The Special Education Organizational Guide stipulates that among the conditions for a student's eligibility for intellectual disability programs is "that the student's IQ not exceed 75 on the Wechsler test, or 69 on the Stanford-Binet (Fifth Edition), or the equivalent of either," together with limitations in at least two areas of adaptive skills. It also requires "not relying on the results of a single scale" and that instruments be administered and interpreted "by qualified specialists" (Ministry of Education, 1436 AH [Arabic]). The earlier Regulations had defined the "educable" range as an IQ between 55 and 75 on the Wechsler or between 52 and 73 on the Stanford-Binet (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). The different figures in the Regulations can be understood in light of the different standard deviation of the older editions of the Stanford-Binet (16 rather than 15); both fall at roughly the same distance from the mean. In the current Guide, however, the current editions of the two scales share the same standard deviation (15), and we found no rationale in the text for the two different figures (75 and 69). Nor did we find any provision directing that the standard error of measurement or the confidence interval be taken into account when these figures are applied.

**A question for reflection:** A student scores 77 on the Wechsler scale, with a confidence interval of 72–83, and shows clear limitations in three adaptive domains. How should the team handle the cut-off score of 75 in light of what you have learned in this section?

## 2.5.1 Differences between scores

Many diagnostic decisions rest on a **difference** between two scores: between ability and achievement in the discrepancy model, between two indexes on an intelligence test, or between a student's scores before and after an intervention.

Each of the two scores carries its own error, and the two errors combine in the difference. A difference is therefore less precise than either score, and it must be relatively large before we can be confident that it is real.

When we find a difference between two scores, we ask two distinct questions. The first concerns **statistical significance**: is the difference too large to be attributed to measurement error alone? The second concerns **rarity**, or the **base rate** (**Base rate**): what proportion of the norming sample itself shows a difference of this size or larger? A difference may be significant, that is, real rather than a product of error, and yet common among typically developing children, in which case it cannot be treated as a sign of a disorder. Good test manuals present the two in separate tables.

**For the specialist: The standard error of the difference**

When the two scores share the same standard deviation, the **standard error of the difference** (**Standard error of the difference, $SE_{diff}$**) is calculated as follows:

$$SE_{diff} = SD \times \sqrt{(2 - r_1 - r_2)}$$

standard error of the difference = standard deviation × square root of (2 – reliability of the first score – reliability of the second score)

**Worked example: A difference between two indexes (from Table 2.8)**

1. The Verbal Comprehension Index is 84, with a reliability of 0.94, and the Visual Spatial Index is 102, with a reliability of 0.92. The difference is 18 points.
2. The standard error of the difference: $15 \times \sqrt{(2 - 0.94 - 0.92)} = 15 \times \sqrt{0.14} \approx 15 \times 0.37 \approx 5.6$
3. The critical difference at the 0.05 level of significance: $1.96 \times 5.6 \approx 11$. The observed difference (18) exceeds it, so it is **statistically significant**.
4. Next we ask about rarity. Suppose, purely for illustration, that the base rate table in the manual shows that about 15% of children in the norming sample have a difference of this size or larger between these two indexes. The difference is real, but it appears in roughly one child in seven, so it is **not rare** and cannot by itself serve as evidence of a disorder.
5. Compare this with the difference between Working Memory (76, reliability 0.93) and Processing Speed (80, reliability 0.90): the standard error of the difference is $15 \times \sqrt{0.17} \approx 6.2$, the critical difference about 12, and the observed difference only 4. We cannot, then, say that processing speed is “stronger” than working memory.

The reliability of the difference score itself has a formula that shows the problem from another angle: $r_{diff} = [(r_1 + r_2) \div 2 - r_{12}] \div (1 - r_{12})$, where $r_{12}$ is the correlation between the two scores. If the two indexes had reliabilities of 0.94 and 0.92 and correlated at 0.60, the reliability of the difference would be $(0.93 - 0.60) \div 0.40 \approx 0.83$, lower than that of either index. The higher the correlation between two scores, the lower the reliability of the difference between them. This is the root of the criticism leveled at decisions based on differences alone, as in the ability–achievement discrepancy model and in patterns of strengths and weaknesses models (to which Chapter 6 returns).

## 2.5.2 Regression toward the mean

When people who obtained extreme scores are measured again, their average score the second time tends to move closer to the population mean. This statistical phenomenon is called **regression toward the mean** (**Regression toward the mean**). It occurs because an extreme score usually combines genuinely high (or low) ability with measurement error that happened to fall in the same direction, and that error does not necessarily recur the second time. The estimated true score we met in the previous "For the specialist" box is a direct application of this idea.

The phenomenon has practical consequences in two places. The first is giftedness: if we select students who scored above 130 on a test with a reliability of 0.90, the expected score of a student who obtained 135, on a retest with the same instrument, is about $100 + 0.90 \times 35 \approx 132$, and it drops further if the second test is a different one that correlates with the first at only about 0.80 (to about 128). A decline in the scores of some selected students at reevaluation is therefore to be expected; it means neither that the original decision was wrong nor that the student has "lost" his giftedness (Lohman & Korb, 2006). The second is disability and learning disabilities: some of those selected for their

very low scores will score higher when remeasured even without any intervention, so this gain cannot be credited entirely to the program in the absence of a comparison group.

### 2.5.3 Floor, ceiling and item gradient

A test's **floor** (**Floor**) is its capacity to discriminate among very low levels of performance, and its **ceiling** (**Ceiling**) is its capacity to discriminate among very high levels. If a test has few easy items, a child with very low performance and a child with low performance will both receive the same minimum score, and the test can no longer tell them apart. Likewise, if there are few difficult items, high performers bunch up at the highest possible score. On many individual intelligence tests, the full-scale score range bottoms out at about 40 and tops out at about 160. That suffices for most children, but not in two situations: distinguishing levels of severity among the more severe cases of intellectual disability, to which Chapter 7 returns, and distinguishing among exceptionally gifted students, to which Chapter 12 returns. For this reason, some publishers have issued **extended norms** (**Extended norms**) for both tails of the distribution.

Closely related is the **item gradient** (**Item gradient**), the amount by which the standard score changes when the raw score increases by one point. If one additional correct answer moves a child from a scaled score of 4 to 7, the test is coarse in that range, and a small scoring error or a momentary lapse can change the whole classification. Bracken proposed minimum standards of technical adequacy that have since become widespread in the literature, among them that a raw score of 1 on a subtest should place the child more than two standard deviations below the mean if the floor is to be adequate, and that a one-point increase in raw score should not produce a change of more than a third of a standard deviation (Bracken, 1987). The problem is most pronounced with young children and with children at the youngest or oldest ages a test covers. If a child is at the edge of a test's age range, examine the norm tables before trusting the score.

### 2.5.4 Item response theory

Everything we have presented so far belongs to **classical test theory**, which assumes a single standard error for all scores on a test. Most modern tests, however, were built, wholly or in part, within another framework: **item response theory** (**Item Response Theory, IRT**). Teachers need only two ideas from it: that a test's precision varies across ability levels and is highest where its items are concentrated; and that this is what makes computerized adaptive testing possible, in which the computer selects for each student the items suited to his level.

**For the specialist: The item characteristic curve and the information function**

The theory describes each item by an **item characteristic curve** (**Item characteristic curve**), which links the student's ability, denoted θ and expressed on a scale resembling the z-score, to the probability of answering the item correctly. The curve's most important features are two parameters: **difficulty** (b), the ability level at which the probability of a correct answer reaches 50%, and **discrimination** (a), the slope of the curve at that point. A third parameter for guessing may be added for multiple-choice questions (Embretson & Reise, 2000).

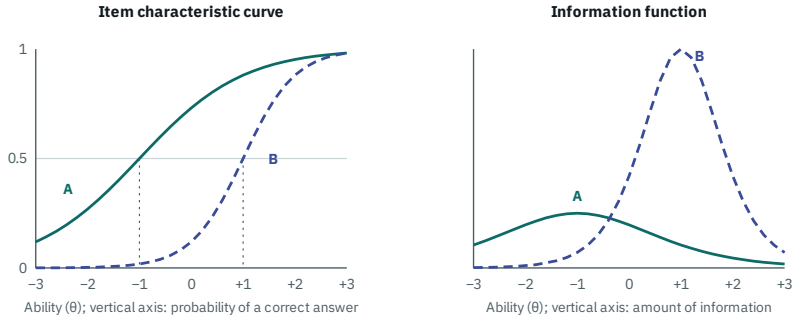



**Figure 2.5:** Two hypothetical items. Item "A" (solid line) is easy and moderately discriminating, and item "B" (dashed line) is difficult and highly discriminating. In the left panel, the probability of a correct answer reaches 50% at θ = −1 for item "A" and at θ = +1 for item "B." In the right panel, each item yields the most information near its difficulty, and the more discriminating item provides more information over a narrower range.

From the curve we derive the **information function** (**Information function**), which shows how much the item contributes to the precision of the ability estimate at each level. What all the items contribute is summed in the **test information function**, and the standard error of the ability estimate at any level equals the inverse of the square root of the information at that level. A test whose items are concentrated around the mean therefore measures

average students with high precision and the two tails with less, which is a more precise formulation of the floor and ceiling problem discussed above. A **computerized adaptive test** (**Computerized adaptive testing, CAT**) begins with an item of moderate difficulty, estimates the student's ability after each response, and selects the next item from the item bank so as to yield the most information at that estimate. It thus reaches acceptable precision with fewer items, and each student faces items close to his level. The validity of this approach, however, depends on calibrating the item bank on a sample resembling the students we are measuring, a requirement that brings us back to Section 2.7.

## 2.6 Validity

If reliability asks, "Does the scale measure consistently?", validity asks, "Does it actually measure what it is meant to measure? And are the decisions we base on its scores sound?" In the modern view, validity is a property of the interpretation of a scale's scores for a particular purpose, and it changes with the purpose and with the person tested. A scale that is valid for screening three-year-olds may not be valid for diagnosing an adolescent. The Standards for Educational and Psychological Testing treat validity as a single, unified concept, sometimes called **construct validity**: the degree to which the evidence supports interpreting the score as representing the intended construct. The construct is what all the evidence, taken together, attests to. The Standards draw this evidence from **five sources** (AERA et al., 2014):

**Table 2.4: The five sources of validity evidence in the Standards for Educational and Psychological Testing, with examples from adapting scales for Arabic-speaking settings.**

| Source of evidence | Question | Example |
|---|---|---|
| **Content** | Do the items cover the facets of the construct adequately, and do they suit the setting in which they are used? | An autism spectrum disorder scale whose items cover the two core domains in the diagnostic manual: difficulties in social communication and interaction, and restricted, repetitive behaviors and interests. For the Arabic version, local experts review each item: does it describe a behavior one would expect to see in a Saudi home or school? |
| **Response processes** | Do test takers or raters actually engage the processes the construct assumes when they respond? | Saudi children are asked to think aloud while solving the items of a reasoning test, or parents are asked what they understand by a statement on a rating scale. If it turns out that a child fails because he does not know a Standard Arabic word in the question, the item is measuring vocabulary, not reasoning |
| **Internal structure** | Do the items and dimensions relate to one another as the theory predicts? | Factor analysis of the Arabic version: does it retain the structure of the original? A factor analysis of the Arabic version of the Childhood Autism Rating Scale arrived at a three-factor model of 14 items (Alotaibi & Alotaibi, 2021), and an Omani study of the Adaptive Behavior Assessment System (ABAS-3) supported a single general-factor model among children with intellectual disability (Emam et al., 2020) |

| Source of evidence | Question | Example |
|---|---|---|
| **Relations to other variables** | Do scores correlate with what they are expected to correlate with, predict the criterion, and distinguish between groups? | The Saudi version of the Wechsler Preschool and Primary Scale of Intelligence, Third Edition, correlated with the Saudi version of the Stanford-Binet at 0.81 (Ejbara, 2024). In a validation of the Arabic version of the Modified Checklist for Autism in Toddlers across nine Arab countries, sensitivity reached 0.86 and specificity 0.80 (Seif Eldin et al., 2008) |
| **Consequences** | What are the effects, intended or unintended, of using the scores in decisions? | Does using a scale for nomination to gifted programs lead to lower-than-expected representation of students from villages or of a particular group? (Chapter 14 returns to this question) |

The older division into "content validity," "criterion validity" and "construct validity" still appears in many test manuals and Arabic textbooks. In that scheme, "criterion validity" corresponds to part of the fourth source, while "construct validity" corresponds to validity as a whole in the modern view. The second and third sources, however, are the most important questions we ask when evaluating an Arabic adaptation: Does the Arab child understand the item as intended? And has the adaptation retained the structure of the original?

## Sensitivity and specificity, and why the value of a positive result changes

For screening and diagnostic tools, validity evidence based on relations to other variables is usually reported through two concepts: **sensitivity** (**Sensitivity**), the proportion of those who have the condition whom the tool correctly detects, and **specificity** (**Specificity**), the proportion of those who do

not have the condition whom the tool correctly rules out. Yet two other figures matter more to teachers and families. The first is the **positive predictive value** (**Positive predictive value**): if a student's result is positive, how likely is it that he actually has the condition? The second is the **negative predictive value** (**Negative predictive value**): if his result is negative, how likely is it that the condition is truly absent? This second figure is the most important question in screening, because a negative result means the student will not be followed up.

These two probabilities depend jointly on the tool and on how prevalent the condition is in the group being tested. Consider a hypothetical example with illustrative figures: a screening tool with 85% sensitivity and 95% specificity is administered once to 1,000 children in a universal screening where the condition has a prevalence of 2%, and once to 200 children referred because of concerns, among whom the prevalence is 30%:

**Table 2.5: The same tool (sensitivity 85%, specificity 95%) in two groups with different prevalence. Hypothetical figures.**

| | Universal screening (1,000 children, prevalence 2%) | Referred group (200 children, prevalence 30%) |
|---|---|---|
| **Have the condition** | 20 | 60 |
| **True positive (detected by the tool)** | 17 | 51 |
| **False negative (missed by the tool)** | 3 | 9 |
| **Do not have the condition** | 980 | 140 |
| **False positive** | 49 | 7 |
| **True negative** | 931 | 133 |

| | Universal screening (1,000 children, prevalence 2%) | Referred group (200 children, prevalence 30%) |
|---|---|---|
| **Positive predictive value** | 17 ÷ 66 ≈ 26% | 51 ÷ 58 ≈ 88% |
| **Negative predictive value** | 931 ÷ 934 ≈ 99.7% | 133 ÷ 142 ≈ 94% |

In universal screening, roughly three-quarters of the positive results are false, yet a negative result is highly trustworthy. In the referred group the picture is reversed: most positive results are correct, but about one in every sixteen children with a negative result actually has the condition. This is why a positive screening result is not a diagnosis, and why a negative result in a high-risk group is not sufficient reason to close the file.

## 2.6.1 Fairness and bias in measurement

The Standards for Educational and Psychological Testing make **fairness** (**Fairness**) a third foundation of measurement alongside validity and reliability/precision, devoting their third chapter, “Fairness in Testing,” to it (AERA et al., 2014). Fairness in this sense is at heart a question of validity: Does a score mean the same thing for a student from a different linguistic, cultural or geographic group, or for a student with a disability? A failure in this respect is called **measurement bias**, and it takes several forms:

**Table 2.6: Forms of measurement bias and how they are examined.**

| Form | What it means | How is it examined? |
|---|---|---|
| **Construct bias** | The construct itself is not defined, or does not manifest, in the same way in the two groups, as when "appropriate social behavior" is measured by the standards of a single culture | Expert review from both cultures, and comparison of factor structures |
| **Differential item functioning (Differential item functioning, DIF)** | Two students of equal ability from different groups do not have the same probability of answering a particular item correctly | Comparing the two groups' performance on each item after matching them on total score or estimated ability; experts then review the items on which a difference appears |
| **Measurement non-invariance** | The relationship between the items and the construct differs across the two groups, so their scores cannot be compared directly | Sequential tests in multigroup confirmatory factor analysis: equivalence of structure, then of loadings, then of intercepts |
| **Predictive bias** | The same score predicts different later performance in the two groups | Comparing the regression lines of criterion on score in the two groups: a difference in slope or intercept is evidence of bias |

A difference in group means may be real; the two groups may genuinely differ in the measured skill for reasons related to opportunities to learn, and the scale may be describing that difference accurately. Bias is when the same score means something different depending on the group. Nor is examining the instrument enough: fairness also includes making the construct accessible to every test taker, as when a student with low vision is given the test in large

print, provided this does not change what it measures. Fairness in the referral and identification stages, and indices of proportional representation, are taken up again in Chapter 14.

## 2.7 Norms, standardization and test adaptation

A standard score means "the student's position among his peers in the norming sample." The **norming sample** (**Norm sample**) is the group to which the scale was administered to derive its norms. If the norming sample does not resemble the student, the score loses its meaning. We therefore ask three questions of any instrument before using it:

1. **Who is in the norming sample?** Does it represent the children we are assessing in age, environment and language? A test normed on American children compares a Saudi child with American children rather than with peers in his own environment.
2. **When was it normed?** Average performance on intelligence tests has risen across generations in many countries, a phenomenon known as the Flynn effect (Flynn, 1984). A meta-analysis, that is, a study that statistically combines the results of many studies, estimated it at about two to three standard-score points per decade (Trahan et al., 2014). The phenomenon, however, is neither universal nor constant in direction; a slowdown or even a reversal has been documented in some countries in recent decades (Bratsberg & Rogeberg, 2018). Where the effect exists, old norms inflate the scores of today's children, masking an existing disability at one end and swelling the number identified as gifted at the other. The practical lesson is the same in both cases: know the year in which the norms were collected, and treat scores near cut-off points with caution when the norms are old.
3. **In what language and culture?** Translation alone is not enough. A question about "snow" or "Thanksgiving" is familiar to a child in one setting and strange to a child in another.

## From translation to standardization

To simplify the path an instrument travels from one culture to another, this book distinguishes three levels. The division is the author's own teaching simplification, built on what international guidelines require (see the next box):

**Table 2.7: Three levels in transferring an instrument from one culture to another (the author's simplification).**

| Level | What happens | What can be done with the scores |
|---|---|---|
| **Translation** | Items are rendered into Arabic literally or nearly so | The original norms cannot properly be used to judge the child |
| **Test adaptation** | Items and examples are modified to suit the culture, with back-translation (translating the Arabic version back into the original language and comparing it with the original), expert review, item analysis, and verification of **measurement invariance** across the two versions and of the absence of **differential item functioning** (see 2.6.1) | Better than translation, but the original norms remain unsuitable |
| **Local standardization** | The adapted version is administered to a representative sample of the local population, new norms are derived, and reliability and validity are studied in that population | Standard scores become meaningful for children in that population |

**For the specialist: What do the International Test Commission guidelines actually say?**

The second edition of the International Test Commission Guidelines for Translating and Adapting Tests consists of eighteen guidelines grouped into six categories. Unlike the transfer levels in our table, they follow the stages of the work (International Test Commission, 2017):

1. **Precondition:** obtaining permission from the copyright holder, and verifying before starting that the construct being measured is meaningful in the new culture.
2. **Test development:** translating and adapting the items with a team qualified in both languages and in the subject matter, and trying out the items on small samples.
3. **Confirmation:** gathering empirical evidence of the equivalence of the two versions: measurement invariance, examination of differential item functioning, reliability, and validity in the new population.
4. **Administration:** controlling administration instructions and conditions so as to reduce sources of bias in the new population.
5. **Score scales and interpretation:** scores are compared across the two populations only to the extent that the established level of equivalence allows.
6. **Documentation:** documenting every change made to the original, and providing guidance for those who use the new version.

Local standardization in our table thus corresponds roughly to the completion of the "confirmation" and "score scales" stages, together with the derivation of new norms. For a version that has not gone through the confirmation stage, a well-crafted translation is not enough.

**In the Saudi context: Where do the norms of instruments in use stand?**

The instruments used in the Kingdom vary widely in where they stand among the three levels. Among the published examples we found:

- **Wechsler Intelligence Scale for Children:** The Saudi standardization cited in the literature is that of the revised edition (**WISC-R**) for ages 6–16, carried out within the gifted identification and care project at King Abdulaziz City for Science and Technology (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024). We found no published Saudi standardization of the fourth or fifth edition at the time of writing. In light of the foregoing, this means that the norms of the locally standardized edition are old, while the norms of the current edition are not local.
- **Wechsler Preschool and Primary Scale of Intelligence, Third Edition:** A Saudi version administered to 788 children aged 4 years to 7 years 3 months in the Jeddah education region between 2018 and 2022, derived from the Jordanian version (Ejbara, 2024). The sample is large, but it comes from a single region.
- **Stanford-Binet Intelligence Scales, Fifth Edition:** Two Gulf versions are in use, compared in a study in the Northern Borders region (Mohammed, 2016 [Arabic]); we found no details of either norming sample in a published source.
- **Gilliam Autism Rating Scale, Third Edition:** An Arabic version studied with 178 participants aged 3 to 22 years from educational and clinical settings in the Kingdom, with a Cronbach's alpha of 0.97 (Alasmari et al., 2024). This is a study of the version's properties; national norms are a later step.

**A question for reflection:** Which of these examples reaches the level of "local standardization" in Table 2.7? And what does each one lack to be suitable for a student in your city?

**Common misconceptions**

- **"An accurately translated version carries the validity, reliability and norms of the original."** Validity, reliability and norms are properties of using scores in a particular population; they do not travel with the translation. New evidence from the new population is essential.
- **"As long as the reliability coefficient is high, the scale is valid."** Reliability is a precondition for validity, not evidence of it (see Figure 2.3). A scale can consistently measure something other than what we intend.
- **"Any large difference between two scores indicates a problem."** The difference may not be statistically significant, or it may be significant yet common in the norming sample (see 2.5.1).

⚠ **Before you trust an Arabic version of an instrument**

- Is the version licensed by the original publisher?
- Has a published adaptation study been conducted, or was it merely translated? Were measurement invariance and differential item functioning examined?
- Does it have local norms? When were they derived? From which regions was the sample drawn?
- Have its reliability coefficients and validity evidence in the local setting been published, and from which of the five sources does this evidence come?

If the answer to most of these questions is no, use the scale's results with great caution, confine them to description rather than diagnosis, and state this explicitly in the report.

## 2.8 Reading a realistic score table

Intelligence test reports usually present a table summarizing the composite scores. Table 2.8 presents hypothetical values for a ten-year-old student, modeled on the indexes of the Wechsler Intelligence Scale for Children, Fifth Edition:

**Table 2.8: Hypothetical scores for a ten-year-old student, modeled on the indexes of the Wechsler Intelligence Scale for Children, Fifth Edition.**

| Index | Standard score | Percentile rank | 95% confidence interval | Descriptive classification |
|---|---|---|---|---|
| **Verbal Comprehension** | 84 | 14 | 78–92 | Low average |
| **Visual Spatial** | 102 | 55 | 94–110 | Average |

| Index | Standard score | Percentile rank | 95% confidence interval | Descriptive classification |
|---|---|---|---|---|
| **Fluid Reasoning** | 98 | 45 | 91–105 | Average |
| **Working Memory** | 76 | 5 | 70–85 | Very low |
| **Processing Speed** | 80 | 9 | 73–91 | Low average |
| — | 86 | 18 | 81–92 | Low average |

The confidence intervals in the table are built around the estimated true score, as in the original scale, which is why they are not symmetrical around the score (see the "For the specialist" box in 2.5). We assumed the following reliability coefficients: 0.94 for Verbal Comprehension and Fluid Reasoning, 0.92 for Visual Spatial, 0.93 for Working Memory, 0.90 for Processing Speed, and 0.96 for the Full Scale IQ (**Full Scale IQ, FSIQ**). How, then, should a teacher read this table?

1. **Look first at the full scale score and the limits of what it tells you.** The full scale score is 86, with an interval of 81–92, in the low average range. It is the most reliable score in the table and the strongest predictor of achievement.
2. **Then examine the differences for significance and rarity before interpreting the pattern.** The gap between the highest and lowest index is 26 points. The questions that concern us are: Is it statistically significant? And is it rare in the norming sample? We saw in 2.5.1 that the difference between Verbal Comprehension and Visual Spatial is significant but not rare, and that the difference between Working Memory and Processing Speed is not significant at all. Only if the differences prove both significant and rare does it make sense to speak of a "pattern," at which point the full scale score alone no longer describes the student adequately.
3. **Note the strengths.** His reasoning and visual thinking are in the average range. This is important instructional information: new material can be

presented visually, and teachers can draw on his ability to make inferences.

4. **Note the weaknesses and what they might mean in the classroom.** Working memory and processing speed are low. He may lose track of long instructions, need more time to finish tasks, and appear inattentive when in fact he cannot keep up with the pace of the class. These are hypotheses awaiting verification through classroom observation.
5. **Read the confidence intervals.** The working memory interval runs from 70 to 85, so describing it as "very low" in the scale's classification is correct for the observed score, but the interval extends into the low average range, so the degree of weakness is not settled.
6. **Ask: What does this mean for diagnosis?** This table alone diagnoses nothing. It does, however, call for considering the possibility of learning disabilities or attention-deficit/hyperactivity disorder, and for gathering further evidence from achievement, observation and measures of attention.

**For the specialist: The scholarly debate over profile analysis**

School psychology has an ongoing debate about the value of **profile analysis**, that is, interpreting the pattern of strengths and weaknesses across indexes. On one side, critics argue that the indexes add little to the full scale score in predicting achievement, that the reliability of differences between them is much lower than that of the full scale score, and that patterns have not proven useful in selecting a particular instructional intervention (McGill et al., 2018). On the other side, its defenders argue that disciplined interpretation of the indexes, grounded in differences that are significant and rare and supported by evidence from outside the test, helps in understanding the student and in guiding subsequent assessment, particularly with twice-exceptional learners, whose uneven abilities the full scale score may conceal (Kaufman et al., 2016). The position this book takes lies between the two: begin with the full scale score and the limits of what it tells you, move to interpreting the pattern only after the differences have been examined statistically, and treat the pattern as a hypothesis requiring further evidence. Chapter 6 returns to this debate in discussing patterns of strengths and weaknesses models.

## 2.9 Ethics and professional responsibility

Those who assess children answer to two kinds of references that should not be confused. The first is a **technical reference**, chief among them the Standards for Educational and Psychological Testing, issued by three American associations (AERA et al., 2014). The Standards specify the validity, reliability/precision and fairness that tests and their use should meet, and devote their third chapter to fairness in testing. In international practice they are complemented by the International Test Commission's guidelines on test use, which set out the competencies needed by test users (International Test

Commission, 2001). The second is the **codes of ethics** that professional bodies issue for their members, among them the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2017), the National Association of School Psychologists' professional standards, which include its principles for professional ethics (National Association of School Psychologists, 2020), and the Council for Exceptional Children's ethics and professional practice standards (Council for Exceptional Children, 2015). Among the principles on which these references agree, several bear directly on the teacher's work:

- **Informed consent:** The family knows why the assessment is being conducted, what will be administered and who will see the results, and gives consent before it begins. In the Kingdom, the Special Education Organizational Guide requires the guardian's written consent after formal notification (Ministry of Education, 1436 AH [Arabic]), and the Regulations stipulate obtaining his permission “for the purpose of diagnosis” and explaining his child's condition to him “in understandable language” (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]).
- **Confidentiality:** A student's results are not discussed in the staff room or sent in messaging groups, and only those who need them to serve the student have access to them. In the Kingdom, the Personal Data Protection Law, issued by Royal Decree No. (M/19) in 1443 AH, applies, and it classifies health data as sensitive data.
- **Test security:** Test pages are not photographed, and students are not coached on test items, since this undermines the validity of their results and those of others.
- **Appropriate use:** A scale is not used for a purpose it was not designed for, and no important decision rests on a single score. This is what the Organizational Guide stipulates in the phrase “not relying solely on the results of a single scale.”
- **Equity:** The assessor takes into account the student's language, culture and other disabilities; a student with low vision is not judged on intelligence by

a visual test (see 2.6.1).

- **Boundaries of competence:** The assessor does not administer a tool on which he has not been trained, nor interpret the results of a student from a linguistic or cultural background he does not know without consulting someone who does. Cultural competence is not acquired by reading a manual, and one of its boundaries is that the assessor knows when he needs a qualified interpreter or a colleague from the same community.
- **Conflicts of interest:** When the person who diagnoses is the same person who provides the paid service he recommends, or when a teacher's evaluation of a student is tied to incentives or to the number of students admitted to the teacher's program, judgment is affected, even if unintentionally. This should be disclosed, and an independent opinion sought where possible.

Among these principles lies a tension that recurs in practice: **the family's right of access** versus **test security**. The family has the right to know everything related to their child's assessment and program, and to request re-diagnosis if they believe he has not been placed in the appropriate program (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). On the other hand, copying and handing over test items exposes the test to circulation and may violate the publisher's rights. The solution toward which professional codes lean is to explain the results to the family in detail and to show them their child's responses and how they were scored in the presence of the specialist, without handing over the test booklet or copying its items.

## Qualification levels: Who administers what?

Many publishers classify their instruments into qualification levels that determine who may purchase and use them. The details vary from one publisher to another, but the general idea is the same: a first level for tools that teachers and others can administer after reading the manual; a second level requiring training in measurement or a professional license; and a third level

for complex tools such as individual intelligence tests and diagnostic instruments, which requires graduate study in a specialized field or an advanced professional license. The Wechsler Intelligence Scale for Children is an example of this last level. What the Saudi documents we reviewed stipulate is summarized in the following table:

**Table 2.9: Who administers what in the Kingdom, according to the published documents we reviewed at the time of writing.**

| Role | What the documents stipulate | What we found no provision for |
|---|---|---|
| **General education teacher** | Member of the school's assessment team (Ministry of Education, 1436 AH [Arabic]) | The tools he is entitled to administer. His practical role is referral, observation and informal tools |
| **Special education teacher** | Member of the assessment team (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]; Ministry of Education, 1436 AH [Arabic]) | A list of the standardized scales he administers |
| **Psychologist in schools and institutes** | The Regulations, under the title “behavioral training teacher (psychologist),” assign him assessment and diagnosis using standardized intelligence scales, adaptive behavior scales and informal methods, as well as re-diagnosis when needed (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). The Guide requires that tools be administered and interpreted by “qualified specialists” and that they be “approved by the Ministry of Education” (Ministry of Education, 1436 AH [Arabic]) | A published list of approved scales, and the qualification criteria required for each scale |

| Role | What the documents stipulate | What we found no provision for |
|---|---|---|
| **Psychologist in the health sector** | Classified and licensed by the Saudi Commission for Health Specialties. In the edition of the classification manual we reviewed, a bachelor's degree holder is classified as an "assistant specialist" working under supervision, and is classified as a "psychology specialist" after no less than three years of clinical experience (Saudi Commission for Health Specialties, 2009 [Arabic]). The psychology licensing examination assesses, among other things, the application of knowledge in psychological assessment and diagnosis (Saudi Commission for Health Specialties, 2024 [Arabic]) | The current edition of the classification manual |
| **Physician** | Participates in the school's assessment team "if available" (Ministry of Education, 1436 AH [Arabic]) | A general provision in education regulations specifying the diagnoses reserved for physicians alone |

We found no published provision in the Kingdom corresponding to the qualification levels publishers apply to the purchase and administration of standardized psychological tests. Until such a provision is issued, the practical reference remains the publisher's own requirements, together with the principle of boundaries of competence described above.

**In the Saudi context: Two ethical dilemmas**

**The first dilemma: consent by silence.** The Organizational Guide states that if the guardian does not respond within a period not exceeding fourteen days from the date of written notification, "this shall be considered permission and consent" (Ministry of Education, 1436 AH [Arabic]). The provision protects the student's right not to have his assessment delayed because a notice was ignored. Yet the principle of informed consent assumes that the family has understood what it is consenting to. What should the team do if it learns that the father cannot read, or that the notice was sent home with the student and never arrived? The provision permits the team to proceed; ethics requires attempting to reach the family by other means, and documenting that attempt, before settling for silence.

**The second dilemma: the mother who asks for the items.** A mother asks for a copy of the booklet of the intelligence test administered to her son, because she wants to "train him on the questions he got wrong" before his reassessment next year. The mother has a right to access and understanding, and the test has a security that protects the validity of her own son's results next time. How would you word a reply that respects both rights?

**For the teacher: Back to Umm Majed**

The teacher can now answer Umm Majed's two questions along these lines: "The number 3 is not a passing grade. It means that Majed's performance on this test is higher than that of about 3% of his peers of the same age. In other words, he needs clear support. As for your question about disability, the test alone cannot answer it. His score is close to the score used in making the decision, and the band around it extends on both sides, so the team will gather information about his everyday skills at home and at school before reaching any judgment."

## Conclusion

What we have presented in this chapter rests on a single unifying idea: a score is an approximate estimate of the student's position among his peers. Standard scores of every kind are different languages for one position on the normal curve; a percentile rank is the proportion of peers the student outperformed; every score is surrounded by error whose extent the confidence interval shows; and the difference between two scores is less precise than either one. It follows that a decision in which the confidence interval crosses the cut-off score is a professional decision requiring other evidence; that a difference between two scores supports interpretation only if it is both significant and rare; and that an extreme score can be expected to regress toward the mean on retesting.

Precision alone is not enough. Whether a score suits its purpose is a question of validity, whose evidence is gathered from multiple sources, and whether it suits the student is a question of norms and fairness: norms are appropriate only if their sample resembles the student and is recent enough, and an accurate translation carries over neither the validity nor the norms of the original. Whenever an assessor uses a translated tool or norms from another

setting, one question remains that no calculation can settle: Does the norming sample resemble the student in front of me? The answer depends on the technical evidence available and on the limitations the report states openly.

## 2.10 Summary, key terms and activities

### Chapter summary

- Norm-referenced measures compare the student with peers and serve diagnosis; criterion-referenced measures compare the student with a mastery standard and serve instruction; and some tools, such as curriculum-based measurement, are interpreted both ways.
- All standard scores are different languages for one position on the normal curve, and they can be converted into one another via the z-score. Descriptive labels vary across tests and their editions.
- Two standard deviations below or above the mean mark off about 2% of people at each tail, and many diagnostic cut-off scores derive from them.
- A percentile rank is the proportion of peers the student outperformed, not the proportion of correct answers, and the distances between ranks are unequal.
- Every score is an estimate surrounded by error, and the confidence interval shows its extent. When the interval crosses the cut-off score, the decision becomes a professional one resting on other evidence.
- The difference between two scores is less precise than either one, and two questions must be asked of it: Is it statistically significant? And is it rare in the norming sample?
- Extreme scores regress toward the mean on retesting, and a test's precision weakens near its floor and ceiling.

- Validity is a single unified concept whose evidence is gathered from five sources, and fairness is a third foundation alongside it and reliability/precision. The predictive value of a result changes as prevalence changes.
- Norms are appropriate only if their sample resembles the student and is recent enough, and an accurate translation carries over neither the validity nor the norms of the original.
- Testing standards are a technical document, the codes of ethics of professional bodies are a separate reference, and local regulations add what is binding on the assessor in his own country.

## Glossary of terms

| | |
|---|---|
| **Norm sample** | The group to which a scale was administered to derive its norms, and against which each student's score is later compared. |
| **Raw score** | The number of correct answers or the total points before conversion; it has no meaning on its own. |
| **Standard score** | A converted score with a fixed mean and standard deviation, the best known having a mean of 100 and a standard deviation of 15. |
| **Percentile rank** | The percentage of the student's peers in the norming sample (of the same age or grade) who obtained a lower score; a percentile rank of 84 means the student's score is higher than those of 84% of them. It is not the percentage of correct answers. |

| | |
|---|---|
| **Standard deviation** | A measure of the spread of scores around their mean. |
| **Cut-off score** | The score that separates two categories in a given decision, such as referral versus no referral, or meeting versus not meeting a criterion. |
| **Reliability/precision** | The consistency of a scale's results across time, items or raters, and the precision of the score in general. |
| **Standard Error of Measurement, SEM** | An estimate of how much an individual's score fluctuates around his true score owing to imperfect reliability. |
| **Confidence interval** | A range of scores within which the true score is expected, with a specified probability, to fall. |
| **Standard error of the difference** | The amount of error expected in the difference between two scores, from which the smallest statistically significant difference is calculated. |
| **Base rate** | The proportion of the norming sample showing a particular difference or pattern, which indicates whether the difference is rare or common. |
| **Regression toward the mean** | The tendency of extreme scores to move closer to the mean on retesting. |
| **Floor and ceiling** | A test's capacity to discriminate among very low and very high levels of performance. |

| | |
|---|---|
| **Item Response Theory, IRT** | A measurement framework that describes each item by a curve linking ability to the probability of a correct response, in which measurement precision varies with ability level. |
| **Validity** | The degree to which evidence supports interpreting a scale's scores for a particular purpose; its evidence is gathered from five sources. |
| **Differential item functioning, DIF** | A difference between two groups of equal ability in the probability of answering an item correctly. |
| **Measurement invariance** | A scale's measuring the same construct in the same way in two groups, so that their scores can legitimately be compared. |
| **Flynn effect** | Change in average performance on intelligence tests across generations, which was upward for most of the twentieth century; where it occurs, old norms inflate scores. |
| **Norm-referenced test** | A test in which a student's score is interpreted by comparison with the scores of peers in the norming sample. |
| **Criterion-referenced test** | A test in which a student's score is interpreted by comparison with a predetermined level of performance or mastery. |
| **Age equivalent** | The age at which children's average raw score equals the student's raw score; it does not mean the student's skills resemble those of children of that age. |

| | |
|---|---|
| **Sensitivity** | The proportion of those who have the condition whom the tool correctly detects. |
| **Specificity** | The proportion of those who do not have the condition whom the tool correctly rules out. |
| **Positive predictive value** | The probability that the condition is actually present in someone with a positive result; it is affected by the prevalence of the condition in the group tested. |
| **Negative predictive value** | The probability that the condition is actually absent in someone with a negative result. |
| **Test adaptation** | Modifying a scale's items and examples to suit a new culture, together with scientific study of the result; it is broader than translation and narrower than local standardization. It differs from "accommodation" in the sense of supports provided to a student during testing. |

## Review questions

1. Classify each of the following results as norm-referenced or criterion-referenced, and justify your answer: (a) "Read 42 words correctly per minute; his end-of-term goal is 50." (b) "His percentile rank is 30 on a standardized reading test." (c) "Mastered 7 of 10 skills on an addition skills checklist." (Objective 1)
2. Convert a T-score of 35 into a standard score with a mean of 100 and a standard deviation of 15, and give its descriptive classification in Table 2.2. (Objective 2)

3. A student obtained a percentile rank of 98 on a mathematics test. What does this mean? What does it not mean? And why is it incorrect to say that a student who moved from rank 45 to 55 progressed as much as a student who moved from 85 to 95? (Objective 3)
4. Two attention scales: the first has a reliability of 0.70 and the second 0.92. Which would you prefer for a diagnostic decision, and why? Calculate the standard error of measurement for each if the standard deviation is 10. (Objective 4)
5. In Table 2.8, calculate the standard error of the difference between Fluid Reasoning (reliability 0.94) and Working Memory (reliability 0.93), then decide: Is the difference between them significant at the 0.05 level? And what question remains before you interpret it? (Objective 4)
6. A screening tool with 90% sensitivity and 90% specificity is administered to 400 children among whom the condition has a prevalence of 5%. Calculate the positive and negative predictive values, and explain to a family what a positive result means for their child. (Objective 5)
7. An Arabic version of a behavior rating scale is being used at a diagnostic center with its American norms. For each of the five sources of validity evidence, write one question you would put to whoever prepared the version, and state the potential risks for a Saudi student. (Objective 6)
8. A colleague posts in the teachers' messaging group a photo of the summary page of a student's diagnostic report, asking what one of the scores means. Which principles does this action touch on? And how would you respond to him? (Objective 7)

### Case for application

Noura is thirteen and in the first year of intermediate school. Her teacher referred her because of marked underachievement over several years. She was given an Arabic-adapted individual intelligence test whose norms were collected about eighteen years ago on a sample from a region other than her own. Her full scale score was 74, with a full scale reliability coefficient of 0.96.

Her Verbal Comprehension index was 86 (reliability 0.94) and her reasoning index 72 (reliability 0.90). An adaptive behavior form completed by the teacher alone indicates a deficit in one domain. One team member holds that “the score is below 75, so the student is eligible for the intellectual disability program.”

Calculate the confidence interval for the full-scale score and the standard error of the difference between the two indexes. Then reach a judgment: is the evidence sufficient for an eligibility decision? Support your judgment with at least three points you have learned in this chapter, and name the missing evidence you would request before deciding. (Objectives 4, 6, and 7)

## Learning activities

1. **A score ruler:** On a sheet of paper, draw a normal curve and place on it the scores of five hypothetical students expressed in different metrics: standard scores, T-scores, scaled scores, and percentile ranks. Then order them from lowest to highest. *Product:* one drawing with a conversion table. *Time:* about 30 minutes. (Objectives 2 and 3)
2. **Translating for the family:** Rewrite Table 2.8 as a five-sentence paragraph that a family with no training in measurement could follow, mentioning the confidence interval and what the table does not mean. If you wish to practice on a real report, do so only with the consent of the family and the school administration, after removing everything that could identify the student, under the supervision of a specialist, and without using the product of the exercise in any decision. *Product:* one paragraph. *Time:* about 20 minutes. (Objectives 3, 4, and 7)
3. **Examining a test manual:** Choose an instrument whose technical manual you can access, or a published study of an Arabic version of an instrument, and extract from it: a description of the norming sample, the year the norms were collected, the reliability coefficients, the validity evidence classified by the five sources, and any examination of differential item

functioning or measurement invariance. Then judge: is it suitable for a student in your city? *Product:* a one-page table. *Time:* about two hours. (Objective 6)

## Chapter references

Mohammed, H. A. A. (2016). [Differences between the two Gulf versions of the Stanford-Binet Intelligence Scales, Fifth Edition, in assessing and diagnosing intelligence and mental abilities among persons with disabilities in the Northern Borders Region] [In Arabic]. *Jazan University Journal of Human Sciences, 5*(Suppl.), 236–254. https://search.mandumah.com/Record/1103082

Personal Data Protection Law, Royal Decree No. (M/19) (9/2/1443 AH), as amended by Royal Decree No. (M/148) (5/9/1444 AH) [In Arabic].

Saudi Commission for Health Specialties. (2009). *[Professional classification manual for health practitioners]* [In Arabic] (5th ed.). https://cams.ksu.edu.sa/sites/cams.ksu.edu.sa/files/attach/Professional_Registration_Classification_Manual_vr.6.pdf

Saudi Commission for Health Specialties. (2024). *Saudi Psychology Licensure Examination (SPSLE): Applicant guide*. https://scfhs.org.sa/sites/default/files/2024-03/SPSLE%20Applicant%20Guide%202024.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alasmari, M., Alduais, A., Qasem, F., Almaghlouth, S., & AlAmri, L. (2024). Predicting language, cognition, and social skills in individuals with autism spectrum disorder using the Arabic validation of Gilliam Autism Rating Scale–Third Edition. *Applied Neuropsychology: Child*. Advance online publication. https://doi.org/10.1080/21622965.2024.2417812

Alotaibi, B. M., & Alotaibi, A. A. (2021). Exploratory and confirmatory factor analyses of the Arabic version of the Childhood Autism Rating Scale. *Research in Autism Spectrum Disorders, 86*, Article 101827. https://doi.org/10.1016/j.rasd.2021.101827

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*(4), 313–326.

Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences, 115*(26), 6674–6678.

Council for Exceptional Children. (2015). *What every special educator must know: Professional ethics and standards* (7th ed.). CEC.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Emam, M. M., Al-Sulaimani, H., Omara, E., & Al-Nabhany, R. (2020). Assessment of adaptive behaviour in children with intellectual disability in Oman: An examination of ABAS-3 factor structure and validation in the Arab context. *International Journal of Developmental Disabilities, 66*(4), 317–326. https://doi.org/10.1080/20473869.2019.1587939

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51.

International Test Commission. (2001). International guidelines for test use. *International Journal of Testing, 1*(2), 93–114.

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). https://www.intestcom.org

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Wiley.

Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted, 29*(4), 451–484.

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108–121.

National Association of School Psychologists. (2020). *The professional standards of the National Association of School Psychologists*. NASP.

Salvia, J., Ysseldyke, J. E., & Witmer, S. (2017). *Assessment in special and inclusive education* (13th ed.). Cengage Learning.

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). Jerome M. Sattler, Publisher.

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and systems of supports* (12th ed.). AAIDD.

Seif Eldin, A., Habib, D., Noufal, A., Farrag, S., Bazaid, K., Al-Sharbati, M., Badr, H., Moussa, S., Essali, A., & Gaddour, N. (2008). Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry, 20*(3), 281–289. https://doi.org/10.1080/09540260801990324

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Part Two

# Methods

How we gather evidence

Chapter 3: Formal standardized methods

Chapter 4: Informal methods

Chapter 5: Alternative methods

Chapter 6: Identification models

# 3 Chapter 3 Formal standardized methods

Intelligence tests, achievement tests, adaptive behavior scales and rating scales: these are the four families that make up the specialist's toolkit, and their numbers dominate most diagnostic reports. In this chapter we examine what each family measures, how it is administered on paper, digitally and remotely, the conditions under which its results hold or collapse, and, finally, the state of its Arabic versions.

**Scene from the assessment room**

Mona, a teacher, stepped into the school psychologist's room for the first time. Colorful kits lined the shelf; on the table sat an easel booklet propped open, a set of red-and-white blocks, a stopwatch, a response booklet crowded with boxes, and two tablets beside it. "All this to measure intelligence?" she asked. He smiled. "This is just one kit. That's an achievement record form, and this is an adaptive behavior form—you'll be filling one out about your student, Sami."

Then he added, "What matters, whatever is in the kit and whether I give the test on paper or on the tablet, is that I will say the same words to Sami, in the same order, and stop the clock at exactly the moment everyone who gave this test to thousands of children before him stopped it."

The psychologist's last remark holds the key to this chapter. A standardized instrument draws its value less from the difficulty of its questions or the polish of its materials than from the uniformity of its administration, for it is uniformity that makes the comparison between a student and the norm sample (also called the standardization sample) a fair one. The norm sample, as defined in Chapter 2, is the representative group to whom the test was given during its development and from whose performance its norms were derived.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Judge, in a given administration scenario, how a change affects the validity of the score: whether it is an accommodation that preserves the comparison with the norms, or a modification or departure from the directions that invalidates it.
2. Classify a given formal instrument into one of the four instrument families and name the diagnostic decision it informs.
3. Link the indexes of an intelligence test to the broad abilities they measure according to CHC theory.
4. Interpret the results of adaptive behavior scales and the T-scores of rating scales, including discrepancies among informants' ratings.
5. Choose the appropriate type of instrument for a described student and justify the choice in terms of the student's language, senses and motor abilities.
6. Evaluate the status of the Arabic version of a formal instrument by asking specific questions about its standardization, reliability, validity and licensing.
7. Decide whether digital or remote administration of an instrument is appropriate for a described student, and specify what should be documented about it in the report.

# Introduction

Scores from standardized instruments appear in nearly every diagnostic report, and many diagnostic and eligibility decisions rest on them. Yet an instrument's value lies in the uniformity of its administration, scoring and interpretation, so any departure from the directions can weaken or invalidate the comparison with the norms. Anyone who takes part in a diagnostic decision needs to know what each family of instruments measures and what it does not, how an accommodation, digital administration or remote administration changes the meaning of a score, and which questions ought to be put to any "Arabic version" before its results are trusted.

We open the chapter by setting out what makes an instrument standardized, including start and discontinue rules and what invalidates a score (Section 3.1). We then present intelligence tests in light of CHC theory—individual, group and nonverbal—together with developmental scales for young children (Section 3.2), followed by achievement tests (Section 3.3). Next we turn to adaptive behavior scales (Section 3.4), and to behavior rating scales, how to read their T-scores and why informants disagree on them (Section 3.5), then to testing accommodations (Section 3.6) and to digital assessment and teleassessment (Section 3.7). We close with how to choose an instrument and with the state of Arabic versions (Section 3.8).

## 3.1 What makes an instrument standardized?

A standardized instrument is one in which everything is administered in a uniform way: the exact wording of the directions, the materials, the order, the timing and the scoring procedure. This uniformity is what allows the student to be compared with the norm sample, because members of that sample took the test in exactly the same way (Salvia et al., 2017). This is why specialists train at length in administration, until they can deliver the directions verbatim while observing the student at the same time.

## Start and discontinue rules

Most individual tests contain items arranged from easiest to hardest, and not every item is given to every student. A fourteen-year-old is not asked the items designed for a six-year-old, and a six-year-old is not taken all the way to the items designed for adolescents. Three rules govern this:

- **Start point:** the item at which the examiner begins, according to the student's age.
- **Basal:** a set number of consecutive correct responses; once the student achieves it, he is assumed to have been able to pass all the easier items before it, which are credited without being administered. If the student errs on the first items from the start point, the examiner goes back to easier items until the basal is established.
- **Discontinue rule:** the examiner stops once the student misses a set number of consecutive items, on the assumption that he would have missed everything after them. The point at which administration stops is called the ceiling.

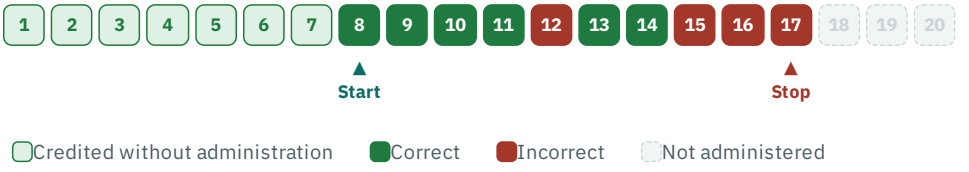



**Figure 3.1:** A hypothetical 20-item subtest with a basal of two consecutive passes and a discontinue rule of three consecutive errors. The student started at item 8 according to his age and passed items 8 and 9, so items 1 through 7 were credited. The examiner then stopped after three errors on items 15, 16 and 17. Raw score: 13. Actual rules vary from test to test and are specified in the manual.

## What invalidates a score?

Any departure from standardized administration makes the comparison with the norm sample less accurate, and some departures invalidate it altogether. Examples include:

- Rephrasing the directions for a student who did not understand them, on a test whose manual does not permit it.
- Allowing extra time on a timed test, since speed is part of what it measures.
- Giving the test to a student who answered its items only a few months earlier; familiarity with the items raises the score without any change in ability (practice effect).
- Translating items on the fly during administration.
- Giving hints, even through facial expressions, as the student nears the correct answer.

**Standardized administration is not rigidity**

Holding to the verbatim directions may seem harsh toward a flustered student. In fact, it is fair to him. Were the examiner to help, the student would obtain a score that looks higher than it really is, and might, because of it, be denied a service he needs. A good examiner reassures the student and encourages his effort without helping him answer.

## 3.2 Intelligence tests

Intelligence tests measure general mental abilities such as reasoning, language comprehension, visual thinking and memory. In diagnosis they serve several purposes: a principal criterion for intellectual disability, part of gifted identification, a basis for understanding the profile of strengths and weaknesses in learning disabilities, and a means of ruling out alternative explanations—showing, for example, that a student's low achievement is not due to low general mental ability.

### What exactly do they measure? CHC theory

Most modern intelligence tests rest on the theory of Cattell, Horn and Carroll (Cattell–Horn–Carroll), known by its authors' initials as **CHC**. The theory holds that intelligence consists of a general factor at the apex, broad abilities beneath it,

and narrower abilities beneath each of those (Schneider & McGrew, 2018). Knowing these abilities helps the teacher read a report, because a test's indexes measure some of them. An index is a composite score derived from two or more subtests that measure the same ability.

In their update of the theory, Schneider and McGrew revised some of what had become settled in older textbooks. The ability once called "long-term retrieval" (**Glr**) they divided into two: **learning efficiency** (**Gl**), the speed with which new information is stored and consolidated, and **retrieval fluency** (**Gr**), the speed with which stored information is recalled. A student may learn slowly yet retrieve fluently, or the reverse, and each pattern carries different implications for teaching. The following figure also includes two abilities close to school achievement, reading and writing and quantitative knowledge, because the next section presents the tests that measure them:

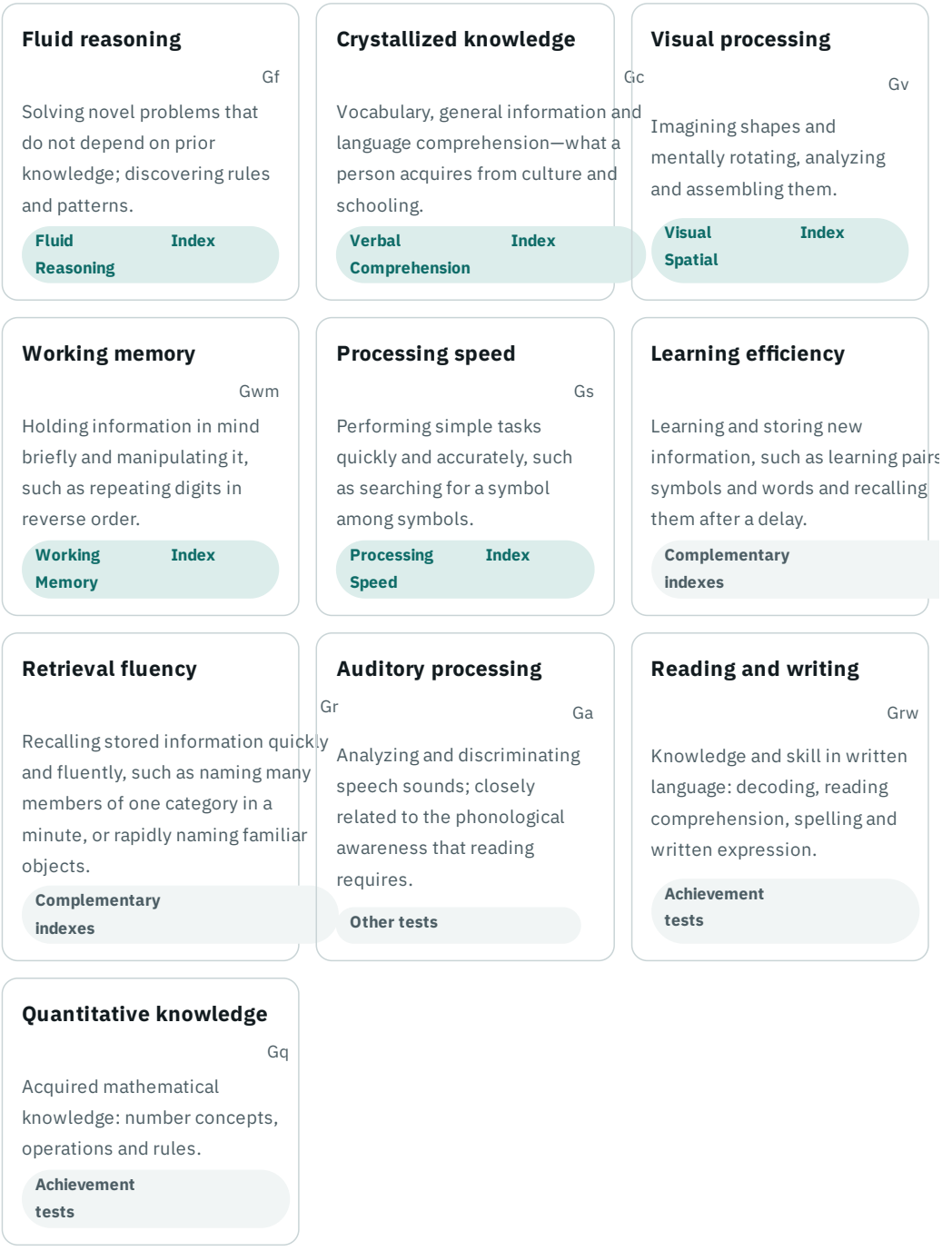



**Figure 3.2:** Ten broad abilities in **CHC** theory as updated by Schneider and McGrew (Schneider & McGrew, 2018), with the corresponding indexes of the fifth edition of the Wechsler scale for children (**Wechsler Intelligence Scale for Children, WISC-V**). The last five abilities are not measured by its primary

indexes: the examiner assesses learning efficiency and retrieval fluency with complementary indexes and turns to other tests for the remaining three.

The theory is not without its critics. Independent factor analyses of intelligence batteries generally find that the general factor accounts for most of the variance shared among the tests, and that several broad abilities are only weakly distinct from it. As a result, interpreting indexes individually adds little to the prediction of achievement, and differences among them are less reliable than the full-scale score (McGill et al., 2018). The cautious examiner therefore begins with the full-scale score and the limits of its meaning, and interprets a difference between two indexes only when it is both statistically significant and rare in the norm sample (Chapter 2 presents this debate).

## The best-known comprehensive individual intelligence tests

**Table 3.1: Comprehensive individual intelligence tests and the state of their Arabic versions.**

| Test | Age | Distinctive features | When preferred | Arabic version |
|---|---|---|---|---|
| **Wechsler Intelligence Scale for Children, Fifth Edition WISC-V (Wechsler, 2014)** | 6 years to 16 years 11 months | Five primary indexes from ten subtests, and a full-scale score from seven. Among the most widely used tests in schools worldwide | A comprehensive general assessment of the profile of strengths and weaknesses | At the time of writing we know of no published Saudi standardization of the fourth or fifth edition. The older revised edition (**WISC-R**) was standardized for ages 6–16 in Saudi Arabia as part of a gifted identification project (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024) |

| Test | Age | Distinctive features | When preferred | Arabic version |
|---|---|---|---|---|
| **Stanford–Binet Intelligence Scales, Fifth Edition SB5 (Roid, 2003)** | 2 to 85+ years | Five factors, each measured in a verbal and a nonverbal domain, and a wide range of items that discriminates among very high and very low abilities | Young children, and very high or very low abilities | Arabic versions exist, including two Gulf versions used in the Kingdom, one prepared by Safwat Farag and the other by Mahmoud Abu Al-Nil and colleagues; a study in the Northern Borders region compared the two (Mohammed, 2016 [Arabic]). We have not verified the size of their norm samples or the year in which they were collected |
| **Kaufman Assessment Battery for Children, Second Edition Normative Update KABC-II NU (Kaufman & Kaufman, 2018)** | 3 to 18 years | Can be interpreted within **CHC** theory or within the Luria model (**Luria**); offers a composite that excludes tests of acquired knowledge, and a nonverbal composite | Students from diverse linguistic and cultural backgrounds, when the aim is to reduce the influence of school knowledge | At the time of writing we know of no published standardized Arabic version |

| Test | Age | Distinctive features | When preferred | Arabic version |
|---|---|---|---|---|
| **Differential Ability Scales, Second Edition DAS-II (Elliott, 2007)** | 2 years 6 months to 17 years 11 months | A General Conceptual Ability score, a Special Nonverbal Composite, and many items at the lower levels | Young children and those with low abilities | At the time of writing we know of no published standardized Arabic version |
| **Cognitive Assessment System, Second Edition CAS2 (Naglieri et al., 2014)** | 5 to 18 years 11 months | Built on the four-process theory—planning, attention, simultaneous processing and successive processing—independent of acquired knowledge | When the question concerns processing more than knowledge | At the time of writing we know of no published standardized Arabic version |

## Nonverbal tests

These tests are administered through gestured directions or with minimal language, and the student responds by pointing or by arranging pieces. They are essential for deaf students, for those who have not mastered the language of the test, and for some children with language disorders or autism spectrum disorder.

**Table 3.2: Nonverbal ability tests and the state of their Arabic versions.**

| Test | Age | Distinctive features | When preferred | Arabic version |
|---|---|---|---|---|
| **Leiter International Performance Scale, Third Edition Leiter-3 (Roid et al., 2013)** | 3 to 75+ years | Fully nonverbal: directions given by gesture, responses by pointing or arranging pieces; measures reasoning, attention and memory | Deaf students, students with limited language, those who do not speak the language of the test, and some children with autism spectrum disorder | At the time of writing we know of no published standardized Arabic version |
| **Universal Nonverbal Intelligence Test, Second Edition UNIT2 (Bracken & McCallum, 2016)** | 5 to 21 years 11 months | Fully nonverbal; measures several abilities (memory, reasoning and quantitative reasoning) rather than a single one | A deaf student, or one who does not speak the language of the test, when a broader assessment than a matrices test is needed | At the time of writing we know of no published standardized Arabic version |
| **Test of Nonverbal Intelligence, Fourth Edition TONI-4 (Brown et al., 2010)** | 6 to 89 | Abstract figural problems that the student answers by pointing, in about 15 to 20 minutes | A quick estimate of nonverbal reasoning that paves the way for a fuller assessment | At the time of writing we know of no published standardized Arabic version |

| Test | Age | Distinctive features | When preferred | Arabic version |
|---|---|---|---|---|
| **Raven's Progressive Matrices** | A wide range, depending on the form | Figural matrices with a missing part; can be administered individually or in groups | Group screening and a quick estimate of fluid reasoning | At the time of writing we have not verified a published Arabic standardization of its recent forms |

**⚠ A nonverbal test is not a complete substitute**

Nonverbal tests are useful when language or hearing stands in the way of measuring ability. Most of them, however, measure a single facet: visual reasoning. They therefore do not describe the student comprehensively, nor are they adequate on their own for a high-stakes decision such as a diagnosis of intellectual disability or an identification of giftedness. The report should always state why such a test was chosen and the limits of what it measures.

In gifted identification specifically, Lohman (2005) showed that these tests predict academic achievement less well than tests of verbal and quantitative reasoning, that they do not erase group differences as is often assumed, and that they may screen out students who excel in verbal and quantitative reasoning. The alternative he proposes is to compare students with peers who have had similar opportunities to learn, within a battery that measures verbal, quantitative and nonverbal reasoning together.

**For the specialist: the Culture-Language Interpretive Matrix**

This interpretive matrix (**Culture-Language Interpretive Matrix, C-LIM**), developed by Flanagan, Ortiz and their colleagues, classifies subtests by their degree of cultural loading and linguistic demand (low, moderate and high). It then compares the student's performance with the range expected of students at the same level of proficiency in the language of the test. If the student's performance falls within that range and declines as cultural and linguistic loading increases, language and culture are likely to explain a large part of the low scores. Studies examining its validity, however, have not found that the pattern of graded decline distinguishes well between second-language learners and other referred students (Kranzler et al., 2010; Styck & Watkins, 2013). Moreover, its classifications were built on English-language tests and English learners, so transferring it to Arabic-speaking settings calls for research. It is best treated as a tool for organizing one's thinking and generating hypotheses (Chapter 14 presents it in greater detail).

## Developmental scales for young children

Before school age, intelligence tests in the familiar sense are used only within limits; greater reliance is placed on developmental scales that measure cognitive, language and motor skills, such as the Bayley Scales of Infant and Toddler Development, Fourth Edition (**Bayley-4**), for ages 16 days to 42 months (Bayley & Aylward, 2019). Results at this age are less stable and less predictive of later ability, so they are read as a description of the child's current developmental level and support decisions about early intervention more than they support a permanent diagnosis. At the time of writing we know of no published standardized Arabic version of the Bayley-4. From age four onward, a study has been published on the properties of a Saudi version of the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (**WPPSI-III**), administered to 788 children aged 4 years

to 7 years 3 months in the Jeddah education directorate between 2018 and 2022. This version was derived from a Jordanian version, and its scores correlated with those of the Saudi version of the Stanford–Binet at 0.81 (Ejbara, 2024).

## Group tests

Some ability tests are given to an entire class at once, with paper and pencil or on a computer. They are faster and less costly, but less precise for the individual, and the examiner cannot observe the student's behavior. They are therefore used for screening, particularly in gifted programs, after which students who exceed the cut-off score—the score set in advance for advancing to the next stage—are assessed with individual instruments. A student who does not reach the cut-off score, however, is not thereby shown to be not gifted: he may have been tired on the day of the test, or a weak reader who did not understand the written questions (Chapters 12 and 13 address this issue in detail).

**Table 3.3: Three group ability tests used in gifted screening.**

| Test | Target population | What it measures | Arabic version |
|---|---|---|---|
| **Cognitive Abilities Test CogAT (Lohman, 2011)** | Kindergarten through Grade 12 | Three batteries—verbal, quantitative and nonverbal reasoning—with the profile across them interpreted | At the time of writing we know of no published standardized Arabic version |
| **Naglieri Nonverbal Ability Test, Third Edition NNAT3 (Naglieri, 2016)** | Kindergarten through Grade 12 | Figural matrices only, so the influence of language is small; the limits of nonverbal tests noted above apply to it | At the time of writing we know of no published standardized Arabic version |

| Test | Target population | What it measures | Arabic version |
|---|---|---|---|
| **Mawhiba Multiple Cognitive Abilities Test** | Grade 3 through Grade 10 (first year of secondary school), at three levels | Mental flexibility, linguistic reasoning and reading comprehension, mathematical and spatial reasoning, and scientific and mechanical reasoning | Developed specifically for the Saudi context and administered by the National Center for Assessment (see the box in 3.8) |

## 3.3 Achievement tests

If intelligence tests ask, "How able is this student to learn?", achievement tests ask, "What has this student actually learned?" They are central to the diagnosis of learning disabilities, because one of the criteria for a learning disability is clearly low achievement in a specific area.

### Comprehensive individual tests

These tests measure several academic areas with precision and break each area into its component skills. If a student reads poorly, they pinpoint where the weakness lies: in decoding, in fluency or in comprehension.

**Table 3.4: Academic areas and their component skills in individual achievement tests.**

| Area | Component skills typically measured |
|---|---|
| **Reading** | Reading familiar words, decoding pseudowords, fluency and reading comprehension |
| **Mathematics** | Calculation, solving word problems and fluency with basic facts |
| **Writing** | Spelling, written expression and writing fluency |
| **Oral language** | Listening comprehension and oral expression |

Three of the best-known such tests internationally are listed below; all are built on the American curriculum and the English language:

**Table 3.5: Three comprehensive individual achievement tests and the state of their Arabic versions.**

| Test | Age | Distinctive features | Arabic version |
|---|---|---|---|
| **Wechsler Individual Achievement Test, Fourth Edition WIAT-4 (NCS Pearson, 2020)** | 4 years to 50 years 11 months | Linked to the Wechsler intelligence scales, which makes it easy to compare ability and achievement on a single sample | At the time of writing we know of no published standardized Arabic version |
| **Kaufman Test of Educational Achievement, Third Edition KTEA-3 (Kaufman & Kaufman, 2014)** | 4 to 25 years 11 months | Allows systematic analysis of error types in reading, spelling and arithmetic | At the time of writing we know of no published standardized Arabic version |
| **Woodcock–Johnson IV Tests of Achievement WJ IV ACH (Schrank et al., 2014)** | 2 to 90+ years | Built on **CHC** theory together with its companion cognitive battery. The publisher has announced a fifth edition (**WJ V**), so check for the latest edition before relying on it | At the time of writing we know of no published standardized Arabic version |

**In the Saudi context: measuring achievement in Arabic**

Achievement is the area most closely tied to language and curriculum. Arabic reading has features with no counterpart in English, such as short-vowel diacritics, letters whose shape changes with their position in the word, and diglossia between Modern Standard Arabic and colloquial dialects. A foreign achievement test therefore cannot be carried over by translation, and we know of no published Arabic standardization of any of the three tests in Table 3.5.

When no standardized Arabic test is available, the alternative is to combine instruments built on the local curriculum, such as curriculum-based measurement (discussed in Chapter 4), with results from national tests, such as the **Nafes** tests in the Kingdom—bearing in mind that these are group tests designed to evaluate education systems, not to diagnose individuals. Research on curriculum-based measurement in Arabic is still preliminary: a pilot study toward standardizing a measure of oral reading fluency with 200 students in the first three grades in Jordan (Abu-Hamour, 2014a), another on monitoring word recognition fluency in first grade (Abu-Hamour, 2014b), and a third on monitoring word reading fluency among 70 second graders in Al Ain, United Arab Emirates (Mahfouz & Mohamed, 2023). At the time of writing we found no published Saudi norms for such measurement.

**Question for reflection:** If no local norms are available, against whom would you compare a student you suspect is a poor reader? And what is lost if you compare him only with his classmates?

## 3.4 Adaptive behavior scales

Adaptive behavior comprises the skills people use in everyday life to look after themselves, interact with others and meet what their community expects of someone their age. A diagnosis of intellectual disability requires that low intelligence be accompanied by deficits in these skills. The American Association

on Intellectual and Developmental Disabilities (**American Association on Intellectual and Developmental Disabilities, AAIDD**) groups them into three domains (Schalock et al., 2021):

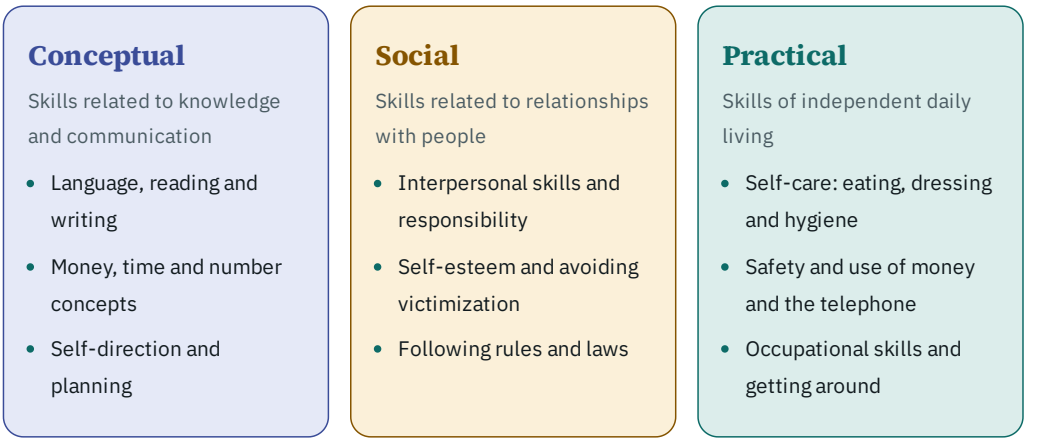



**Figure 3.3:** The three domains of adaptive behavior, with examples of the skills in each.

**What the student does, not what he can do**

Adaptive behavior scales measure what the student typically does in daily life. A child who knows how to tie his shoes but does so only when reminded and pressed is rated as not performing the skill independently. This distinction is what makes these scales a complement to intelligence tests, which measure the most a student can do under optimal conditions.

## Two principal scales

**Table 3.6: A comparison of the Vineland scales and the Adaptive Behavior Assessment System.**

| | Vineland Adaptive Behavior Scales, Third Edition Vineland-3 (Sparrow et al., 2016) | Adaptive Behavior Assessment System, Third Edition ABAS-3 (Harrison & Oakland, 2015) |
|---|---|---|
| **Age** | Birth to 90 years | Birth to 89 years |

| | Vineland Adaptive Behavior Scales, Third Edition Vineland-3 (Sparrow et al., 2016) | Adaptive Behavior Assessment System, Third Edition ABAS-3 (Harrison & Oakland, 2015) |
|---|---|---|
| **Forms** | Parent interview, parent-completed form and teacher form | Rating forms for parents, teachers and adult self-report |
| **Domains** | Communication, daily living skills, socialization, motor skills for young children, and an optional maladaptive behavior domain | Conceptual, social and practical, comprising eleven skill areas |
| **Overall score** | Adaptive Behavior Composite, with a mean of 100 and a standard deviation of 15 | General Adaptive Composite, with a mean of 100 and a standard deviation of 15 |
| **Distinctive features** | The semi-structured interview format allows the examiner to probe and ask for examples | Its domains map directly onto the AAIDD definition |
| **Arabic version** | We know of no published standardized Arabic version of the third edition whose data we were able to verify. Arabic catalogs list a work titled "Vineland Adaptive Behavior Scale: Saudi Norms," whose author, year and sample we were unable to verify | The Arabic teacher form was studied in the Sultanate of Oman with 410 children with intellectual disability; the results supported a single general-factor structure, and reliability was high (Emam et al., 2020). We know of no published standardized Saudi version |

### Scene from a Vineland interview

The specialist begins with Sami's mother not by reading her statements to answer yes or no, but with an open question: "Tell me about Sami's morning on a school day, from the time he wakes up until he leaves the house." Then he listens and asks about the details: "Does he choose his own clothes? Does he put them on without help? What about buttons?" From her answers he scores several items at once, then returns to the remaining items with further questions. The interview thus feels like a natural conversation, and it yields more accurate information than the form.

**For the teacher: the teacher form in adaptive behavior scales**

You will be asked about skills you see at school: Does the student follow two-step directions? Does he take care of his belongings? Does he ask for help when he needs it? You may also be asked about skills you do not see, such as his use of money. In that case, use the "no opportunity to observe" option if the form provides one, or write a note to the examiner—and do not guess.

**In the Saudi context: what do the regulations require of formal instruments?**

For a student to be eligible for intellectual disability programs, the Special Education Organizational Guide requires an IQ no higher than 75 on the Wechsler test or 69 on the Stanford–Binet (Fifth Edition), "or the equivalent of either," together with deficits in at least two adaptive skill areas, measured "according to formal or informal adaptive behavior tests" (Ministry of Education, 1436 AH [Arabic]). Its definition lists nine adaptive areas: communication, self-care, home living, social skills, use of community resources, self-direction, health and safety, functional academic skills, and leisure and work skills. These are closer to the skill areas of the Adaptive Behavior Assessment System than to the three domains the AAIDD uses today. The Guide further requires the administration of "instruments and scales approved by the Ministry of Education," "not relying solely on the results of a single scale," and administration and interpretation by "qualified specialists." We found no published list naming these approved instruments and their editions.

For giftedness, if the Ministry of Education's gifted classes are not filled by students classified through the National Program for Gifted Identification, the remaining places are filled by students who meet two or more criteria, one of which is a mental abilities test with a score "of no less than 120" (Ministry of Education, n.d. [Arabic]).

Two features of these requirements stand out. First, the upper limit for intellectual disability on the Wechsler (75) allows for measurement error around a score of 70, which is consistent with current diagnostic manuals. Second, the cut-offs on the two scales are not equivalent. The older Regulations set them at 75 on the Wechsler and 73 on the Stanford–Binet (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), which are equivalent if the latter is computed with a standard deviation of 16, as in earlier editions of

the Stanford–Binet. The fifth edition, however, has a standard deviation of 15, like the Wechsler, yet its cut-off became 69 rather than 75. Moreover, the text specifies neither the Wechsler edition nor the norms to be used.

**Question for reflection:** Two students each score 74, one on recent norms and the other on norms collected twenty years ago. Do they stand in the same position with respect to this requirement?

## 3.5 Behavior rating scales

On these scales, someone who knows the student—or the student himself—rates how often specific behaviors occur on a graded scale (for example: never, sometimes, often, always). They fall broadly into two types:

- **Broadband scales:** these cover many emotional and behavioral areas at once, such as anxiety, depression, hyperactivity, attention problems, aggression and social skills. They are useful for universal screening and for obtaining the overall picture before focusing on a particular problem. Among the best known are the Behavior Assessment System for Children, Third Edition (**Behavior Assessment System for Children, BASC-3**) (Reynolds & Kamphaus, 2015), and the Child Behavior Checklist from the Achenbach system (**Child Behavior Checklist, CBCL**) (Achenbach & Rescorla, 2001).
- **Narrowband scales:** these focus on a single disorder and measure it in depth, such as the Conners scales, fourth edition **Conners 4** (Conners, 2022), for attention-deficit/hyperactivity disorder (**Attention-Deficit/Hyperactivity Disorder, ADHD**), and the scales for autism spectrum disorder (presented in Chapter 9).

**Table 3.7: Common rating scales and the state of their Arabic versions.**

| Scale | Type and ages | Informants | Arabic version |
|---|---|---|---|
| **BASC-3** | Broadband; 2 years to 21 years 11 months on the parent and teacher forms | Parents, teachers and the student himself | At the time of writing we know of no published standardized Arabic version |
| **Achenbach System, school-age forms ASEBA: CBCL/TRF/YSR** | Broadband; 6 to 18 years | Parents (**CBCL**), teachers (**TRF**) and students aged 11 to 18 (**YSR**) | At the time of writing we know of no published standardized Arabic version |
| **Strengths and Difficulties Questionnaire SDQ** | Brief 25-item broadband screener; 2 to 17 years in age-specific versions, self-report version for ages 11–17 | Parents, teachers, and the student himself on the self-report version | An Arabic version was validated with Yemeni children aged 5 to 12 and discriminated well between a clinical sample (87) and a community sample (100) (Alyahri & Goodman, 2006). We know of no published Saudi norms |
| **Conners 4** | Narrowband: inattention, hyperactivity and associated problems; 6 to 18 years | Parents, teachers and the student | At the time of writing we found no published peer-reviewed Arabic standardization |

| Scale | Type and ages | Informants | Arabic version |
|---|---|---|---|
| **Vanderbilt ADHD Diagnostic Rating Scales** | Narrowband, with items screening for co-occurring problems | Parents and teachers | A Saudi Arabic version (**VADRS-A**) whose properties were studied with 1,812 informants (504 parents and 1,308 teachers) rating children aged 5 to 12 (Alqahtani et al., 2024) |

## How is a T-score read?

Most of these scales use the T-score, a standard score with a mean of 50 and a standard deviation of 10; a score of 60, for example, lies one standard deviation above the mean of the norm sample. In the **BASC-3**, the clinical scales, on which a higher score means more problems, and the adaptive scales, on which a lower score means weaker skills, are read as follows:

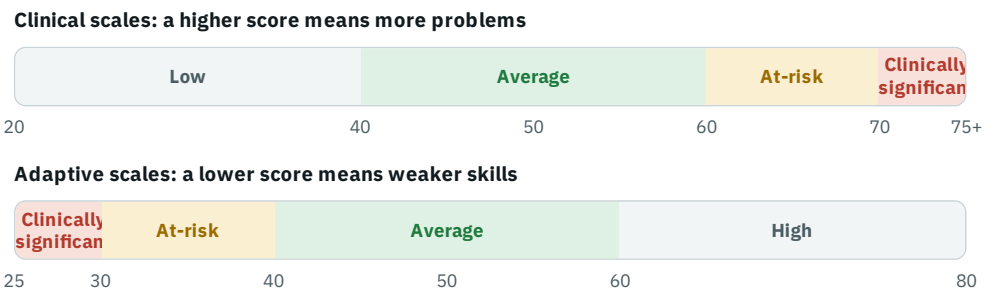



**Figure 3.4:** T-score interpretation ranges on the **BASC-3** scales. On the clinical scales, 60 to 69 is at-risk and 70 or above is clinically significant. On the adaptive scales the direction is reversed: 31 to 40 is at-risk and 30 or below is clinically significant. Lengths are approximate, to illustrate direction.

## Validity indexes within the scale

Some rating scales contain indexes that flag questionable response patterns. Some rise when the respondent chooses unusually negative responses, as if exaggerating the problem; others rise when the respondent gives contradictory answers to similar items. When these indexes are elevated, the scale's results are interpreted with caution, or not interpreted at all.

## When informants disagree

The person who provides information about the student—the father, the mother, the teacher or the student himself—is called the informant. One of the most important things for a teacher to know is that different informants' ratings of the same student vary considerably, and that this is normal. A well-known meta-analysis pooled the results of dozens of studies (Achenbach et al., 1987). The mean correlation between informants who see the child in the same setting, such as two parents, was about 0.60; between those who see the child in different settings, such as a parent and a teacher, about 0.28; and between the child's self-ratings and others' ratings, about 0.22. A more recent review confirmed that these discrepancies are systematic and that, alongside measurement error, they carry information about the settings in which the behavior appears (De Los Reyes et al., 2015).

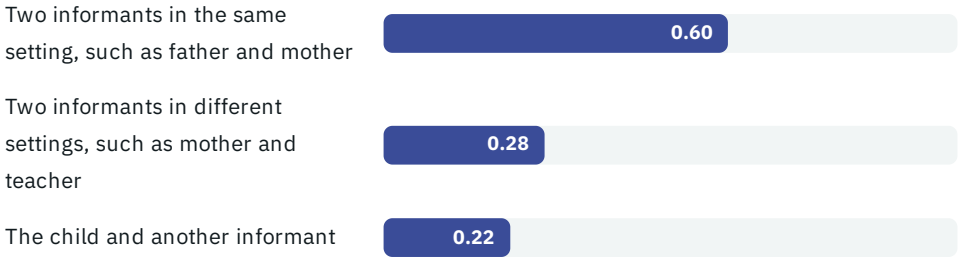



**Figure 3.5:** Mean correlations between different informants' ratings of a child's behavioral and emotional problems, from Achenbach and colleagues (Achenbach et al., 1987). The more the settings differ, the lower the agreement.

This does not mean that one of the informants is necessarily wrong. A child may genuinely behave differently at home and at school. Discrepant ratings are therefore information to be understood and interpreted. The discrepancy itself may even point to something important: a student whom the teacher rates as highly distractible and whose mother rates as attentive may turn out, on examination, to have a hearing loss whose effects show in a noisy classroom and disappear in the quiet of home.

## 3.6 Accommodations during testing

With students who have sensory, motor or language disabilities, the examiner faces a dilemma: administer the test as it stands, and it may measure the disability rather than the ability; change it, and the score may lose its connection to the norms. The way out lies in distinguishing between two kinds of change: accommodation and modification. Accommodation here should not be confused with adapting a test for a new culture (test adaptation), discussed in Chapter 2; that is carried out once, on the instrument as a whole, whereas accommodation is made for a particular student in a particular session.

**Table 3.8: Accommodation versus modification during testing.**

| | Accommodation | Modification |
|---|---|---|
| **Definition** | A change in presentation, response mode or setting that does not alter what the test measures | A change that affects what the test measures or its level of difficulty |
| **Examples** | Enlarged print for a student with low vision, a quiet room, breaks between subtests, pointing to an answer instead of saying it | Reading the items of a reading test aloud to the student, extra time on a speeded test, simplifying the directions on a test of language comprehension |
| **Effect on the score** | Comparison with the norms usually remains possible, if the manual permits it | Comparison with the norms is not valid; results are interpreted descriptively |

In either case, the examiner records in the report what was changed and why, so that readers know how to understand the score.

**Common misconceptions**

- **"Accommodation is cheating that gives the student an advantage over others."** A sound accommodation removes a barrier unrelated to what the test measures, bringing the score closer to the student's true ability. What does confer an advantage is modification, which is why a student tested with modifications is not compared with the norms.
- **"A nonverbal test is free of cultural influence."** Matrices, figures and working with timed materials are all experiences that vary with schooling and environment. A nonverbal test reduces the influence of language, but it does not remove the influence of culture and education.
- **"An accurate translation is enough to use a foreign test."** Translation carries over the items, not the norms. The score is still compared with children from another country who speak another language, and the difficulty of an item may itself shift in translation.
- **"A computerized test is inherently more accurate and objective."** The computer standardizes administration and scoring, but it does not guarantee validity. A digital format needs evidence of equivalence with the paper version, as the next section explains.

## 3.7 Digital assessment and teleassessment

Formal instruments are no longer only paper forms and test kits. Many are now administered on tablets or scored on online platforms, and some are administered remotely by video link. This raises one question in several guises: does the score still mean what it used to mean? The International Test Commission, together with the Association of Test Publishers, has issued guidelines for technology-based

assessment that address these questions, from platform design to data security and the documentation of results (International Test Commission & Association of Test Publishers, 2022).

## Digital administration and equivalence with paper

In the digital administration of the Wechsler Intelligence Scale for Children, for instance, the examiner holds a tablet that displays the directions, records responses, keeps time and applies start and discontinue rules automatically, while the student sees the stimuli on a second tablet and touches it to respond. This reduces examiner errors in calculation and timing, but it changes the medium the student works with. Touching a screen is not the same as pointing to a page, and tapping is not the same as writing with a pencil.

For this reason, norms for the paper format are not applied to the digital format until **equivalence** has been established. This is done through a study in which the test is given in both formats, either to two matched groups or to the same group in counterbalanced order, after which means, standard deviations, reliability and factor structure are compared. If the differences fall below a small threshold set in advance by the researchers, the same norms may be used; otherwise the digital format needs norms of its own. Speeded and motor tasks are the most sensitive to a change of medium, so they may remain on paper until the equivalence of their digital versions is demonstrated. Examiners should check the equivalence evidence for themselves in the technical manual of the edition they use.

## Computerized adaptive tests

In a computerized adaptive test, the computer selects each item on the basis of the student's previous responses: a harder item after a correct answer, an easier one after an error, stopping once the precision of the ability estimate reaches a preset level. The approach rests on item response theory (introduced in Chapter 2). Its advantages are that it is shorter than a fixed-form test and that its precision is fairly even across ability levels, which matters for the gifted student who reaches the ceiling of a fixed test and for the low-ability student who sits at its floor. Its prerequisite is a large item bank calibrated for difficulty on a representative

sample, something still rare among Arabic tests. And because each student answers different items, comparing students item by item makes no sense; comparison rests on the ability estimate alone.

## Curriculum-based measurement platforms

Digital platforms make it possible to administer brief measures of reading and math fluency (discussed in Chapter 4), score them and plot a student's progress automatically. This saves teachers' time and makes the data available to the team. But a polished graph does not settle two questions: are the passages the student reads truly equivalent in difficulty? And where do the norms the platform uses for comparison come from? Many platforms display norms collected in another country and in another language.

## Teleassessment: when is it appropriate, and when not?

In teleassessment, the examiner is in one place and the student in another, and the two communicate by video link. A study of children from a nonclinical sample found that scores on the Wechsler Intelligence Scale for Children were comparable across remote and in-person administration, under controlled conditions and with an assistant seated beside the child (Wright, 2020). That does not mean teleassessment suits every student or every test. The following table summarizes what cautious practice agrees on:

**Table 3.9: When teleassessment is suitable and when in-person administration is preferable.**

| Situation | Usual judgment | Reason |
|---|---|---|
| **An interview with parents, or a rating scale completed by a teacher or parent** | Usually suitable | Does not depend on test materials or strict timing, and semi-structured interviews work well by video |

| Situation | Usual judgment | Reason |
|---|---|---|
| **A verbal intelligence or achievement test for a school-age student, on a publisher-approved platform, with a trained assistant beside the student** | Possible under conditions | Equivalence evidence exists under controlled conditions, but it does not automatically extend to other conditions or other populations |
| **A young child below school age** | In-person preferred | Attention and behavior are hard to manage in front of a screen, and measures for this age rely on concrete materials and play |
| **Motor subtests or subtests that rely on concrete materials (blocks, puzzle pieces, pencil)** | In-person preferred, or the subtests are omitted and this is stated explicitly | Changing the materials changes the task, and the norms may no longer apply |
| **A student with a hearing or visual impairment** | In-person preferred | Audio and video quality add a new sensory barrier, so the test measures the connection rather than the ability |
| **A high-stakes eligibility decision based on a single test administered remotely** | Not sufficient | The result is confirmed through in-person testing or other multisource evidence |

## Test security, data privacy and consent

Digital assessment and teleassessment carry risks unknown to paper-based administration. Test items may be photographed or recorded, or heard and seen by family members, and thereby lose their usefulness for other students. Items from a paper test are therefore not shown over a camera unless the publisher permits it, and the family agrees in advance that nothing will be recorded and on who will be present in the session and in what role. As for data, the results of psychological and educational assessment are personal data, and health-related results count as sensitive data under the Kingdom's Personal Data Protection Law,

overseen by the Saudi Data and Artificial Intelligence Authority (Personal Data Protection Law, 1443 AH [Arabic]). Before any platform is used, then, three questions need answers: Where are the data stored? Who can access them? When are they deleted? Added to this is the requirement in education regulations for the guardian's written consent after formal notification (Ministry of Education, 1436 AH [Arabic]); the consent should state explicitly how the assessment will be administered and whether any part of the session will be recorded.

**For the teacher: What should you look for in a report when a test was given digitally or remotely?**

The report should state the mode of administration (paper, digital in person, or remote); the platform and its version; who was with the student and in what role; the quality of the connection and any interruptions; what departed from standardized administration and which subtests were omitted; the evidence that justifies using the norms; and an explicit statement of how cautiously the results should be interpreted. If all of this is missing, ask the examiner before building on the numbers.

## AI-supported screening tools

Tools are emerging that promise early detection of autism, dyslexia or attention problems by analyzing video clips, eye movements, voice, handwriting or school records. At best they are screening tools, to be followed by diagnosis on broader evidence, and they should face the same questions as any other measure, along with some of their own:

- **Whom were they trained on?** A model that learned from data on children in another country, or from a single group, may err systematically with others. Bias in the training data turns into bias in the decision.
- **How accurate are they with your students?** Reported sensitivity and specificity need to be verified in a sample resembling your students, and positive predictive value shifts with the local prevalence (as Chapter 2 shows).

- **Do they work in Arabic and its dialects?** A tool that analyzes speech or writing in another language does not carry over to Arabic without fresh validation.
- **Can their decisions be explained?** A tool that delivers a verdict without showing its basis is hard for the team to review or challenge.
- **Where do the data go?** Children's video clips and voices are highly sensitive personal data.

The same holds for identifying the gifted. A scoping review of artificial intelligence applications in gifted education found only three empirical studies using machine learning for identification, and it cautioned that models trained on past nomination and rating data may reproduce the very gaps they are meant to close. It therefore recommended treating their outputs as advisory evidence subject to mandatory human review, and verifying their accuracy separately for each group of students (Alsamani & Alsamiri, 2026).

A subtler danger remains: the inclination to believe what a machine says simply because a machine said it. When a tool's result contradicts what the team knows about the student from observation, interviews and work samples, what is needed is more evidence before the number is accepted.

## 3.8 How is an instrument chosen?

Before choosing any formal instrument, the examiner works through a list of questions. What decision is needed? Does the student's age fall within the instrument's range? Are its norms recent and representative? Does it have a standardized Arabic version, and what is that version's standing? Is its reliability adequate for this decision? Is there validity evidence for this purpose? Am I qualified to administer it? On the last question, test publishers classify their instruments by the qualifications required to purchase and administer them. We found no published text in the Kingdom specifying who may purchase or administer standardized psychological tests according to these levels. What is available is that education regulations assign the administration of intelligence and adaptive behavior measures to the school psychologist on a multidisciplinary

team (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), and require that instruments be administered and interpreted by "qualified specialists" (Ministry of Education, 1436 AH [Arabic]). The examiner then considers the characteristics of the student, which determine the kind of instrument that fits:




**Figure 3.6:** Student characteristics and assessment conditions that guide the choice of instrument type.

**Brief case: Turki, age 9**

Turki has been deaf since birth and communicates in sign language. His teacher referred him because he is far behind in reading. Had a verbal intelligence test been given to him through an interpreter, his score would have come out low, because the test would have measured his knowledge of spoken Arabic rather than his ability to reason. The psychologist chose a multi-ability nonverbal test with pantomimed directions, and his score fell in the average range. A sign language specialist then assessed his communication skills with him. In its report, the team noted that the test used had no Arabic norms, that his score should be read as an indicator of reasoning ability alone, and that his reading delay was more likely related to limited access to spoken Arabic than to his cognitive ability. It recommended reading instruction suited to deaf students.

> **What evidence could change the decision?** If the sign language specialist's assessment showed delays in Turki's communication skills themselves despite early exposure to sign language, or if his reading progress remained limited after a sufficient period of reading instruction suited to deaf students, that would call for considering a co-occurring language or learning difficulty that hearing loss alone does not explain.

## The state of Arabic versions: questions to ask before trusting a score

The tables in this chapter show that, for most instruments in wide international use, we know of no published standardized Arabic version, and that those that do exist vary greatly in how well they are documented. The phrase “Arabic version” in a report or an advertisement is not enough on its own. Before you trust a score from an Arabic version, ask:

**Six questions about any Arabic version**

1. **Who developed it, where and when?** The organization or researcher, the country, and the year the data were collected, not the year of printing.
2. **What is the norming sample?** Its size, its representation of regions, genders and school stages, and the ages it covers. Egyptian or Jordanian norms are not Saudi norms.
3. **Is it a translation, an adaptation or a standardization?** Translation alone leaves the norms foreign; standardization yields local norms.
4. **How reliable is it?** Reliability coefficients and the standard error of measurement in the Arabic sample itself.
5. **What is the evidence of its validity?** Did it retain the original factor structure? Was measurement invariance between the two versions examined? Does it correlate with local criteria?
6. **Is it licensed?** A version not licensed by the original publisher raises both a legal and a scientific issue, because no one can guarantee that it matches the original.

**In the Saudi context: instruments built in Arabic from the outset**

Most of what is used for diagnosis in Arabic consists of foreign instruments that have been translated or adapted. Some instruments, however, were built for the Arab setting from the start and so avoided the problems of translation, although they have limits of their own:

- **Mawhiba Multiple Cognitive Abilities Test (MMCAT).** This is the test underlying the National Program for Gifted Identification, launched in 2011 as a partnership among the King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba), the Ministry of Education and the National Center for Assessment. It measures four domains—mental flexibility, linguistic reasoning and reading comprehension, mathematical and spatial reasoning, and scientific and mechanical reasoning—at three levels spanning grade 3 through grade 10 (first year of secondary school), and a student may sit it once at each level. It is administered by the National Center for Assessment, part of the Education and Training Evaluation Commission, through optional registration with a fee paid by the guardian except for subsidized students (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]). We were unable to find the official score scale in published form; a peer-reviewed study of 5,146 students in Al-Ahsa reported an operational cut-off score of 1540 and found that the total score conceals divergent cognitive profiles that remain hidden when it is used alone (Almulhim et al., 2026). Chapter 12 discusses this test in detail.
- **Arabic curriculum-based measures.** These were built on Arabic passages and Arabic word lists rather than translations of foreign texts, but they are still at the stage of preliminary studies with limited samples (see the box in 3.3).

In the sources we were able to verify, we found no comprehensive individual intelligence test built in Arabic from the outset and standardized on a recent Saudi sample. This is a genuine research gap that deserves attention.

**Question for reflection:** What does a student gain from an instrument built in Arabic from the outset, and what might the student lose if its norms are not recent and its validity evidence is unpublished?

## Conclusion

Standardized instruments vary in what they measure: cognitive ability, achievement, what the student typically does in daily life, and what informants observe of the student's behavior. What unites them, as this chapter has shown, is that a score carries meaning only under the conditions of its administration. It loses its value when the directions are departed from, when a modification alters what the test measures, or when a digital format is given paper norms without evidence of equivalence. It follows that interpretation begins with the overall score; that a difference between two indexes is interpreted only if it is both significant and rare; that disagreement among informants is understood as information about the student and the student's settings; and that every accommodation or modification, and every remote administration, is recorded in the report.

Two matters remain open to debate: what can be inferred from the broad abilities on intelligence tests, given their limited distinctness from the general factor; and what can be expected of nonverbal tests, which help when language or hearing is a barrier but do not remove the influence of culture. Prior to both lies the open question of Arabic instruments. For most widely used instruments, no published standardized Arabic version is known, and the label “Arabic version” confers no confidence that has not been earned through evidence about its developer, sample, reliability, validity and licensing.

# 3.9 Summary, key terms and activities

## Chapter summary

- The value of a standardized instrument lies in its uniform administration; any departure from the directions weakens or invalidates comparison with the norms.
- Start and discontinue rules ensure that the student answers items suited to his or her level, and they differ from one test to another.
- Modern intelligence tests measure multiple abilities according to **CHC** theory, which now separates learning efficiency from retrieval fluency. But the broad abilities are only modestly distinct from the general factor, so interpretation begins with the overall score, and a difference between two indexes is interpreted only if it is both significant and rare.
- Nonverbal tests help when language or hearing is a barrier, but most measure a single aspect, predict achievement less well, and do not remove the influence of culture.
- Achievement tests are tied to language and curriculum and cannot be transferred through translation.
- Adaptive behavior scales measure what the student typically does, not what the student can do, and they are an essential criterion for intellectual disability.
- Discrepancies among informants' ratings are natural and expected; they are information to be understood and interpreted.
- An accommodation does not change what the test measures, whereas a modification does; both are recorded in the report.
- A digital format is not given paper norms without evidence of equivalence; teleassessment suits some instruments and students but not others, and it is documented in the report.
- For most widely used instruments, we know of no published standardized Arabic version; before an “Arabic version” is trusted,

questions should be asked about its developer, sample, reliability, validity and licensing.

## Glossary of terms

| | |
|---|---|
| **Standardized administration** | Administering an instrument with the same directions, materials, timing and scoring procedures used with the norming sample. |
| **Basal** | The number of consecutive correct responses after which the student is assumed to know the preceding items, which are then credited. |
| **Discontinue rule** | The number of consecutive errors after which the examiner stops administering the subtest. |
| **Ceiling** | The point at which administration stops once the discontinue rule is met; the student is assumed to fail the items beyond it. |
| **CHC theory** | A hierarchical model of cognitive abilities comprising a general factor, broad abilities and narrow abilities. |
| **Learning efficiency (Gl)** | A broad ability in **CHC** theory concerning the speed of storing and consolidating new information. |
| **Retrieval fluency (Gr)** | A broad ability in **CHC** theory concerning the speed of retrieving stored information. |
| **Adaptive behavior** | The conceptual, social and practical skills a person performs in daily life. |

| | |
|---|---|
| **Broadband scale** | A rating scale that covers many behavioral and emotional domains at once. |
| **Narrowband scale** | A rating scale that focuses on a single disorder or domain and measures it in depth. |
| **T-score** | A standard score with a mean of 50 and a standard deviation of 10, used in most rating scales. |
| **Informant** | The person who provides information about the student on a rating scale or in an interview, such as a parent, a teacher or the student. |
| **Cut-off score** | A predetermined score above which a student moves on to the next stage of assessment. |
| **Accommodation** | A change in presentation or response mode that does not alter what the test measures. |
| **Modification** | A change that affects what the test measures or its difficulty, so that comparison with the norms is no longer valid. |
| **Mode equivalence** | Established evidence that the digital format of a test yields scores with the same meaning as the paper format, so that the paper norms apply to it. |
| **Computerized adaptive test** | A test in which the computer selects each item on the basis of the student's previous responses and stops when the ability estimate reaches a specified precision. |
| **Teleassessment** | Administering an assessment instrument with the examiner in one place and the student in another, by video link. |

## Review questions

1. Why is strict adherence to the directions a matter of fairness to the student rather than harshness? (Objective 1)
2. In Figure 3.1, if the student had made an error on item 8 at the start, what would the examiner do? (Objective 1)
3. Assign each of the following instruments to its family, and name one diagnostic decision it contributes to: **ABAS-3**, **WIAT-4**, **Conners 4** and **UNIT2**. (Objective 2)
4. A student scored low on the Processing Speed Index and average on the other indexes. What might you notice about this student in class? And why is this difference alone not enough to conclude that a particular ability is impaired? (Objective 3)
5. Why is a low intelligence test score not enough to diagnose intellectual disability? What does an adaptive behavior scale add? (Objective 4)
6. On a hyperactivity rating scale, the teacher's rating yielded a T-score of 72 and the mother's a T-score of 55. How would you interpret this? What is the next step? (Objective 4)
7. Classify each of the following as an accommodation or a modification: enlarging pictures, reading word problems aloud on a math test, and extending the time on a subtest that measures processing speed. (Objective 1)
8. You read in a report that the student was given "the Arabic version of the Wechsler scale." What questions would you ask about this version before trusting the score? Which do you consider most important, and why? (Objective 6)
9. A family in a remote village asks for their five-year-old son to be assessed remotely because traveling to the diagnostic center is not feasible. What can be assessed remotely, and what cannot? (Objective 7)

### Case for application

**Razan, 13, first year of intermediate school (grade 7).** Razan has cerebral palsy that affects the movement of her hands; her speech is intelligible but slow, and she uses a computer with a large switch. A recent screening found her hearing and

vision intact. Her achievement in math and science is good on oral tests and weak on timed written tests. The school wants to assess her cognitive ability to decide whether she needs a special education program or only accommodations in the general classroom. The family lives in a governorate far from the diagnostic center and asks whether the assessment can be done remotely.

1. Which type of intelligence test would you choose for Razan, and which subtests would you avoid? Justify your choice. (Objective 5)
2. Propose two accommodations under which her score could still be compared with the norms, and one change that would make the comparison invalid. (Objective 1)
3. In her case, what could be gathered remotely, and what should be administered in person? And what should the report say about this? (Objective 7)

## Learning activities

1. **Instrument profile:** Choose an intelligence or achievement test whose record form and manual, or published technical manual, are available to you. Prepare a profile of it covering its underlying idea, format, who administers it, duration, types of scores and limitations. Then complete the "Arabic version" field using the six questions in 3.8, citing the source of each piece of information and writing "unknown" wherever you find no source. *Product:* a one-page profile. *Approximate time:* two hours.
2. **Simulating start and discontinue rules:** Design a subtest of 15 graded questions in a subject you teach, set start and discontinue rules for it, then administer it to an adult colleague and record what happened in the style of Figure 3.1. *Product:* the test, its rules and a diagram of what happened. *Approximate time:* an hour and a half.
3. **Adaptive behavior interview:** Write five open-ended questions to gather information about a child's everyday living skills, in the style of the Vineland interview. Then try them out with a parent, with that parent's consent, without recording any information that could identify the child, and for training purposes only, so that the answers are not used in any decision about the child.

If that is not possible, try them with a colleague playing the parent. *Product:* the questions, and your notes on what the open-ended questions added. *Approximate time:* one hour.

4. **Documentation paragraph:** Suppose an intelligence scale was administered remotely to an 11-year-old student, assisted by a teacher at the student's school; the connection dropped once, and a subtest relying on blocks was omitted. Write the paragraph you would add to the report to document this, drawing on the "For the teacher" box in 3.7. *Product:* a paragraph of 80 to 120 words. *Approximate time:* half an hour.

## Chapter references

Mohammed, H. A. A. (2016). [Differences between the two Gulf versions of the Stanford-Binet Intelligence Scales, Fifth Edition, in assessing and diagnosing intelligence and mental abilities among persons with disabilities in the Northern Borders Region] [In Arabic]. *Jazan University Journal of Human Sciences, 5*(Suppl.), 236–254. https://search.mandumah.com/Record/1103082

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Personal Data Protection Law, Royal Decree No. (M/19) (9/2/1443 AH), as amended by Royal Decree No. (M/148) (5/9/1444 AH) [In Arabic].

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/الموهوبين%20فصول%20دليل.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الخاصة%20التربية%20للتنظيمي%20الدليل.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Abu-Hamour, B. (2014a). A pilot study for standardizing curriculum-based measurement oral reading fluency (CBM ORF) in Arabic. *Journal of the International Association of Special Education, 15*(1), 16–26.

Abu-Hamour, B. (2014b). Using Arabic word identification fluency to monitor first-grade reading progress. *Dyslexia, 20*(2), 167–174. https://doi.org/10.1002/dys.1472

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Almulhim, N., Aljughaiman, A. M., & Alfaiz, F. S. (2026). Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis. *Journal of Intelligence, 14*(8), Article 166. https://doi.org/10.3390/jintelligence14080166

Alqahtani, M. M. J., Al Saud, N. M., Alsharef, N. M., Alsalhi, S. M., Al-Hifthy, E. H., AlHadi, A. N., Ad-Dab'bagh, Y., Alenazi, F. A., Alotaibi, B. M., Alsaeed, S. M., Arnout, B. A., ALQasem, L., Alhossein, A., Alqahtani, Y. J., AlGhamdi, S. A., Alrahili, N., Varnham, J., & Asiri, S. A. (2024). Psychometric properties of the Arabic Vanderbilt Children's ADHD Diagnostic Rating Scale (VADRS-A) in a Saudi population sample. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 12*(1), 72–83. https://doi.org/10.2478/sjcapp-2024-0008

Alsamani, O. A., & Alsamiri, Y. A. (2026). Artificial intelligence in gifted and talented education: A scoping review. *Frontiers in Psychology, 17*, Article 1898467. https://doi.org/10.3389/fpsyg.2026.1898467

Alyahri, A., & Goodman, R. (2006). Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment. *Eastern Mediterranean Health Journal, 12*(Suppl. 2), S138–S146.

Bayley, N., & Aylward, G. P. (2019). *Bayley Scales of Infant and Toddler Development* (4th ed.). Pearson.

Bracken, B. A., & McCallum, R. S. (2016). *Universal Nonverbal Intelligence Test* (2nd ed.). PRO-ED.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence* (4th ed.). PRO-ED.

Conners, C. K. (2022). *Conners 4th Edition*. Multi-Health Systems.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.). Harcourt Assessment.

Emam, M. M., Al-Sulaimani, H., Omara, E., & Al-Nabhany, R. (2020). Assessment of adaptive behaviour in children with intellectual disability in Oman: An examination of ABAS-3 factor structure and validation in the Arab context. *International Journal of Developmental Disabilities, 66*(4), 317–326. https://doi.org/10.1080/20473869.2019.1587939

Harrison, P. L., & Oakland, T. (2015). *Adaptive Behavior Assessment System* (3rd ed.). WPS.

International Test Commission, & Association of Test Publishers. (2022). *Guidelines for technology-based assessment*. International Test Commission.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement* (3rd ed.). NCS Pearson.

Kaufman, A. S., & Kaufman, N. L. (2018). *Kaufman Assessment Battery for Children, Second Edition, Normative Update*. NCS Pearson.

Kranzler, J. H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review, 39*(3), 431–446.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111–138.

Lohman, D. F. (2011). *Cognitive Abilities Test* (Form 7). Riverside.

Mahfouz, M. A. S., & Mohamed, A. H. H. (2023). Using word reading fluency curriculum-based measurements to monitor students' reading progress in grade 2. *Education Sciences, 13*(2), Article 217. https://doi.org/10.3390/educsci13020217

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108–121.

Naglieri, J. A. (2016). *Naglieri Nonverbal Ability Test* (3rd ed.). Pearson.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System* (2nd ed.). PRO-ED.

NCS Pearson. (2020). *Wechsler Individual Achievement Test* (4th ed.).

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children* (3rd ed.). Pearson.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Riverside.

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale* (3rd ed.). Stoelting.

Salvia, J., Ysseldyke, J. E., & Witmer, S. (2017). *Assessment in special and inclusive education* (13th ed.). Cengage Learning.

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and systems of supports* (12th ed.). AAIDD.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales* (3rd ed.). Pearson.

Styck, K. M., & Watkins, M. W. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition among referred students. *School Psychology Review, 42*(4), 367–382.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Wright, A. J. (2020). Equivalence of remote, digital administration and traditional, in-person administration of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V). *Psychological Assessment, 32*(9), 809–817. https://doi.org/10.1037/pas0000939

# 4 Informal methods

Chapter 4

A standardized test captures a snapshot of a student in a single session; systematic observation, weekly measurement, and the analysis of errors in everyday work trace the course of performance over time. In this chapter we examine these methods, which lie more firmly in the teacher's hands than any others, and we set out the evidence they yield and the safeguards that govern their use.

---

**Scene from a second-grade classroom**

Hessa, a teacher, is worried about her student Ghadeer. Every time she asks her to read aloud, Ghadeer stops at almost every word, reads “سافرَ” (sāfara, “he traveled”) for “سفَر” (safar, “travel”), and sometimes jumps to another line without noticing. Hessa wants to refer her for assessment, but the specialist asks her: “How many words does Ghadeer read per minute? Is she improving? What kinds of errors does she make?” Hessa has no answer beyond “Her reading is weak.”

The specialist's questions are the ones this chapter sets out to answer. A teacher who holds organized data on her student is in a position to judge whether a referral is needed, and to give the assessment team evidence that no test administered in a single session could supply.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Write an operational definition of a target behavior that another observer can record accurately.
2. Select the appropriate recording method for a given behavior and justify the choice.
3. Calculate percentage agreement between two observers and Cohen's kappa, and use both to judge the adequacy of an observation.
4. Draft semi-structured interview questions that elicit specific situations and the student's strengths.
5. Administer a curriculum-based measure, graph its data, and determine the instructional decision that the decision rules support.
6. Classify a student's errors in reading, writing, and arithmetic into patterns, and derive an instructional goal from each pattern.
7. Construct an analytic rubric for a writing task and plan how to check the consistency of its application.
8. Judge, in a new case, whether informal evidence is sufficient for a particular decision, whether instructional or diagnostic, and justify that judgment.

# Introduction

A standardized test captures a snapshot of a student in a single session and compares that performance with a norming sample. The teacher, by contrast, sees the student perform day after day on actual curriculum tasks, and is therefore able to gather repeated data that no one else can: structured observation of behavior, weekly measurement of fluency, and analysis of errors in written work. With such data, a general statement like “Her reading is weak,” as in Ghadeer's case, becomes evidence that reveals the student's level, the extent of improvement under instruction, and the faulty rule the

student is applying. Yet the closeness of these methods to the classroom usually comes at the cost of national norms, so it is necessary to establish which decisions they can support and which they cannot.

We begin by explaining the need for informal methods (Section 4.1). We then present structured observation step by step, from the operational definition through the choice of recording method to the check on interobserver agreement (Section 4.2), followed by the interview (Section 4.3). Next we take up curriculum-based measurement, the graphing of its data, and decision making in light of those data (Section 4.4); error analysis, with particular attention to oral reading errors (Section 4.5); and portfolios and rubrics (Section 4.6). We then consider the limits of these methods (Section 4.7) and close with forms ready for use (Section 4.8).

## 4.1 Why informal methods?

Informal methods are ways of gathering information grounded in what actually happens in the classroom, rather than in norms derived from a norming sample (see Chapter 2): observing behavior, interviewing people who know the student, measuring performance on tasks drawn from the student's curriculum, and analyzing the student's work. They are as important as formal methods, and they complement them in areas those methods cannot reach:

**Table 4.1: Formal standardized methods and informal methods.**

| Aspect | Formal standardized methods | Informal methods |
|---|---|---|
| **Core question** | Where does the student stand relative to peers? | What is the student doing here and now? What do we teach tomorrow? |
| **Frequency** | Once, or every few years | Daily or weekly |
| **Link to the curriculum** | General; may not match what the student is studying | Taken directly from what the student is studying |

| Aspect | Formal standardized methods | Informal methods |
|---|---|---|
| **Who administers** | Usually a qualified specialist | Usually the teacher |
| **Weakness** | A momentary snapshot; may not suit the student's language and culture | Usually no national norms; quality depends on the accuracy of the person administering |

## 4.2 Structured observation

Observation is part of every teacher's work, but observation that can serve as diagnostic evidence differs from everyday impressions in three ways: it begins with a clear definition of the behavior, it uses a specific recording method, and it can be repeated by another person who will arrive at the same result.

### Step one: the operational definition

An operational definition describes behavior in terms of what can be seen and heard, so that two people reading it would agree on whether the behavior occurred. Compare two definitions of the same behavior:

**Non-operational definition**

"Mishal is aggressive with his classmates."

What does "aggressive" mean? Shouting? Hitting? An angry look? Each observer will interpret it in his or her own way.

**Operational definition**

"Physical aggression: any contact between Mishal's hand or foot, or an object he is holding, and a classmate's body, with enough force to make a sound or to push the classmate from where he stands. Contact during organized play is not included."

A definition can be tested with a simple question: if someone who does not know Mishal read it, could that person record the behavior accurately?

## Step two: choosing a recording method

The recording method depends on the nature of the behavior. Does it have a clear beginning and end? Does it occur often? Is the number of occurrences what matters, or their duration? The best-known methods are the following (Cooper et al., 2020):

**Table 4.2: Observation recording methods and the behaviors each one suits.**

| Method | What is recorded | Suited to | Example |
|---|---|---|---|
| **Anecdotal record** | A free narrative of what happened, in sequence | The first stage, to grasp the overall picture | "In the third period, Mishal got up from his seat and went to the window…" |
| **Antecedent–behavior–consequence recording** | The antecedent event, the behavior, and the consequence that follows | Understanding the function of the behavior, that is, what it achieves for the student (discussed in Chapter 5) | Antecedent: he was asked to complete a worksheet. Behavior: he tore up the sheet. Consequence: he was sent out of the class |
| **Frequency recording** | The number of times the behavior occurs within a set period | Brief behaviors with a clear beginning and end | Number of times a student raises a hand to participate in a lesson |
| **Duration recording** | How long the behavior lasts | Extended behaviors | How long a child cries after arriving at kindergarten |

| Method | What is recorded | Suited to | Example |
|---|---|---|---|
| **Latency recording** | The time between giving an instruction and the start of the response | Delays in starting tasks | How long a student takes to begin writing after being asked to |
| **Interval recording** | Did the behavior occur in each short time interval? | Very high-frequency behaviors, or behaviors without a clear beginning and end | Time on task across a whole lesson |

Section 4.8 provides two blank forms for the methods most often used in schools: an interval recording form and an antecedent–behavior–consequence form.

## Interval recording: three methods, different results

In interval recording, the observation period is divided into short, equal intervals, say ten seconds each, and in every interval the observer records whether or not the behavior occurred. There are three ways of doing this, and they yield different results for the same behavior. Consider Figure 4.1, which shows one minute of observing a student who is out of his seat:

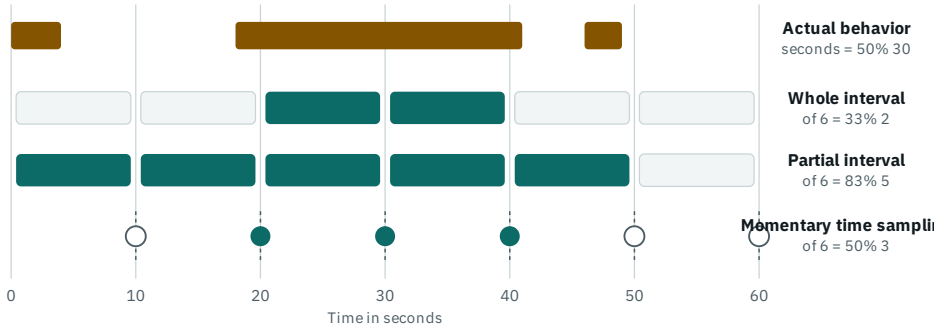



**Figure 4.1:** One minute divided into six intervals. The top bar shows when the student was actually out of his seat. In **whole-interval** recording, an interval is scored only if the behavior lasts throughout it, so the method tends to underestimate the behavior. In **partial-interval** recording, an interval is scored if the behavior occurs at any point within it, so the method tends to overestimate the behavior. In **momentary time sampling**, the observer looks only at the end of each interval; filled circles mark the moments when the student was found out of his seat, and open circles the moments when he was found seated. Its estimate comes closer to the true value as the observation lengthens and the number of intervals grows.

The practical rule: partial-interval recording suits behaviors we want to reduce, because it does not miss their occurrence. Whole-interval recording suits behaviors we want to increase, such as time on task, because it does not inflate them. Momentary time sampling is well suited to a teacher who is teaching and observing at the same time, since all it requires is to look up at the end of each interval, say every minute.

## Step three: checking interobserver agreement

How do we know that our observation is accurate? We ask a second observer to watch the student at the same time, independently, and then calculate the percentage of agreement between the two. The simplest way to calculate it in interval recording is:

$$\text{Percentage agreement} = \text{agreements} \div (\text{agreements} + \text{disagreements}) \times 100$$

A figure of 80% or higher is usually considered acceptable in research and practice (Cooper et al., 2020)

**Worked example**

1. A teacher and her assistant observed a student for 20 intervals; they agreed on 17 intervals and disagreed on 3.
2. Percentage agreement = 17 ÷ (17 + 3) × 100 = 85%.
3. On its face the figure is acceptable, so the definition is clear enough. Had it fallen below 80%, the two would have reviewed the operational definition together before continuing. This figure can be misleading, however, when the behavior is very rare or very frequent, as the next box shows.

**For the specialist: when percentage agreement misleads**

Suppose the twenty intervals in the previous example were distributed as in Table 4.3, and that the recorded behavior is rare, such as a student leaving his seat during a quiet lesson:

**Table 4.3: Distribution of the two observers' records across twenty intervals.**

| | Assistant: occurred | Assistant: did not occur | Total |
|---|---|---|---|
| **Teacher: occurred** | 1 | 2 | 3 |
| **Teacher: did not occur** | 1 | 16 | 17 |
| **Total** | 2 | 18 | 20 |

Percentage agreement is 85%, as we calculated, but most of that agreement concerns intervals in which nothing happened. The two observers agreed that the behavior occurred in only one of the four intervals in which either of them saw it, that is, just 25%. For this reason, it is recommended that agreement for rare behavior be calculated on occurrence intervals alone, and for very frequent behavior on nonoccurrence intervals alone (Cooper et al., 2020).

A more general solution is the kappa coefficient (**Cohen's kappa**), which removes from observed agreement the portion expected by chance (Cohen, 1960):

**Calculating kappa**

1. Observed agreement = (1 + 16) ÷ 20 = 0.85.
2. The teacher recorded occurrence in 15% of the intervals, and the assistant in 10%. Agreement expected by chance = (0.15 × 0.10) + (0.85 × 0.90) = 0.015 + 0.765 = 0.78.
3. Kappa = (0.85 − 0.78) ÷ (1 − 0.78) = 0.07 ÷ 0.22 ≈ 0.32.

Once chance is removed, the agreement that seemed acceptable falls into the band that the common classification labels merely "fair" (**fair**), the same classification that treats values between 0.61 and 0.80 as "substantial" agreement (Landis & Koch, 1977). These cutoffs are flexible conventions, but the lesson holds: before trusting a high percentage of agreement, check how often the behavior occurred in the first place.

## 4.3 The interview

The interview is a way of gathering information that cannot be seen in the classroom: developmental history, the student's behavior at home, what the family has already tried, and the student's own view. Interviews vary in how structured they are:

- **Structured interview:** specific questions in a fixed order and fixed wording, as in some standardized diagnostic interviews conducted with parents during assessment for autism spectrum disorder. It is precise and its results are comparable, but it is less flexible.
- **Semi-structured interview:** predetermined topics, with freedom in wording and follow-up. For example, the specialist might say to a mother, "Tell me about your child's morning on a school day," and then follow up with questions about the details, as in the Vineland Adaptive Behavior Scales interview.

- **Unstructured interview:** an open conversation that follows what the interviewee raises; it is useful for building rapport and grasping the overall picture.

**For the teacher: a short family interview guide**

1. What does your child enjoy, and what is he or she good at?
2. What concerns you about your child's learning or behavior? When did you first notice it?
3. Describe the last time it happened: What happened before it? What did your child do? How did the situation end?
4. Is it different at home than in other places?
5. What have you tried? What has worked, even a little?
6. Is there any health information, or any change in family life, that would help us understand your child?

Notice that the first and fifth questions look for strengths and successes alongside the problem. The third question asks for a specific situation, because situations yield more precise information than general descriptions.

Do not neglect to interview the student as well. A ten-year-old can tell you what is hard, what helps, and when he feels bored or anxious. He may tell you something no one else knows: “I can't see the board well from where I sit.” This interview carries even more weight with adolescents, who can describe what is going on inside them better than any other source, and who may reject a plan they were never consulted about.

## 4.4 Curriculum-based measurement

Curriculum-based measurement (**Curriculum-Based Measurement, CBM**) is a set of very brief measures, each taking from one to a few minutes, that the teacher administers to a student regularly, usually weekly, to monitor progress in basic skills. It was developed by Stanley Deno and his colleagues at the University of Minnesota in the 1970s to give teachers a quick and reliable way of knowing whether their instruction was working (Deno, 1985).

Its central idea is that certain brief tasks work like a thermometer. Just as body temperature indicates general health without diagnosing the illness, the number of words a student reads correctly in one minute is strongly related to reading skill as a whole and changes as that skill improves, even though it measures only one aspect of it.

This kind of measurement sits between the formal and the informal. Its content comes from the curriculum, but its procedures are standardized: fixed directions, set timing, and uniform scoring rules. Some of its English-language measures have extensive norms by grade and season of the school year. Its scores are therefore interpreted in two ways at once: normatively, by comparing the student with peers in the screening conducted at set points in the year, and against a growth goal set for the student and tracked week by week.

### The best-known measures

**Table 4.4: The best-known curriculum-based measures.**

| Area | Scale | How it is administered | What is scored |
|---|---|---|---|
| **Reading** | Oral reading fluency | The student reads a grade-level passage aloud for one minute; this is repeated with three passages | Number of words read correctly per minute, taking the median of the three passages |

| Area | Scale | How it is administered | What is scored |
|---|---|---|---|
| **Reading** | Maze | A passage in which every few words one word has been deleted, with three choices offered at each blank; two to three minutes | Number of correct choices; related to reading comprehension |
| **Mathematics** | Math computation | A sheet of mixed problems from the grade-level curriculum; two to four minutes | Number of correct digits in the answers, which is more precise than the number of problems |
| **Writing** | Written expression | The student is given the opening of a story and writes for three minutes | Number of words written, or number of correct word sequences in spelling and grammar |

## Scene from practice: reading fluency

Hessa sits down with Ghadeer. In front of Ghadeer is a copy of a passage she has never read; in front of Hessa is a second copy. “Read this story aloud as well as you can,” she says. “If you come to a word you don't know, I'll tell it to you. Begin.” She starts the stopwatch. As Ghadeer reads, Hessa puts a slash through every word she misreads or skips, and whenever Ghadeer pauses on a word for three seconds, Hessa supplies it and counts it as an error. After one minute she says, “Stop,” and marks the last word read. Then she does the arithmetic: Ghadeer read 46 words with 12 errors, so her score is 34 words correct per minute. Hessa records all of this on a scoring sheet like Form A in Section 4.8, together with the type of each error, because she will need it for the error analysis.

## Graphing data and making decisions

The strength of this kind of measurement lies in its repetition and its graphing. Each week the teacher plots the student's score on a graph and draws on it a line representing the goal, called the aimline, then compares the direction of the student's data points with the direction of the line. Several decision rules are in common use. One is the four-point rule: if four consecutive points fall below the aimline, the instructional approach is changed; if four consecutive points fall above it, the goal is raised (Hosp et al., 2016). Another is the trend-line rule, which compares the slope of the line through the student's points with the slope of the aimline: if it is shallower, the approach is changed; if it is steeper, the student is on track to reach the goal.

These rules, however, are approximate tools, and how far they can be trusted depends on the stakes of the decision built on them:

- **A low-stakes instructional decision**, such as modifying the teaching approach or increasing instructional time: an error here becomes apparent and can be corrected within weeks, so the point rule or the trend line serves as a practical guide. Administration manuals advise against reaching a judgment before several measurements have been collected after the baseline, because any single measurement fluctuates. This advice rests more on practitioner experience than on empirical research.
- **A high-stakes decision**, such as referral for an eligibility evaluation, a judgment that the student has “not responded” to intervention, or the discontinuation of a service: this requires more than a few weeks and a point rule. It calls for a longer monitoring period, passages that have actually been tried out and shown to be equivalent, administration that follows the same directions every time, and evidence from other sources as well, such as observation, error analysis, and a standardized test when needed.

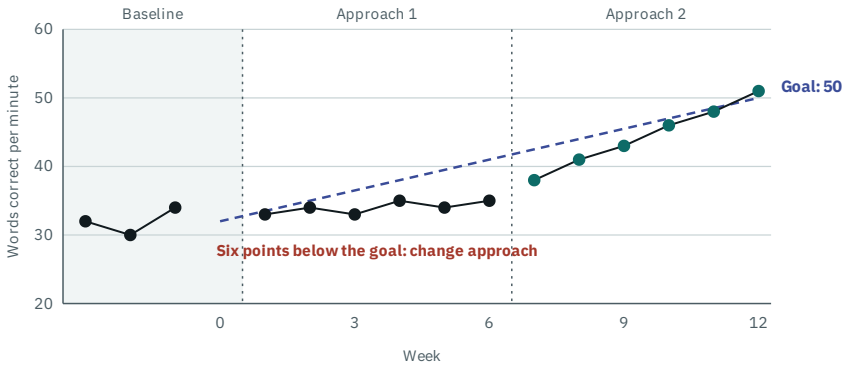



**Figure 4.2:** A hypothetical graph monitoring Ghadeer's reading fluency. The baseline, that is, the measurements taken before the new instruction begins, consists of three measurements with a median of 32 words; the goal was set at 50 words after 12 weeks, and the dashed aimline was drawn between the two. Under the first approach all the points stayed below the line, so the teacher waited until she had collected six measurements and then changed the approach in week 6. Under the second approach the slope of the points became steeper than that of the aimline, and they drew closer to it week by week until they reached the goal. This is what the trend-line rule measures. Changing the approach here is a low-stakes instructional decision; were an eligibility decision to be based on these data, a longer period and additional evidence would be required.

**For the specialist: what supports the decision rules?**

Ardoin and colleagues reviewed the literature on decision rules in curriculum-based measurement of oral reading: 171 documents, including articles, chapters, and administration manuals, of which 102 met the review's criteria. They focused on the recommended number of measurements before reaching a judgment and on guidelines for interpreting an individual student's data. They concluded that most decision-making practices rest on expert opinion and that their psychometric and empirical support is very limited (Ardoin et al., 2013). For this reason, we do not offer in this chapter a number of measurements that suits every case, since the research does not establish such a number. The grounds for caution are well understood in measurement logic:

- **Fluctuation:** measurement error in a single session is large compared with the growth expected in a week, so the fluctuation of one measurement can raise or lower a data point by as much as several weeks of growth.
- **Passage variability:** passages treated as equivalent may in fact differ in difficulty, so the slope shifts with their order, and what is really an effect of the passage is credited to the student's progress. In Arabic this problem is compounded by differences between passages in vowel marking and word length.
- **Unstable slopes over short periods:** a trend line calculated from only a few weeks can change considerably when a single point is added or removed.

The matter becomes more complicated still when these data are used to determine who has “not responded” to intervention. The common criteria for doing so, such as low level, slow slope, and the dual discrepancy that

combines the two, identify different groups of students, and agreement among them is weak (Barth et al., 2008). Chapter 6 presents these criteria, their thresholds, and the critiques of them in detail.

**What does the graph tell us about diagnosis?**

Had Ghadeer failed to improve even under the second approach, despite good, intensive instruction delivered as planned, that would have been important evidence justifying her referral for a comprehensive evaluation. Her improvement, by contrast, suggests that her weakness stemmed from instruction that did not suit her, which weakens the hypothesis of a learning disability. This idea is the foundation of the response to intervention model (**RTI**), which Chapter 6 presents in detail.

**Monitoring does not delay the family's right to an evaluation**

Collecting intervention data must never become a reason to postpone or prevent an evaluation the family requests. This was stated explicitly in an official memorandum from the Office of Special Education Programs at the U.S. Department of Education (U.S. Department of Education, 2011). We found no provision in published Saudi regulations addressing the relationship between tiered intervention and a family's request for an initial evaluation. The Regulations do, however, require that the family be invited to take part in the diagnosis, and they give the family the right to request a re-diagnosis if it believes its child has not been placed in the appropriate program (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 76). Informal data, then, are attached to the referral, where they strengthen and support it.

**In the Saudi context: where do we get the goal?**

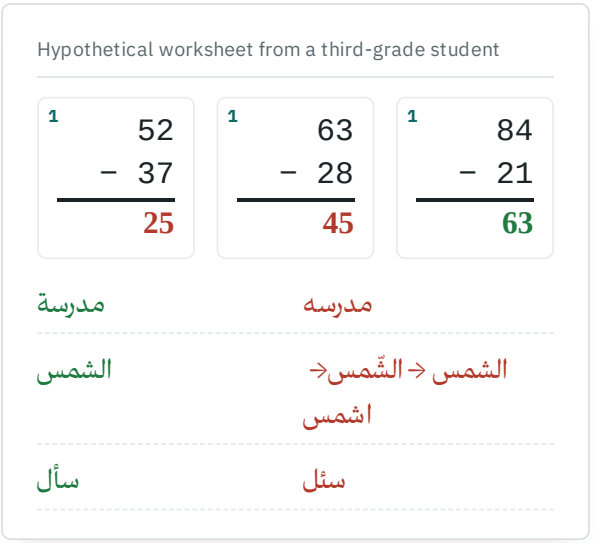
For reading in English, the United States has extensive reading fluency norms for every grade and season of the school year, from which a teacher can tell where a student stands and what goal is reasonable (Hasbrouck & Tindal, 2017). For Arabic, published studies are few and recent, and most come from outside the Kingdom. They include a pilot study to standardize an Arabic oral reading fluency measure, administered to 200 students in Grades 1 through 3 in Jordan (Abu-Hamour, 2014a); a second study by the same researcher that used Arabic word recognition fluency to monitor the progress of first graders (Abu-Hamour, 2014b); and a third that used word reading fluency measures to monitor the progress of 70 second graders in the city of Al Ain in the United Arab Emirates (Mahfouz & Mohamed, 2023). At the time this book was prepared, we found no published Saudi norms for curriculum-based measurement.

The practical alternative is for the school to build **local norms**: it administers the same measure to all students in a grade three times a year and calculates the percentile ranks of their scores, that is, the percentage of the student's classmates in the same grade who scored lower than the student did. A percentile rank of 84 thus means that the student outperformed 84% of them. In this way the school knows where each student stands among peers in that school. Passages should be vowelized or unvowelized according to what the student studies in that grade, because short-vowel marks change the difficulty of reading Arabic considerably. The team should also decide in advance whether errors in case endings on final letters count as errors, and then hold to that decision across all measurements; otherwise the scores are no longer comparable.

**A question for reflection:** If your school wanted to build local norms for second-grade reading fluency this year, who would administer the measure, where would the passages come from, and how would you verify that they are equivalent?

## 4.5 Error analysis

A score reveals how many items a student got wrong; error analysis reveals how and why. Errors are seldom random. They often betray a faulty rule that the student applies consistently, and once we discover that rule, we know what needs to be taught.




1. **A subtraction pattern:** in each place, the student subtracts the smaller digit from the larger regardless of its position. In the first problem he subtracted 2 from 7 and wrote 5, instead of regrouping and subtracting 7 from 12. He got the third problem right because it requires no regrouping. The error, then, lies in his understanding of place value; he knows his subtraction facts.
2. **Tā' marbūṭa and hā':** he writes what he hears when pausing at the end of the word (مدرسه for مدرسة). The remedy is to teach the morphological rule that distinguishes tā' marbūṭa from hā'.
3. **The "sun" lām:** he omits the lām that is not pronounced (اشمس for الشمس). This is a common phonological spelling error that indicates reliance on hearing when writing.
4. **The hamza:** he chooses the wrong seat for a medial hamza (سئل for سأل). This is among the most difficult rules of Arabic spelling and calls for systematic instruction.

**Figure 4.3:** Error analysis in arithmetic and spelling. Each error carries information about the rule the student is applying, and the remedy targets the rule that produced the answer.

The first pattern is known in the literature as the "smaller-from-larger" error, one of the best-known error patterns in arithmetic (Ashlock, 2010).

## Oral reading errors: from "meaning" to decoding

A widely used approach to analyzing oral reading errors is miscue analysis, which asks of each error: Did it preserve the meaning? Does it fit the structure of the sentence? Does it resemble the printed word in form? (Goodman, 1969). Today, however, this analysis calls for a critical reading. It rests on a view of reading as the parallel use of three "cues": meaning, structure, and the visual form of the word. Contemporary reading science has moved away from this view, having shown that skilled readers recognize words from their letters accurately and rapidly, and that reliance on context to recognize words is a mark of weak readers, who use it to compensate for poor decoding (Castles et al., 2018). The practical danger is that the question "Did it preserve the meaning?" can reward guessing, masking a decoding weakness rather than exposing it, which matters particularly in the diagnosis of dyslexia (taken up in Chapter 8).

**Misconception: "As long as the error didn't change the meaning, it's fine"**

A student read "The child went to the mosque" where the text said "The boy went to the mosque." By the logic of the "three cues," this is a good error, because it preserved meaning and structure. But the student never read the printed word at all: he guessed it from the sentence. If this pattern recurs, it is a sign that he cannot yet decode unfamiliar words and is covering for it with good comprehension. The right diagnostic question is therefore "Did he read what is written?" and not "Did he understand what is meant?" Comprehension is measured with other tasks.

In Arabic, it is more useful to analyze an error from two angles: decoding and morphological structure. Take Ghadeer's reading of "سافَرَ" (sāfara) for "سَفَر" (safar). The two words share a root (s-f-r), so the root letters are intact; the error lies in the pattern. She read a word on the faʿal pattern as though it were on the fāʿala pattern, adding a long vowel that is not in the text and turning the noun

into a verb familiar to her. The diagnostic questions here are: Does she read short vowels and long vowels as written? Does she substitute a familiar word from the same root for the printed word? Does this recur in both vowelized and unvowelized passages? If it does, the instructional goal is accurate reading of short and long vowels and recognition of common patterns, with priority given to reading the word over urging her to "understand the sentence." Table 4.5 summarizes a practical classification of these errors:

**Table 4.5: A classification of Arabic oral reading errors by decoding and structure.**

| Error type | Example (written → read) | What it may indicate |
|---|---|---|
| **Short-vowel error** | "كَتَبَ" → "كُتِبَ" (kataba → kutiba) | Weakness in reading short-vowel marks, or reading the word's outline without its vowels |
| **Adding or omitting a long vowel** | "سفَر" → "سافرَ" (safar → sāfara) | Inattention to long-vowel letters, and confusion between morphological patterns |
| **Substituting a letter for one that looks similar** | "بنت" → "بيت" (bint, "girl" → bayt, "house") | Weakness in discriminating diacritical dots and in analyzing the word into its letters |
| **Substituting a letter for one that sounds similar** | "صار" → "سار" (ṣāra → sāra) | Weakness in phonological discrimination between similar sounds |
| **Substituting one form for another from the same root** | "الطلاب" → "الطالب" (al-ṭullāb, "the students" → al-ṭālib, "the student") | Reading the root and guessing the structure, as in confusing the singular with the broken plural |

| Error type | Example (written → read) | What it may indicate |
|---|---|---|
| **Substituting a word that fits the context** | "الولد" → "الطفل" (al-walad, "the boy" → al-ṭifl, "the child") | Guessing from context to compensate for weak decoding; not counted as a success even if the meaning is preserved |

A judgment requires more than a single error. A pattern emerges when the same type of error recurs across more than one passage and more than one session, which is why Form A in Section 4.8 includes columns for tallying error types in each fluency measurement.

## 4.6 Portfolios and rubrics

### The portfolio

A portfolio is a selected collection of a student's work, gathered over a period of time for a specific purpose, in whose selection and reflection the student takes part (Paulson et al., 1991). It differs from a mere pile of papers in that it is built around a purpose and has criteria that determine what goes into it and how it is evaluated. There are three common types:

- **Working portfolio:** contains work in progress and drafts, and shows how the student is advancing.
- **Showcase portfolio:** contains the student's best work, and shows the most the student can do.
- **Assessment portfolio:** contains specified pieces of work matched against specified criteria, and is used in decision making.

In diagnosis, the portfolio is especially useful in two areas: monitoring the progress of a student with difficulties over time, and bringing to light abilities of a gifted student that tests do not reveal, such as a science project, a story the

student wrote, or a design the student created. It is therefore an important tool in identifying gifted students, and twice-exceptional learners in particular, that is, students who combine giftedness with a disability or disorder, each of which may mask the other (Chapters 12 and 13 address this in detail).

## Rubrics

To make the evaluation of student work more consistent, rubrics are used: they set out levels of performance and describe each level in words. A note on terminology is in order here. A rubric describes the quality of a piece of work the student has produced, whereas a rating scale asks a teacher or parent to rate the frequency or severity of a student's behaviors, as with the behavior scales presented in Chapter 3. An analytic rubric rates each aspect of the work separately, while a holistic rubric gives a single rating for the work as a whole. Here is an example of a brief analytic rubric for written expression:

**Table 4.6: A brief analytic rubric for written expression.**

| Criterion | 4 Advanced | 3 Proficient | 2 Developing | 1 Beginning |
|---|---|---|---|---|
| **Ideas and organization** | A clear idea, with a connected beginning, middle, and end | A clear idea, with minor weaknesses in coherence | An incomplete idea and weak organization | Disconnected sentences with no clear idea |
| **Language and vocabulary** | Varied and precise vocabulary | Appropriate vocabulary with some repetition | Limited and repetitive vocabulary | Very little vocabulary, or colloquial vocabulary |
| **Spelling and punctuation** | Rare errors that do not hinder comprehension | Few errors | Frequent errors that sometimes hinder comprehension | Many errors that hinder comprehension |

For a rubric to be dependable, it is tried out on a number of work samples that two teachers rate independently. Their agreement is calculated in the way presented for observation (4.2), after which they discuss the points of disagreement and revise the wording of the levels until they agree.

## Limits of portfolios and performance assessment in high-stakes decisions

What serves instruction is not necessarily fit for a consequential decision. The clearest illustration is the experience of the U.S. state of Vermont, which in the early 1990s built a statewide assessment system on portfolios in writing and mathematics. The program's evaluation found that its effect on instruction was positive, but that the quality of the scores it produced, including agreement among raters, was not sufficient for using them to make judgments about students and schools (Koretz et al., 1994). The lesson is that a portfolio is strong evidence of what a student can do, and it can support an eligibility or admission decision once raters have been trained, tasks standardized, and interrater agreement calculated, and when it stands alongside other evidence.

The same holds for performance assessments used in identifying gifted students, such as the DISCOVER model (**DISCOVER**), in which trained assessors observe students as they solve problems in different domains. In a study of 248 students in Grades 3 through 5 at two private schools in Beirut, ratings in the spatial domain correlated highly with scores on Raven's Progressive Matrices, whereas ratings in the linguistic domain correlated only weakly with those scores. The researcher concluded that the model could be used in Lebanon, while noting the need for further research to support the findings (Sarouphim, 2009). The evidence from the Arab region is thus promising, but it is limited by a small sample and private schools, and it calls for local validation before the tool is relied on for admission decisions.

## 4.7 Limits of informal methods

For all their considerable value, these methods have limits that anyone who uses them should keep in mind:

- **They usually lack norms:** on their own, they cannot tell us whether a student's performance departs from that of peers to an unusual degree, unless local norms have been built for them.
- **Their quality is in the hands of the person administering them:** a vague definition, a reading passage harder than the others, or an unclear rubric will each render the results unreliable.
- **They may carry observer bias:** a teacher who expects a problem may see more of it than is there. Checking agreement between independent observers is among the most important ways of limiting this bias.
- **Their decision rules are approximate:** they are adequate for instructional decisions, but a high-stakes decision requires other evidence alongside them, as we saw in 4.4.

For this reason, these methods are used in the diagnostic process both before and after referral: before it, to document the problem and attempt to address it in the classroom; after it, to complete the picture provided by formal tests and to plan instruction. Rarely are they sufficient on their own for a diagnostic decision. Even models that base their decisions on these data in a systematic way, such as response to intervention, require an adequate monitoring period, equivalent passages, and evidence from other sources.

**In the Saudi context: where do informal methods fit in the regulations?**

The Regulations for Special Education Institutes and Programs stipulate that assessment and diagnosis use both formal and informal methods; that the results of a single measure not be relied on alone in determining a pupil's eligibility; that information be gathered from a variety of sources; and that, where necessary, the pupil be placed under observation for a full semester (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 80). The same Regulations cite the interview, observation, and checklists as examples of the informal tools used by the school psychologist (Articles 48 and 49). Among the conditions for admitting a pupil to learning disabilities programs is that "it be established that regular educational services are inappropriate or of limited effectiveness" (Article 7), which is precisely what curriculum-based measurement data document when they are collected regularly. The Special Education Organizational Guide retained these conditions and did not mention response to intervention by name (Ministry of Education, 1436 AH [Arabic]).

In practice, two recent studies have described how this model is applied in Saudi schools from the teachers' perspective. Eight female learning disabilities teachers reported barriers that included limited time, weak collaboration, a shortage of data and training, and the absence of formal adoption of the model (Altheiabi et al., 2025). Another study, with thirteen teachers, found that practices resembling the model were largely informal and lacked systematic progress monitoring, diagnostic tools, and organizational guidance (Almulla & Mahama, 2026). Both studies are qualitative with small samples; they describe experiences, and how widespread those experiences are has yet to be measured.

**A question for reflection:** How would you document "the limited effectiveness of regular educational services" for a student in your school in a way the assessment team would accept?

## 4.8 Forms for use

This section contains four blank forms prepared by the author, each designed to be printed on a separate page; readers may copy them for their own educational and professional use. Training in administration should precede use of any form. The forms are used to gather student data in accordance with the procedures adopted by the school, and the Special Education Organizational Guide requires the written consent of the parent or guardian, following formal notification, before assessment and diagnostic procedures begin (Ministry of Education, 1436 AH [Arabic]). In training activities, a code is written in place of the student's name, and no decision is based on the results.

- **Form A:** Oral reading fluency record sheet, with columns for classifying errors (4.4 and 4.5).
- **Form B:** Interval recording form, with a calculation of interobserver agreement (4.2).
- **Form C:** Antecedent–behavior–consequence recording form (4.2).
- **Form D:** Evidence-supported referral form, organized according to the **RIOT/ICEL** framework (Christ, 2008). This framework gathers evidence through four methods (review of records, interview, observation, and testing) across four domains (instruction, curriculum, environment, and the learner); Chapter 1 presents it in detail. Its value here is that it shows the referring teacher which cells are empty before the assessment team discovers them.

## Form A: Oral reading fluency record sheet

Curriculum-based measurement in reading

Student or code: _____ Grade: _____

Date: _____ Examiner: _____

Purpose: ☐ Screening ☐ Baseline ☐ Progress monitoring, week: _____

Passages: ☐ Vowelized ☐ Unvowelized ☐ Partially vowelized

**Standardized directions (read verbatim every time):** "Read this story aloud as well as you can. If you come to a word you don't know, I will tell it to you. Begin." Timing starts when the student says the first word; after 60 seconds the examiner says "Stop" and places a mark after the last word read.

- **Counted as an error:** substituting a word, omitting a word, and a word on which the student hesitates for 3 seconds and which the examiner then supplies.
- **Not counted as an error:** self-correction within 3 seconds, repeating a word, and inserting a word not in the text (a note is made).
- **Team decision made in advance:** Are case-ending vowel errors at the end of a word counted as errors? ☐ Yes ☐ No (applied consistently across all measurements).

**Table 4.7: Form A, recording three passages and classifying their errors.**

| Passage (number or title) | Words read | Errors | Words correct per minute | Number of errors by type (Table 4.5) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Short vowel | Long vowel | Letter similar in shape or sound | Form from the same root | Word that fits the context |
| 1. | | | | | | | | |
| 2. | | | | | | | | |
| 3. | | | | | | | | |
| **Median** | | | | **Total across the three passages:** | | | | |

Qualitative notes (expression, pauses, line tracking, skipped lines): _____

The median value is plotted on the progress-monitoring graph. No change in performance is inferred from a single measurement.

### Form B: Interval recording form

Ten minutes, six intervals per minute

Student or code: _____ Date: _____

Time: from _____ to _____ Class period or activity: _____

Observer: _____ ☐ First observer ☐ Second observer

Target behavior: _____

Operational definition (what is seen and heard, and what is excluded): _____

_____

Method: ☐ Partial interval ☐ Whole interval ☐ Momentary time sampling Interval length (seconds): _____

**Table 4.8: Form B, interval grid. Write ✓ if the behavior is scored under the chosen method, and – if it is not.**

| Minute | 1 | 2 | 3 | 4 | 5 | 6 | Behavior intervals |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |

Behavior intervals: _____ of _____ = _____ %

**Interobserver agreement (the two forms are compared interval by interval):**

Agreements: _____ Disagreements: _____

Percentage agreement: _____ % Agreement on occurrence intervals only: ____ %

Kappa (optional): _____

Partial interval for behavior to be reduced, whole interval for behavior to be increased, and momentary time sampling for someone who is teaching and observing at the same time (4.2).

### Form C: Antecedent–behavior–consequence recording form

Descriptive recording in natural settings

Student or code: _____ Observer: _____

Period: from _____ to _____

Target behavior and its operational definition: _____

**Table 4.9: Form C, event log. Record what actually happened, and hold back interpretation.**

| Date and time | Setting and activity, and who was present | Antecedent: What happened immediately before? | Behavior: What did the student do? (as seen and heard) | Consequence: What happened immediately after? | Notes |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

**After several days of recording:**

Most frequent antecedent: _____

Most frequent consequence: _____

Tentative hypothesis about the function of the behavior: _____

The data on this form are descriptive and correlational: they suggest a hypothesis about the function of the behavior that still awaits verification, and Chapter 5 takes up the methods for doing so.

### Form D: Evidence-supported referral form

Organized according to the RIOT/ICEL framework

Student: _____ Grade: _____

Age: _____ Referring teacher: _____

Date of referral: _____

Reason for referral in behavioral terms (what the student does or does not do, when, and how often):

_____

Student's strengths: _____

**What was tried before referral:**

1. Intervention: _____ Duration in weeks: _____ Sessions per week and length: _____

2. Intervention: _____ Duration in weeks: _____ Sessions per week and length: _____

Baseline (median): _____ Goal: _____

Number of measurements after baseline: _____ Student's slope compared with the slope of the aimline: _____

☐ Progress-monitoring graph attached

☐ Passages or worksheets are at a single level and have been tried out

☐ Intervention was implemented as planned (documentation available)

Most recent hearing screening, date and result: _____ Most recent vision screening, date and result: _____

Attendance and punctuality this year: _____

**Table 4.10: Form D, evidence matrix. In each cell, record the source of the evidence and its date; leave the cell blank if no evidence is available.**

| Domain ↓ / Method → | Review (R): records and previous work | Interview (I): teachers, family, and student | Observation (O): in class and other settings | Test (T): curriculum-based measurement or other |
|---|---|---|---|---|
| **Instruction (I)** | | | | |
| **Curriculum (C)** | | | | |

| Domain ↓ / Method → | Review (R): records and previous work | Interview (I): teachers, family, and student | Observation (O): in class and other settings | Test (T): curriculum-based measurement or other |
|---|---|---|---|---|
| **Environment (E)** | | | | |
| **Learner (L)** | | | | |

Date of written notification to parent or guardian: Date of written consent: _____

The Special Education Organizational Guide sets the period for the parent or guardian's response at no more than 14 days from the date of written notification (Ministry of Education, 1436 AH [Arabic]). The approved procedures in force at the time of use should be checked.

Referring teacher | Special education teacher | School principal

## Conclusion

In this chapter we have seen that informal methods are a source in their own right of a different kind of evidence, comparable in value to the standardized test: evidence that is closely tied to the curriculum, collected repeatedly, descriptive of what actually happens in the classroom, and of direct service to instruction. Their value, however, depends on the discipline of their procedures. Observation requires an operational definition, a recording method suited to the behavior, and interobserver agreement that chance has not inflated; the strength of curriculum-based measurement lies in graphing its data over time; and error analysis looks past the number of a student's errors for the rule the student is applying. In practice, these data are attached to a referral to strengthen it, and monitoring a student's response to intervention must not become a pretext for delaying an evaluation the family has requested.

The open question concerns how far these methods can be relied on in high-stakes decisions. Decision rules in curriculum-based measurement rest more on expert opinion than on research, and portfolios and rubrics reveal what

tests do not, especially among gifted students, yet before they are used on their own they need evidence of agreement among raters. These methods are at their strongest when combined with other evidence.

## 4.9 Summary, key terms, and activities

### Chapter summary

- Informal methods stay close to the classroom and the curriculum, are repeated, and serve instruction, but they usually lack national norms.
- Structured observation begins with an operational definition, selects a recording method suited to the behavior, and checks agreement between observers. When behavior is rare or very frequent, simple percentage agreement is inflated by chance, so the kappa coefficient or agreement on occurrence intervals is used instead.
- The three interval recording methods yield different results for the same behavior, and the one chosen should fit the purpose.
- Curriculum-based measurement is brief, repeated, and standardized in its procedures, and its strength lies in graphing the data. Its decision rules draw more on expert opinion than on research, so they are suited to instructional decisions; a high-stakes decision requires a longer period, equivalent passages, and other evidence.
- Error analysis reveals the faulty rule a student is applying. In reading, an error is analyzed for what it shows about decoding and morphological structure before considering whether it preserved meaning.
- Portfolios and rubrics show what tests do not, especially among gifted students; using them alone in a high-stakes decision first requires checking agreement among raters.

- Response to intervention must not delay an evaluation the family has requested, and informal data are attached to the referral to strengthen it.

## Glossary of terms

| | |
|---|---|
| **Operational definition** | A description of behavior in terms of what is seen and heard, such that two observers can agree on whether it occurred. |
| **Interval recording** | Dividing the observation time into short intervals and recording whether the behavior occurs in each. |
| **Momentary time sampling** | Recording whether the behavior is occurring only at the moment each interval ends. |
| **Interobserver agreement** | The percentage of agreement between independent observers recording the same behavior. |
| **Cohen's kappa** | An index of interobserver agreement that removes the agreement expected by chance; it is more accurate than simple percentage agreement when behavior is rare or very frequent. |
| **Curriculum-based measurement** | Brief, repeated measures with standardized procedures, drawn from the curriculum, that monitor progress in basic skills. |
| **Baseline** | The first measurements, collected before new instruction begins, against which later progress is |

| | |
|---|---|
| | compared. |
| **Aimline** | A line on the progress-monitoring graph connecting the starting level with the goal, against which the student's data points are compared. |
| **Error analysis** | Examining the patterns in a student's errors to uncover the faulty rules the student is applying. |
| **Miscue analysis** | Analysis of oral reading errors according to whether they preserve meaning, fit the sentence structure, and resemble the printed word. It rests on a model that contemporary reading science has moved away from. |
| **Portfolio** | A selected collection of a student's work, gathered over time according to a purpose and criteria. |
| **Rubric** | A tool that sets out the criteria for evaluating a completed piece of work and describes in words the levels of performance on each criterion. It differs from a rating scale, which estimates the frequency or intensity of a behavior. |

## Review questions

1. Write an operational definition of "inattention," then test it by asking: Could a stranger record it accurately? **(Objective 1)**
2. You want to monitor how long a student stays engaged with his task during a mathematics lesson. Which interval recording method would you choose, and why? **(Objective 2)**

3. Two observers agreed on 22 intervals and disagreed on 8. Calculate the percentage agreement and decide whether it is acceptable. Then suppose the first observer recorded the behavior in 15 intervals and the second in 13, and that they agreed on its occurrence in 10 of them: calculate the kappa coefficient and interpret it. **(Objective 3)**
4. Draft three questions for a semi-structured interview with the father of a middle school student who is late in handing in his homework, so that one question elicits a specific situation and another looks for strengths. **(Objective 4)**
5. In Figure 4.2, if the points stayed below the aimline after the approach was changed, what would you conclude? What would be the next step? **(Objective 5)**
6. Why might six weekly measurements be enough to adjust the teaching approach, but not enough to decide that a student has "not responded" to intervention? Give two measurement-related reasons. **(Objectives 5 and 8)**
7. A student wrote: 45 + 38 = 713. What faulty rule is he applying? What instructional goal would you derive from it? **(Objective 6)**
8. A student read "يلعبون" (yalʿabūn, "they play") for "لعبوا" (laʿibū, "they played") in more than one passage, and "البحر" (al-baḥr, "the sea") for "النهر" (al-nahr, "the river") once. Classify each error according to Table 4.5, and explain why the second error should not be regarded as "good" simply because it preserved the general meaning. **(Objective 6)**
9. Why is a portfolio important in identifying a gifted student who does not excel on tests? And what must be verified before it is used in an admission decision? **(Objective 8)**

**Brief answers to the calculation questions:** Question 3: Percentage agreement = 22 ÷ 30 × 100 ≈ 73%, which is below 80% and therefore not acceptable until the definition has been reviewed. The agreement expected by chance = (0.50 × 0.43) + (0.50 × 0.57) = approximately 0.50, so kappa = (0.73 − 0.50) ÷ (1 − 0.50) ≈ 0.47, that is, "moderate" agreement by the common classification.

Question 7: He adds each place value separately and writes the full sum without carrying to the next place (5 + 8 = 13, and 4 + 3 = 7), so the instructional goal is the concept of regrouping in place value.

## Case for application

### Bandar, 13 years old, first grade of middle school

**The story:** Bandar's mathematics teacher complains that he is "careless": he does not finish in-class exercises, and he makes many errors on quizzes, even though his homework is usually complete. Over eight weeks, and with his parent's consent, the teacher worked with the learning disabilities teacher to gather the data shown in Table 4.11.

**Table 4.11: Bandar's informal data.**

| Source | Result |
|---|---|
| **Beginning-of-year screening (math computation, two minutes)** | His median performance falls at approximately the 12th percentile rank among students in his grade at the school |
| **Baseline (three weeks)** | 14, 12, and 15 correct digits, with a median of 14. The goal was set at 26 after 12 weeks, that is, one digit per week |
| **Five weeks of small-group instruction, three sessions a week** | 14, 15, 13, 16, 15. The aimline points for these weeks were 15, 16, 17, 18, 19. The worksheets had not been tried out in advance to verify their equivalence |
| **Observation (momentary time sampling every 30 seconds)** | On task 45% of the time in three mathematics lessons, and 80% in one science lesson. Agreement with a second observer was 90% in a single lesson, and kappa was not calculated |

| Source | Result |
|---|---|
| **Analysis of 30 problems from his notebook** | Most errors occur in copying numbers from the board, such as writing 68 for 86, followed by errors with fractions. Multiplication in the printed textbook problems is usually correct |
| **Interview with Bandar** | He sits at the back of the class and says the numbers on the board “look alike” to him, and that he is too embarrassed to ask to move closer |
| **Interview with his mother** | He studies with his older brother and completes all of his homework. She does not recall when his vision was last checked |

**Task:** (1) What instructional decision does the four-point rule support at this point? And what makes you hesitant to rely on it alone? (2) Are these data sufficient to refer Bandar for a comprehensive evaluation for learning disabilities in mathematics? Justify your judgment. (3) Identify at least two missing pieces of evidence, and locate each in the matrix of Form D. **(Objectives 3, 5, and 8)**

## Learning activities

Every activity carried out with students is conducted with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a specialist; its results are not used in any decision concerning the student.

1. **Paired observation:** With a colleague, observe a student for ten minutes using partial-interval recording, each of you recording on a separate copy of Form B. Then calculate the percentage agreement, agreement on occurrence intervals, and the kappa coefficient, and discuss where you

disagreed. *Product:* the two forms, a 2 × 2 table, and a paragraph explaining the difference among the three indices. *Approximate time:* one hour. *Desk-based alternative:* observe a student in a published video intended for teacher training.

2. **Measuring fluency:** Choose three passages from the reading textbook for a given grade, administer them to three students using Form A, calculate the median for each student, and classify their errors according to Table 4.5. *Product:* the three forms and a paragraph on the most prominent error pattern. *Approximate time:* an hour and a half. *Desk-based alternative:* prepare six passages at a single grade level, compare them in number of words, word length, and vowelization, then propose which of them could serve as equivalent passages for progress monitoring and justify your choice.
3. **Analyzing a notebook:** Collect five worksheets from a single student in arithmetic or spelling, identify at least two error patterns in them, and propose an instructional activity for each pattern. *Product:* a table of the patterns, their examples, and the activity proposed for each. *Approximate time:* 45 minutes. *Desk-based alternative:* analyze the errors in Figure 4.3 and in Questions 7 and 8.
4. **Building a rubric:** Construct a four-level analytic rubric for a writing task from the curriculum of a grade of your choice. Then, with a colleague, use it to evaluate five texts written by students, with their names removed; calculate the percentage agreement on each criterion, and revise the wording of the levels on which you disagreed. *Product:* the rubric before and after revision, and the agreement table. *Approximate time:* an hour and a half.

# Chapter references

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Abu-Hamour, B. (2014a). A pilot study for standardizing curriculum-based measurement oral reading fluency (CBM ORF) in Arabic. *Journal of the International Association of Special Education, 15*(1), 16–26.

Abu-Hamour, B. (2014b). Using Arabic word identification fluency to monitor first-grade reading progress. *Dyslexia, 20*(2), 167–174. https://doi.org/10.1002/dys.1472

Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications, 13*(1), Article 974. https://doi.org/10.1057/s41599-026-07376-9

Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties, 30*(1), 75–92. https://doi.org/10.1080/19404158.2025.2473425

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18. https://doi.org/10.1016/j.jsp.2012.09.004

Ashlock, R. B. (2010). *Error patterns in computation: Using error patterns to help each student learn* (10th ed.). Allyn & Bacon.

Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences, 18*(3), 296–307.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Christ, T. J. (2008). Best practices in problem analysis. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 159–176). National Association of School Psychologists.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219–232.

Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly, 5*(1), 9–30.

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms* (Technical Report No. 1702). Behavioral Research and Teaching, University of Oregon.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement* (2nd ed.). Guilford.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13*(3), 5–16. https://doi.org/10.1111/j.1745-3992.1994.tb00443.x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Mahfouz, M. A. S., & Mohamed, A. H. H. (2023). Using word reading fluency curriculum-based measurements to monitor students' reading progress in grade 2. *Education Sciences, 13*(2), Article 217. https://doi.org/10.3390/educsci13020217

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60–63.

Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted Lebanese students: Is DISCOVER effective? *Journal for the Education of the Gifted, 33*(2), 275–295. https://doi.org/10.1177/016235320903300206

U.S. Department of Education, Office of Special Education Programs. (2011). *A response to intervention (RTI) process cannot be used to delay-deny an evaluation for eligibility under IDEA* (OSEP Memo 11-07).

# 5 Chapter 5 Alternative methods

Assessment sometimes reaches beyond the question “What does the student know now?” to more important ones: “How does the student learn when given help?”, “Why does the student behave this way?” and “What can the student produce in a real task?” Traditional tests fall short on these questions, and this chapter presents the methods developed to answer them.

---

**Three scenes from one school**

In first grade, Nizar, a student who recently moved from another country, scored low on an Arabic vocabulary test. Does he have a language disorder, or has he simply not had enough exposure to Arabic yet?

In fourth grade, a student throws his book on the floor in almost every math lesson. And every time he is punished, things get worse.

In sixth grade, Tala, a student with average test scores, single-handedly designed a model of a solar-powered irrigation system for the school garden.

Each of these scenes raises a question that no traditional test can answer, and the methods we take up in this chapter were developed to answer precisely these three questions. Yet, like all assessment methods, they are bound by limits that anyone basing a decision on them must keep in mind, and so we pause on those limits for each method.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Apply the test–teach–retest model to a specific task and record the student's response on a graduated prompting scale.
2. Interpret the results of a dynamic assessment for a student from a different linguistic background, distinguishing what points to a lack of opportunity from what calls for assessment of a possible disorder.
3. Formulate, from antecedent, behavior and consequence data, a hypothesis about the function of a behavior, and specify the evidence needed to confirm it.
4. Design an authentic performance task with clear scoring criteria, and critique the reliability of its ratings if it were used in a high-stakes decision.
5. Choose between play-based assessment and a standardized developmental test for a child under three, and justify the choice.
6. Propose sources of evidence for assessing a student with severe and multiple disabilities for whom standardized instruments are unsuitable.
7. Judge whether the evidence an alternative method provides is sufficient for a particular decision in a new case, and justify the judgment.

# Introduction

Traditional tests mostly answer a single question: what does the student know now, compared with peers? Yet many diagnostic decisions hinge on other questions. Is a low score a sign of a disorder, or the result of limited opportunities to learn? Why does a behavior persist when punishment has done nothing to stop it? What can a student produce in a real task that a short test does not reveal? And what does a young child, or a student with a severe disability, actually do in daily life when comparison with peers is not meaningful? The three scenes at the start of the chapter illustrate these questions, and each calls for an assessment method suited to it, with limits that must be understood before any decision is built on it.

We open the chapter by setting out the need for alternative methods and distinguishing them from "alternate assessment" in its regulatory sense (Section 5.1), then present dynamic assessment, with its models, its two formats and its link to response to intervention (Section 5.2). We next turn to functional behavioral assessment, the common functions of behavior and its steps (Section 5.3), then to authentic and performance assessment and their place in identifying gifted learners (Section 5.4), followed by play-based assessment and assessment in early childhood (Section 5.5). We close with the assessment of learners with severe and multiple disabilities (Section 5.6).

## 5.1 Why do we need alternative methods?

Traditional tests, formal and informal alike, usually measure the **product** of learning at a given moment: what does the student know now? This is useful, but it assumes that all students have had equal opportunities to learn, and that performance in the testing situation reflects performance in life. When that assumption does not hold, we need other questions:

**Table 5.1: Alternative methods and the question each one asks.**

| Method | The question it asks | Where it is most useful |
|---|---|---|
| **Dynamic assessment** | How does the student learn when we teach? And how much does the student benefit from help? | Students from different linguistic and cultural backgrounds; distinguishing lack of opportunity from disorder |
| **—** | What drives the student to this behavior? And what maintains it? | Behavior problems, and building behavior intervention plans |
| **Authentic assessment** | What can the student produce in a real, meaningful task? | Giftedness and creativity, and skills that do not show up on short tests |
| **Play-based assessment** | How does the young child behave in natural activity? | Early childhood, and children who do not respond to testing situations |
| **Assessment in routines and environments** | What does the student actually do, and with what help, in everyday situations? | Learners with severe and multiple disabilities for whom standardized instruments are unsuitable |

These methods have a basis in Saudi regulations. In the section on assessment and diagnosis, the Regulations for Special Education Institutes and Programs call for the use of "formal and informal methods" and for "not relying on the results of a single measure when determining a pupil's eligibility," and they permit placing a pupil under observation "for a full semester" when needed (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 80). Alternative methods are what fill this space with organized evidence, rather than leaving observation as a general impression.

⚠ **A note on terminology: "alternative methods" are not "alternate assessment"**

In this chapter we use "alternative methods" to mean methods that complement traditional testing by asking other questions. In the American literature, however, the term **alternate assessment** has a specific regulatory meaning: a statewide achievement test taken by students with the most significant cognitive disabilities, in which their performance is measured against alternate achievement standards rather than grade-level standards. The Every Student Succeeds Act caps the share of students assessed this way at 1% of all students tested in the state in a given subject (Every Student Succeeds Act, 2015). When readers come across the term in a report or study from abroad, it usually carries this narrow regulatory meaning. We discuss the assessment of this group of students in Section 5.6.

## 5.2 Dynamic assessment

**Dynamic assessment** rests on an idea from the Russian psychologist Lev Vygotsky known as the **zone of proximal development**: the distance between what a child can do alone and what the child can do with help from someone more experienced (Vygotsky, 1978). Two children may earn the same score when each works alone, yet one moves ahead quickly after a simple hint, while the other progresses only with considerable help. This difference does not appear on a traditional test, but it reveals a great deal about their capacity to learn.

This is why dynamic assessment departs from traditional assessment: the examiner gives up neutrality, deliberately intervening and teaching, and then measures the effect of that teaching. Reuven Feuerstein and his colleagues developed this idea into an instrument called the **Learning Potential Assessment Device** (LPAD), which relies on what Feuerstein called **mediated learning experience**: an adult stands as a mediator between the child and the

task, directing the child's attention, organizing information, and helping the child grasp the strategy and transfer it to other situations (Feuerstein et al., 1979).

## The best-known model: test, teach, retest

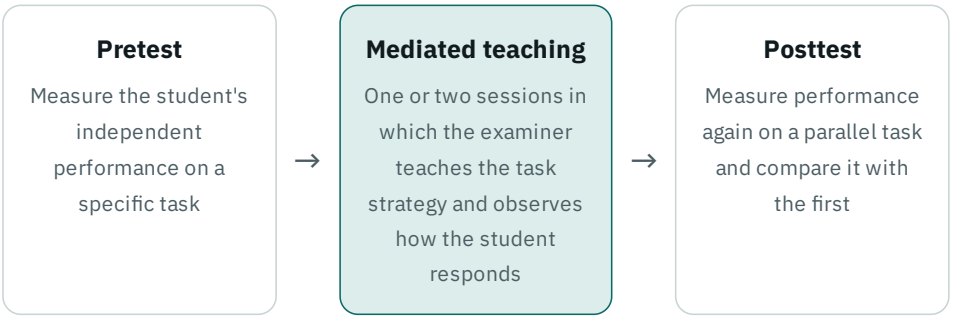



**Figure 5.1:** The test–teach–retest model. More important than the final score are the amount of improvement and the way the student responds during teaching.

During mediated teaching, the examiner records observations of **modifiability**, that is, how much the student benefits from instruction: Did the student pay attention? Apply the new strategy? Need a great deal of repetition? Transfer it to a slightly different task?

## Distinguishing difference from disorder

One of the most important uses of dynamic assessment is to distinguish a student who performs poorly because of insufficient exposure to the language or skill from one who performs poorly because of a genuine disorder. A study of children from diverse linguistic backgrounds found that children with typical language development improved markedly after two brief sessions on strategies for naming objects, whereas children with a language disorder improved only slightly, even though both groups started with similarly low scores (Peña et al., 2001).

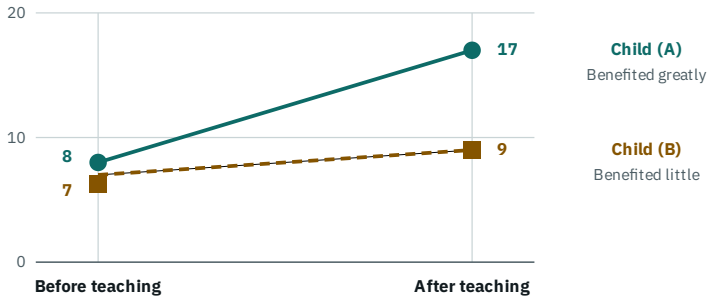



**Figure 5.2:** A hypothetical example of two children who both started with low scores (8 and 7 out of 20). Child (A), shown by the solid line and circular markers, jumped to 17 after two sessions, which suggests that the low initial score stemmed from a lack of opportunity. Child (B), shown by the dashed line and square markers, remained low despite the same teaching, which calls for a deeper assessment of a possible disorder.

Let us return to Nizar in the opening scene. His low score on the vocabulary test says nothing about its cause. So, over two short sessions, the examiner teaches him the names of a new set of objects, along with a strategy: ask for the name of anything he does not know, and link each new name to its category (this is a fruit, this is a kitchen utensil). The examiner then measures how well he learns a parallel set that was not taught. If his performance resembles that of Child (A) in Figure 5.2, learning quickly, transferring the strategy and needing few prompts, this suggests that his low score reflects limited exposure to Arabic, and the appropriate response is intensive language support and monitoring. If it resembles Child (B), the next step is to complete the assessment before any judgment of a language disorder, and among its most important questions is how his first language developed. A true language disorder appears in both languages, whereas difficulty acquiring a second language is confined to the new language.

## Graduated prompts

One form of dynamic assessment has the examiner offer **graduated prompts**, moving from general to specific, and count how many prompts the student needed to reach the answer. A student who needs one general prompt differs greatly from a student who must be given the solution step by step.

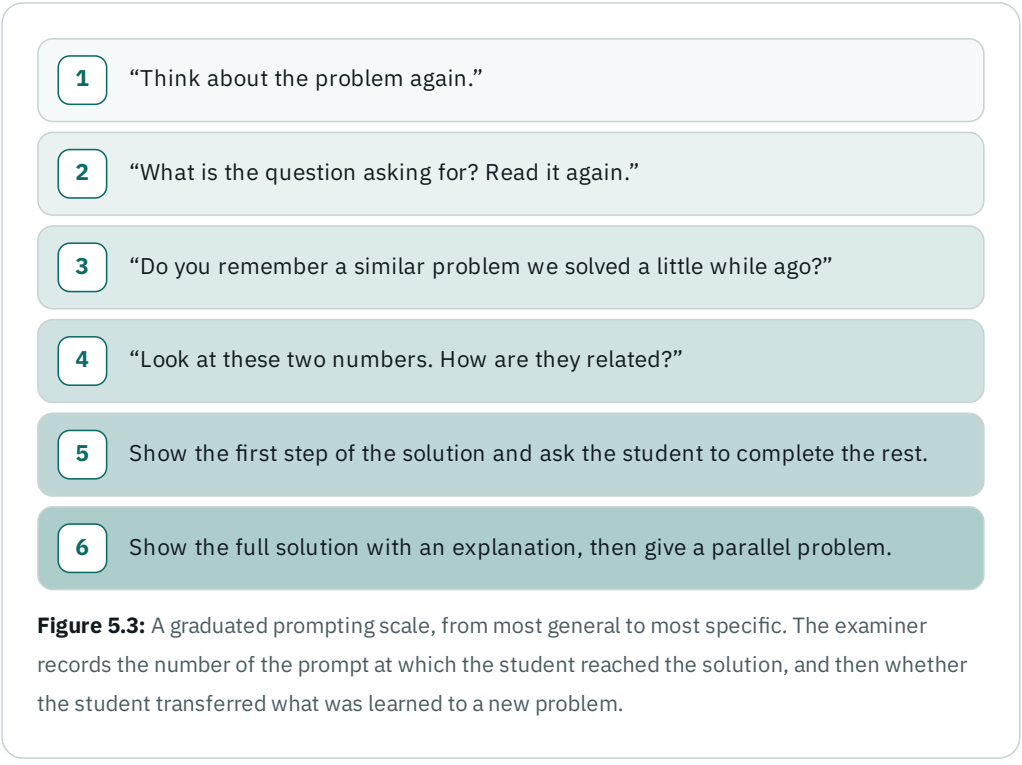



**Figure 5.3:** A graduated prompting scale, from most general to most specific. The examiner records the number of the prompt at which the student reached the solution, and then whether the student transferred what was learned to a new problem.

## Two formats of dynamic assessment

Sternberg and Grigorenko distinguish two main formats on which most applications of dynamic assessment are built (Sternberg & Grigorenko, 2002). The first places all the teaching between two tests; the second distributes it across the items themselves, as in the prompting scale above:

**Table 5.2: The two formats of dynamic assessment and what each measures.**

| Format | How it is administered | What is recorded | When it fits |
|---|---|---|---|
| **"Sandwich" format** | A pretest, then one or more teaching sessions, then a posttest on a parallel task | The difference between the two tests, and transfer of learning to a new task | When the aim is to learn the effect of intensive teaching on a specific skill, such as learning new vocabulary |
| **"Cake" format** | Each item is presented; if the student errs, graduated prompts are given on that same item until the student reaches the solution | The number and type of prompts on each item, and their decline as the items progress | When the aim is to build an accurate picture of how much help, and what kind, the student needs |

Dynamic assessment also varies in how standardized it is. Some forms are flexible and clinical, with the examiner adapting the mediation to what he or she sees in the student, as in Feuerstein's approach. Others are standardized, with the examiner offering the same prompts in the same order to every student. The flexible forms yield richer information about how the student learns, but the standardized ones make comparisons across students easier and come closer to the requirements of reliability.

## Dynamic assessment and response to intervention

In its underlying idea, dynamic assessment converges with the **response to intervention** (RTI) model discussed in Chapter 6: both look beyond the student's level to ask how the student responds when taught. Grigorenko argued that the two belong to a single family of methodologies whose defining feature is the merging of assessment and intervention into one activity, and she urged RTI proponents to draw on the successes and failures accumulated in dynamic assessment, which predates RTI by decades (Grigorenko, 2009). The

practical difference lies in time span and setting: RTI is monitored in the classroom over weeks through repeated measurement, whereas dynamic assessment compresses the same question into one or two sessions run by an examiner. Dynamic assessment can therefore serve as a quick preliminary indicator, but it is no substitute for monitoring the response to genuine intervention in the classroom.

**⚠ Limits of dynamic assessment**

It takes longer than traditional testing, depends heavily on the examiner's skill, and usually lacks norms. It is therefore used as a complement to traditional assessment, and its results are presented descriptively in the report, showing how the student responded to teaching, rather than being reduced to a standard score (Haywood & Lidz, 2007).

What reviews say about its predictive validity is mixed. Caffrey and two colleagues reviewed twenty-four studies and found that its ability to predict later achievement varied with its form, the student population and the type of criterion: prediction was better when help did not depend on the student's responses, when it was applied to students with disabilities rather than to at-risk or typically developing students, and when the criterion was a criterion-referenced test or an independent dynamic assessment rather than a norm-referenced test or teacher ratings (Caffrey et al., 2008). In a more recent systematic review of fourteen studies on identifying dyslexia, dynamic assessment explained variance in prediction that traditional tests did not, but classification accuracy improved when it was combined with traditional measures, was highest when it targeted skills close to reading such as decoding, and its predictive power declined with age (Dixon et al., 2023). The bottom line: dynamic assessment adds real information, but it cannot by itself serve as the basis for an eligibility decision.

**For the specialist: how is benefit from teaching measured?**

The simplest way to measure benefit is to subtract the pretest score from the posttest score. But a difference score is less reliable than either of the two scores, and the higher the correlation between the two tests, the lower the reliability of the difference, because the error in both measurements accumulates in it while the true variance shrinks. The difference is also affected by regression toward the mean: a student whose pretest score was very low because of measurement error will appear to have "benefited" even without teaching. Designers therefore turn to alternatives: relying on the number of prompts needed (the cake format), on the posttest score after statistically controlling for the pretest, or on items that test transfer of learning to an untaught task. The practical rule for anyone reading a dynamic assessment report: read "the score rose by 5 points" together with the kind of help the student needed and the student's ability to transfer.

## 5.3 Functional behavioral assessment

When a problem behavior recurs, people tend to explain it by a trait of the student: "stubborn," "a troublemaker," "wants attention." **Functional behavioral assessment** (FBA), by contrast, starts from a different assumption: **every behavior that persists does so because it gets the student something**. Our task is to find out what it gets, that is, what its **function** is. Once we know the function, we can teach the student a more appropriate way to achieve it, and the problem behavior decreases.

### Four common functions of behavior

Functions of behavior are classified along two dimensions: Is the student getting something or escaping something? And does it come from other people or from internal sensation (Iwata et al., 1994)? A single behavior may be

maintained by more than one function in the same student.

| | From other people or the environment | From within the body or the senses |
|---|---|---|
| **Getting something** | **Attention or a tangible item**<br>He screams, and the teacher comes over and talks to him. He cries, and he is given the tablet. | **Sensory stimulation**<br>He rocks his body or makes sounds because the sensation itself is pleasant, and he does it even when alone. |
| **Escaping something** | **A task or situation**<br>He throws the book and is sent out of the lesson, so he avoids the difficult task. | **An unpleasant sensation**<br>He covers his ears and leaves the crowded hall to escape the noise, or hits his head when he is in pain. |

**Figure 5.4:** Four common functions of behavior. The same behavior may serve different functions in different students, so its treatment is built on its function before its form.

**Misconception: "He does it to get attention"**

This is the most common explanation in schools, and the most deceptive. Teachers turn to the student after most disruptive behaviors, so attention always seems to be a "consequence" of the behavior. But the presence of attention after a behavior does not mean that attention is what maintains it. The student may be escaping the task, with attention merely a side effect. As we will see under the limits of descriptive assessment below, this confusion is documented in research.

## Steps of functional assessment

1. **Define the behavior operationally,** that is, describe it in terms of actions that can be observed and counted, such as "throws the book on the floor" rather than "aggressive" or "uncooperative."
2. **Gather indirect information:** interviews with teachers and the family, and questionnaires about when and where the behavior occurs and who is present.
3. **Direct descriptive observation:** recording what precedes and follows the behavior in its natural settings over several days, known as antecedent-behavior-consequence recording (ABC).
4. **Formulate a hypothesis** about the function, based on the pattern that emerges in the data.
5. **Test the hypothesis,** either by trying an intervention built on it and observing its effect, or through an experimental **functional analysis** that systematically varies conditions and is carried out by a behavior analyst (Hanley et al., 2003).
6. **Build a behavior intervention plan** that teaches the student an acceptable replacement behavior serving the same function.

The sources of information in these steps differ in the strength of the evidence they provide about function, as Table 5.3 shows:

**Table 5.3: Levels of evidence in functional behavioral assessment.**

| Level | What it provides | Its limits | Who usually carries it out |
|---|---|---|---|
| **Indirect (interviews and questionnaires)** | A quick picture of the conditions surrounding the behavior, and initial hypotheses | Depends on the respondent's memory and interpretation, and respondents may disagree | The teacher, the specialist, and the counselor |

| Level | What it provides | Its limits | Who usually carries it out |
|---|---|---|---|
| **Descriptive (antecedent-behavior-consequence recording)** | An actual pattern from the natural setting | Correlational: shows what accompanies the behavior, not what causes it | The teacher after training, and the specialist |
| **Experimental (functional analysis)** | Causal evidence: one condition is changed and its effect on the behavior observed | Deliberately evokes the behavior, so it requires a specialist, family consent and safety safeguards | A behavior analyst |
| **A trial of an intervention based on the hypothesis** | Practical evidence: does the behavior decrease when the function is met in an acceptable way? | Slower, and the behavior may improve for other reasons that coincide with the intervention | The school team under specialist supervision |

## Example: Badr and the math book

Badr is the fourth grader from the opening scene who throws his book during math lessons. His teacher recorded what came before and after his behavior over ten days, then classified what happened after it, because the consequence that follows a behavior is what maintains it. Here are examples from her record:

**Table 5.4: Examples from Badr's antecedent-behavior-consequence record.**

| What happened before the behavior | Behavior | What happened after |
|---|---|---|
| **A worksheet of division problems was handed out** | Threw the book and said, “I don't want to” | Was sent to the vice principal's office |
| **Was asked to solve a problem on the board** | Threw the book | Was seated at the back of the class with no task |
| **A group mental math activity began** | Threw the book and banged the desk | Was sent to the school counselor |
| **Free art lesson** | The behavior did not occur | Worked quietly |

| 8 | 1 | 1 | 0 |
|---|---|---|---|
| Ended with removal from the task | Ended with teacher attention | Ended with no change | In non-academic lessons |

The pattern is clear: the behavior occurs when he is asked to do a math task, and it usually ends with his removal from that task. Sending him to the vice principal, which was meant as punishment, in fact gave Badr exactly what he wanted, and this explains why the behavior increased. The hypothesis is formulated as follows:

When **Badr is asked to do a written or public math task**, **he throws the book and refuses to work**, and this leads to **his removal from the task**. The function of the behavior is therefore most likely **escape from the task**.

Yet this hypothesis remains **tentative**, because it rests on a descriptive record that shows what accompanied the behavior, not what caused it. It can be confirmed only by trying an intervention built on it, for instance teaching Badr to ask for a short break, honoring the request, and seeing the throwing decline, or by a functional analysis conducted by a specialist.

The question the assessment raises next matters for diagnosis: why does Badr escape from math in particular? The answer may reveal an as yet undetected learning disability in mathematics, making the behavior a symptom rather than a separate problem. It is therefore recommended that his math skills be assessed alongside a behavior intervention plan that teaches him to ask for help or a short break instead of throwing the book (O'Neill et al., 2015).

⚠ **Limits of descriptive assessment**

Antecedent-behavior-consequence data are correlational: they record what happens after a behavior, not what maintains it. Thompson and Iwata compared the results of descriptive assessment with those of functional analysis for twelve participants, and the most frequent consequence in the descriptive record matched the actual function in only three cases. Attention was the most frequent consequence following the behavior for eight of them, yet functional analysis confirmed that attention maintained the behavior for only two of those eight (Thompson & Iwata, 2007). In other words, descriptive records tend to inflate the attention function, because adults turn toward disruptive behavior whatever its function.

Because traditional functional analysis is lengthy and requires controlled conditions, shorter formats suited to schools and clinics have been developed. Among the best known is the **interview-informed synthesized contingency analysis** (IISCA), in which an open-ended interview with the family or teacher is used to identify the conditions most likely to evoke and maintain the behavior; these are then combined into a single test condition and compared with a control condition in which the same reinforcers are available freely (Hanley et al., 2014). Across thirty applications of this format, the analysis took about twenty-five minutes on average (Jessel et al., 2016). A recent review of a full decade of functional analysis research (2012–2022) shows a trend toward shorter sessions, more use in outpatient clinics, and the pairing of analysis with complementary assessments (Melanson & Fahmie, 2023). One criticism of the synthesized format is that it shows what maintains the behavior in combination, without isolating each function separately.

In all its formats, functional analysis deliberately evokes the problem behavior, so it should be conducted only by a qualified specialist, with written consent from the family and with stated criteria for ending the session if the behavior becomes dangerous.

**In the Saudi context: who conducts functional assessment?**

The Regulations assign assessment and diagnosis to a multidisciplinary team whose members include a "behavioral training teacher (psychologist)," and they list among this member's duties the use of formal and informal assessment tools "such as interviews, observation and checklists" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Articles 48–49 and 81). Interviews and descriptive observation, the mainstays of the first two levels in Table 5.3, thus lie at the heart of this team member's role. As for experimental functional analysis, we found no published regulatory text specifying who conducts it in schools or requiring any particular qualification for it.

*A question for reflection:* If your school has no behavior analyst, what can the team safely do to test a hypothesis such as the one about Badr?

**Functional assessment and diagnosis**

Functional assessment does not diagnose a disorder, but it complements diagnosis in two ways: it prevents a behavior with a clear function from being read as evidence of an emotional disturbance, and it sometimes uncovers a hidden need, such as a learning disability, sensory sensitivity or a communication weakness.

## 5.4 Authentic and performance assessment

**Authentic assessment**, and the closely related **performance assessment**, asks students to carry out a task resembling what people do in real life, rather than answer discrete questions. Instead of choosing the correct answer to a question about the area of a rectangle, the student is asked to plan a small garden on a set budget and calculate how much grass and fencing it will need (Wiggins, 1990). Performance is scored with a **rubric**, which is not the same as the "rating scales" a teacher or parent completes about a student's behavior: a rubric sets out **scoring criteria** and describes levels of mastery on each criterion.

**Table 5.5: Traditional testing and authentic assessment.**

| Aspect | Traditional test | Authentic assessment |
|---|---|---|
| **Task** | Short, discrete questions | An integrated task in a real context |
| **What it measures** | Mostly knowledge and recall | Application, planning, creativity and communication |
| **Answer** | Usually one correct answer | There may be several good solutions |
| **Scoring** | An answer key | A rubric with clear scoring criteria |
| **Weakness** | May not reflect the ability to apply knowledge | Time-consuming and affected by the scorer and the choice of task, so its reliability needs special attention |

### Authentic assessment in identifying gifted learners

Authentic assessment has a special place in identifying giftedness, because the abilities of many gifted learners do not show up on standardized tests, especially those from less advantaged linguistic or economic backgrounds and

those whose disabilities mask their abilities, namely twice-exceptional learners (discussed in Chapter 13). One of the best-known models here is the DISCOVER project, developed by June Maker, in which trained observers watch children solve problems that range from closed to fully open-ended across several domains, and record their thinking and their solutions (Maker, 1996).

This model has evidence from an Arab setting. In a study of 248 boys and girls in grades three to five at two private schools in Beirut, DISCOVER ratings of spatial intelligence correlated highly with scores on Raven's Progressive Matrices, while its ratings of linguistic intelligence correlated weakly with them, an expected pattern given that the Raven is a nonverbal test. The model identified 14.5% of the sample, with no differences between boys and girls, and interviews with families and teachers largely supported the results. The researcher concluded that the model is suitable for identification in Lebanon, while calling for further research (Sarouphim, 2009). Note, however, that the sample came from private schools in a single capital city, so the findings cannot be generalized to public or rural schools, and at the time of writing we had found no published study of the model's validity in Saudi schools.

Tala in the opening scene illustrates what this kind of assessment can reveal. Her average test scores do not capture the problem-solving, design ability and persistence that the real task displayed. This does not mean that she is necessarily gifted, but it does mean that she deserves a broader assessment drawing on other evidence (Chapter 12 covers the identification of gifted learners in detail).

## Reliability and validity of ratings when the stakes rise

The very appeal of authentic assessment makes measurement questions all the more pressing when high-stakes decisions rest on it, such as admission to a program or exclusion from it. Large-scale experience has documented three problems:

- **Scorer reliability:** In the US state of Vermont's portfolio assessment program in writing and mathematics, agreement among scorers was so low that the scores could not be used for dependable comparisons between schools and students, even though teachers reported that the program had a positive effect on their teaching practices (Koretz et al., 1994).
- **Task variability:** A student's performance on one performance task differs considerably from performance on another task in the same domain, and in the studies of Shavelson and colleagues this source of error was larger than differences among scorers. A dependable judgment therefore requires several tasks, however rich any single task may be (Shavelson et al., 1993).
- **Validity beyond appearance:** A task's being "real" does not establish its validity. Linn and two colleagues accordingly proposed criteria for validating complex assessments, among them consequences, fairness, generalizability and transfer, cognitive complexity, content quality and coverage, and cost (Linn et al., 1991).

The practical upshot: authentic assessment enters the giftedness decision as one piece of evidence among others, not as the sole gateway. It is strengthened by training scorers on the rubric with worked examples, by having each product scored by two independent scorers whose agreement is calculated, and by basing the judgment on more than one task.

## 5.5 Play-based assessment and assessment in early childhood

Young children, especially before school age, cannot sit quietly in front of an unfamiliar examiner who asks them questions, and when they are made to do so, the test measures their fear or boredom more than their abilities. The way around this is to observe the child in the activity most familiar to him or her: play.

In **play-based assessment**, the child plays in a room with carefully selected toys, first with a parent and then with a trained facilitator, while a multidisciplinary team observes, for example a speech-language pathologist, an occupational therapist, a psychologist and a special education teacher. This model yields information on four developmental domains at once (Linder, 2008):




**Figure 5.5:** The four developmental domains the team observes during a single play-based assessment session.

Some standardized instruments are built on the same idea, such as the Autism Diagnostic Observation Schedule (ADOS-2), which uses standardized play activities to elicit and observe communication and social interaction behaviors. The difference is that play-based assessment is more flexible and covers all areas of development, whereas the ADOS-2 is standardized and focused on the features of autism spectrum disorder (Chapter 9 discusses it in detail).

**For the early childhood teacher**

Free play time in kindergarten is a daily opportunity for assessment that the child does not even notice. Watch: Whom does the child play with? Does the child imitate other children? Engage in pretend play, such as feeding a doll? How does the child react when another child takes away a toy? Record your observations systematically: the date and time, the situation, and exactly what the child did and said, in words that describe what you saw rather than what you infer.

## Assessment from birth to three: what is different?

Assessing an infant or a child under three is not assessing an older child with smaller tools. Development in these years is rapid and uneven, the child is not assessed in isolation from the family, and the first aim is to determine who needs **early intervention** and what they need; classification comes later. In the United States, for example, this stage is covered by Part C of the Individuals with Disabilities Education Act (IDEA), under which eligibility rests on a developmental delay defined by each state or on a diagnosed condition likely to result in delay, and the child receives an individualized family service plan rather than an individualized education program (Individuals with Disabilities Education Act, 2004). Table 5.6 summarizes what follows from the characteristics of this stage:

**Table 5.6: Characteristics of assessment from birth to three and what they mean for the assessor.**

| Characteristic of the stage | What it means for the assessor |
|---|---|
| **Development is rapid and irregular, and performance is affected by sleep, hunger, illness and familiarity** | Scores at this age are less stable and weaker predictors of the future than at school age. No lasting judgment should rest on a single session, and assessment is repeated at short intervals |

| Characteristic of the stage | What it means for the assessor |
|---|---|
| **Many instruments for young children lack adequate floors and appropriate item gradients** | Two children who differ in ability may obtain the same lowest score. Before choosing an instrument, check the technical manual for the floor and item gradient at the child's age (Bracken, 1987) |
| **The family is the most important source of information and the primary partner in services** | The family is involved in planning the assessment, carrying it out and interpreting its results, and is asked about its priorities, concerns and daily routines |
| **The child shows his or her best in a familiar environment** | Observation at home or in daycare during daily routines, such as eating, dressing and playing, is preferred over the testing room alone |
| **Developmental domains are intertwined** | The child is assessed by a multidisciplinary team, using multiple methods that combine parent report, observation, and a standardized instrument |

The recommended practices in early intervention affirm these principles: involving the family, using multiple methods and sources, assessing in natural environments, and matching instruments to the child's language and culture (Division for Early Childhood, 2014). This does not mean dispensing with standardized instruments. When the extent of delay must be determined for an eligibility decision, an individually administered standardized developmental instrument is used, such as the Bayley Scales of Infant and Toddler Development, Fourth Edition (**Bayley-4**), which covers ages 16 days to 42 months in the cognitive, language, and motor domains and is supplemented by measures of social-emotional and adaptive behavior (Bayley & Aylward, 2019). The standardized instrument and play do not compete: the first answers “How much delay?” relative to peers, and the second answers “How does the child function?” in everyday life.

**In the Saudi context: Early intervention and early detection**

**The legal framework:** Article 8 of the Law on the Rights of Persons with Disabilities provides for "designing and implementing early intervention programs for children with disabilities" and for providing and adapting "modern assessment and diagnostic tools" (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic]). The Special Education Organizational Guide defines two groups in early intervention: **developmentally delayed** children, who show a delay "of two standard deviations below the mean in two or more domains" or "at a rate of 25%," and children **at risk**, who have been exposed to at least three environmental risk factors, provided that the child is no older than six (Ministry of Education, 1436 AH [Arabic]).

**An example of the delay percentage:** Ghaith is two and a half (32 months old), and his communication performance is equivalent to that of a 24-month-old. The delay is 8 months out of 32, or 25%, so he reaches the stated threshold in this domain. Caution is needed with this method, however: age equivalents are not equal units, a delay of eight months at age two is not the same as at age five, and an age equivalent is less reliable than a standard score. It is therefore advisable to pair the percentage calculation with a standard score from a standardized instrument and with observation of the child in daily routines.

**Detection at birth:** The "National Newborn Screening Program for the Prevention of Disability" has operated since 2005. It screens newborns within the first 24–72 hours, and the number of conditions covered has stood at seventeen since 1435 AH (Ministry of Health, 2016 [Arabic]). A national newborn hearing screening program has operated since 2016, and in 2021 its coverage reached about 92.6% of newborns (Alothman et al., 2024). A recent systematic review, however, identified challenges, including low awareness, loss to follow-up after referral, and weak data-

tracking systems (Alanazi & Nicholson, 2025). Detection at birth is thus no substitute for monitoring development afterward, and most cases of developmental delay are not detected by newborn screening.

*A question for reflection:* If a two-year-old girl were referred to you for delayed speech, from what sources would you gather evidence before deciding that she meets the Guide's threshold for "developmental delay"?

## 5.6 Assessing learners with severe and multiple disabilities

The Kingdom's Special Education Organizational Guide includes "multiple disabilities" and "deafblindness" among its categories (Ministry of Education, 1436 AH [Arabic]). Students in these two categories, along with those with severe intellectual disability, are the ones most failed by standardized instruments. Most tests assume that the student can see the materials, hear the instructions, point or speak to respond, and sustain attention for several minutes at a stretch. If any of these conditions is missing, the test no longer measures what it was designed to measure. Even when it can be administered, the student's score often falls at the test's floor, so it neither distinguishes one student from another nor says anything about what the student can actually do. Adapting the administration may enable the student to respond, but it takes the test outside its standardization conditions, so the score can no longer be validly compared with the norms.

For this group, the question therefore shifts from "Where does the student stand relative to peers?" to "What does the student do now, with what help, and what is the next step that can be taught?" All the alternative methods converge in answering it:

**Table 5.7: Sources of evidence when standardized instruments are not suitable.**

| Source | What it answers | Example |
|---|---|---|
| **Observation during daily routines on repeated occasions** | What the student actually does, and how much help is needed at each step | At breakfast: Does he reach toward the cup? Does he open his mouth as the spoon approaches? |
| **Interviews with the family and those who work with the student daily** | The student's best performance, and how he expresses wanting, refusal, and distress | “He calms down when he hears his brother's voice”; “He turns his head away when he doesn't want the food” |
| **Assessment of pre-symbolic communication** | By what means he communicates: body movement, vocalization, gesture, a tangible object, or a picture | He touches the cup to ask for water, so the cup becomes a tactile symbol to build on |
| **Functional vision and hearing assessment** | How he uses his available vision and hearing in real-life situations, alongside clinical acuity | From what distance and on which side does he turn toward a light-up toy? |
| **Preference assessment** | What he likes that could serve as a reinforcer in instruction | Presenting two items together and recording which one he chooses across several trials |
| **Response to structured instruction** | Does he learn a specific skill when it is taught in consistent steps, and how quickly? | Daily data on the number of successful trials on a single skill, which is the very logic of dynamic assessment |

In this group, the student's health has a large effect on performance: fatigue, seizures, medication, and body positioning can change the picture from one day to the next. Evidence is therefore gathered over several days and at different times, and each observation is recorded together with whatever is affecting the student's condition that day. The assessment involves a team that includes the family, the special education teacher, and specialists in movement, communication, vision, and hearing as the case requires, which is also what the Organizational Guide's requirement of "not relying on the results of a single measure" demands (Ministry of Education, 1436 AH [Arabic]). The chapters on specific categories treat the assessment of severe intellectual disability and of sensory and multiple disabilities in greater detail.

## Conclusion

What unites the methods we have examined in this chapter is that they shift the question from the student's standing among peers to what the student's performance reveals under specific conditions: response to instruction in dynamic assessment, the function of behavior in functional assessment, production in a real task in authentic assessment, and what young children and learners with severe disabilities do in their daily routines. It follows that behavior should be understood before it is addressed, since what is meant as punishment may be the very thing that maintains the behavior if its function is escape, and that a low score from a student with little exposure to the language of instruction should be treated as a question about opportunity before it is taken as evidence of a disorder.

These methods have their limits as well. The predictive value of dynamic assessment varies, so it cannot by itself serve as the basis for an eligibility decision; a hypothesis built on a descriptive record remains tentative until an intervention trial or a functional analysis confirms it; rater reliability and task variability make authentic assessment one piece of evidence among others; and scores in the early years are unstable. The value of these methods lies in completing and enriching the picture drawn by conventional tests.

# 5.7 Summary, key terms, and activities

## Chapter summary

- Alternative methods pose questions that conventional tests do not answer: How does the student learn? Why does the student behave this way? What does the student produce in a real task? What does the student do in daily life? They are distinct from "alternate assessment" in its U.S. regulatory sense.
- Dynamic assessment measures response to instruction in a "sandwich" or "cake" format and helps distinguish lack of opportunity from a genuine disorder. It adds predictive value to conventional tests, but that value varies, so it cannot by itself serve as the basis for an eligibility decision.
- Dynamic assessment and response to intervention belong to the same family: the first compresses the question into a single session, while the second tracks it in the classroom over weeks.
- Functional assessment looks for the function of a behavior, and the common functions are four: obtaining something or escaping it, whether from other people or from sensory experience. What is meant as punishment may be what maintains the behavior if its function is escape.
- A descriptive record is correlational and tends to overstate the attention function, so a hypothesis built on it is tentative until an intervention trial or a functional analysis conducted by a specialist confirms it.
- Authentic assessment reveals abilities that do not show up on short tests, and it is important in identifying gifted students. Rater reliability and task variability, however, require that it be one piece of evidence among others, based on more than one task and two independent raters.

- Play-based assessment suits young children. Assessment below age three centers on the family, the natural environment, and the team, and it combines play with a standardized developmental instrument, with caution about the instability of scores at this age.
- With learners who have severe and multiple disabilities, the question shifts from comparison with peers to what the student does now and with what help; evidence is gathered from routines, the family, communication assessment, preference assessment, and response to structured instruction.

## Glossary of terms

| | |
|---|---|
| **Alternate assessment** | In the U.S. system: an achievement test based on alternate achievement standards for students with the most significant cognitive disabilities; it does not refer to the “alternative methods” of this chapter. |
| **Zone of proximal development** | The distance between what a child can do alone and what the child can do with help from someone more experienced. |
| **Dynamic assessment** | Assessment in which the examiner deliberately intervenes and teaches, then measures the effect of the teaching and how the student responds to it. |
| **Mediated learning experience** | Learning in which an adult stands as a mediator between the child and the task, directing the child's attention, organizing information, and helping the child transfer what has been learned. |

| | |
|---|---|
| **Graduated prompts** | Aids the examiner provides, moving from the most general to the most specific, with a record of how many the student needed to reach the solution. |
| **Sandwich and cake formats** | Two formats of dynamic assessment: the first places teaching between a pretest and a posttest, and the second provides prompts within each item. |
| **Modifiability** | The extent to which a student benefits from mediated teaching and can transfer what has been learned. |
| **Functional behavioral assessment** | A process for determining the function of a problem behavior and the conditions that trigger and maintain it. |
| **Behavior function** | What a behavior achieves for the person, such as gaining attention or escaping a task. |
| **Descriptive assessment** | Recording what precedes and follows a behavior in its natural settings without changing the conditions; it shows what accompanies the behavior, not what causes it. |
| **Functional analysis** | Systematically changing conditions to test a hypothesis about the function of a behavior experimentally. |
| **Interview-informed synthesized contingency analysis** | An abbreviated form of functional analysis that combines the conditions identified in an open-ended interview into a single test condition compared with a control condition. |

| | |
|---|---|
| **Behavior intervention plan** | A plan based on the results of a functional assessment that teaches the student a replacement behavior serving the same function and changes the conditions that trigger the problem behavior. |
| **Authentic assessment** | Assessment that requires performing tasks resembling real-life tasks, scored with rubrics. |
| **Rubric** | A scoring tool that specifies the scoring criteria and the levels of mastery on each criterion, used in assessing open-ended tasks. It differs from rating scales, with which a teacher or parent rates a student's behavior. |
| **Play-based assessment** | Observing a young child during play to assess the various developmental domains. |
| **Early intervention** | Services provided to young children who are developmentally delayed or at risk of delay, and to their families; each country's legal system sets the age range and eligibility criteria. |

## Review questions

1. How does the examiner's role in dynamic assessment differ from that in a standardized test? And why is a rise in score between pretest and posttest not, by itself, sufficient evidence of benefit?
2. Two students who recently arrived from another country obtained low scores on a vocabulary test. Design a simple dynamic assessment to help distinguish the student whose low score reflects limited exposure to Arabic from the one who may have a language disorder, and specify what you need to know about their home language before reaching a judgment.

3. True or false? Explain: “Since dynamic assessment measures the ability to learn directly, it is sufficient on its own to determine a student's eligibility for learning disabilities services.”
4. A girl screams at recess when other children approach her. Propose two possible functions for this behavior, and specify the data that would distinguish between them.
5. Why did Badr's behavior increase despite punishment? Why does the hypothesis about the function of his behavior remain tentative? And what would confirm it?
6. Design an authentic performance task in science for fifth graders, with three scoring criteria and clear performance levels. Then explain what would be needed for its results to serve as evidence in an admission decision for a gifted program.
7. When would you prefer play-based assessment to a standardized developmental test for a two-year-old? And when would you still need the standardized test?
8. A six-year-old girl has a severe physical disability and severe visual impairment and does not speak. Name four sources of evidence on which to base her assessment, and explain what each source contributes.

## Case for application: Hattan, age 14, Grade 8

**The story:** Hattan's teachers referred his behavior to the school team because he “provokes the teacher until he gets sent out of class,” and they believe he “likes his classmates to laugh at him.” A team member recorded what preceded and followed his behavior in English and science classes over eight days; mathematics and physical education classes were not recorded. Hattan's grades in mathematics are good, and he participates orally there, but his grades in reading-dependent subjects are low, and his reading has never been assessed.

**Table 5.8: Summary of Hattan's antecedent-behavior-consequence record (eight incidents over eight days).**

| What happened before the behavior | Frequency | What happened after |
|---|---|---|
| **He was asked to read a passage or question aloud in front of the class** | 5 | Sent out of class (4); reprimanded by the teacher in front of classmates, who laughed (1) |
| **Group work began in which students took turns reading the instructions** | 2 | Classmates laughed and the teacher excused him from his turn to read (1); reprimanded by the teacher (1) |
| **The teacher explained the lesson orally** | 1 | Nothing changed |

**Tasks:** (a) Formulate two competing hypotheses about the function of the behavior, and show which data support each. (b) Is this evidence sufficient for a confident hypothesis? Justify your judgment, and identify the missing evidence and how you would gather it. (c) What would you propose assessing in addition to the behavior, and why?

**Table 5.9: Mapping chapter objectives to questions and activities.**

| Objective | Questions | Activities |
|---|---|---|
| **1 Applying the test-teach-retest model** | 1, 2 | 1 |
| **2 Interpreting a dynamic assessment of a student from a different linguistic background** | 2 | 1 |
| **3 Formulating a hypothesis about the function of behavior and identifying the evidence to confirm it** | 4, 5, Hattan's case | 2 |
| **4 Designing an authentic task and critiquing the reliability of its scoring** | 6 | 3 |

| Objective | Questions | Activities |
|---|---|---|
| **5 Choosing between play and a standardized developmental test** | 7 | — |
| **6 Sources of evidence for learners with severe and multiple disabilities** | 8 | 4 |
| **7 Judging the sufficiency of evidence from an alternative method** | 3, Hattan's case | 3 |

## Learning activities

For any activity carried out with a student or child: it is conducted with the consent of the family and the school administration, without any data that identify the student, and under professional supervision; no decision is based on its results, and they are communicated only through official channels.

1. **A dynamic session** (about two hours): Choose a task, such as sorting pictures or solving shape puzzles, and apply the test-teach-retest model to it with a child, recording the number of the prompt needed on each item on a scale like the one in Figure 5.3. *Product:* a table of prompts and a paragraph interpreting modifiability and the limits of what you can conclude. *Desk alternative:* conduct the session with an adult colleague on a task unfamiliar to them, such as learning a writing system they do not know.
2. **Functional assessment** (five days of recording and an hour of analysis): Choose a recurring behavior of a student, record what comes before and after it, then classify the results and formulate a hypothesis modeled on the one about Badr. *Product:* the record table, the hypothesis, and a two-step plan for confirming it through an intervention trial. *Desk alternative:* record in the same way a behavior of your own, such as procrastinating on a task or browsing your phone while studying.

3. **An authentic task and the reliability of its scoring** (about three hours): Design an open-ended task for solving a problem from the school environment, with a three-criterion rubric, and try it out with a small group. Then have a colleague score the same products independently of you. *Product:* the task and its rubric, the percentage of agreement between you on each criterion, and a paragraph on the reasons for disagreement and what is needed to improve the rubric.
4. **An assessment plan for a student with multiple disabilities** (desk activity, about one hour): Drawing on Table 5.7, write an assessment plan for the girl in Question 8: who takes part, where and how often she is observed, what questions are put to the family, and how the evidence is recorded. *Product:* a one-page plan.

## Chapter references

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Health. (2016). *[The national newborn screening program for the prevention of disability: Operational manual]* [In Arabic]. https://www.moh.gov.sa/en/Ministry/About/Health%20Policies/019.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alanazi, A. A., & Nicholson, N. (2025). Challenges of newborn hearing screening programs in Saudi Arabia: A systematic review. *Audiology Research, 15*(2), Article 34. https://doi.org/10.3390/audiolres15020034

Alothman, N., Elbeltagy, R., & Mulla, R. (2024). Universal newborn hearing screening program in Saudi Arabia: Current insight. *Journal of Otology, 19*(1), 35–39. https://doi.org/10.1016/j.joto.2024.01.002

Bayley, N., & Aylward, G. P. (2019). *Bayley Scales of Infant and Toddler Development* (4th ed.). Pearson.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*(4), 313–326.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*(4), 254–270.

Division for Early Childhood. (2014). *DEC recommended practices in early intervention/early childhood special education 2014*. http://www.dec-sped.org/recommendedpractices

Dixon, C., Oxley, E., Nash, H., & Gellert, A. S. (2023). Does dynamic assessment offer an alternative approach to identifying reading disorder? A systematic review. *Journal of Learning Disabilities, 56*(6), 423–439. https://doi.org/10.1177/00222194221117510

Every Student Succeeds Act, 20 U.S.C. § 6301 et seq. (2015).

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device*. University Park Press.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111–132.

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*(2), 147–185.

Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. *Journal of Applied Behavior Analysis, 47*(1), 16–36.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge University Press.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 et seq. (2004); implementing regulations for infants and toddlers (Part C) 34 CFR Part 303.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. (Original work published 1982)

Jessel, J., Hanley, G. P., & Ghaemmaghami, M. (2016). Interview-informed synthesized contingency analyses: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis, 49*(3), 576–595.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13*(3), 5–16. https://doi.org/10.1111/j.1745-3992.1994.tb00443.x

Linder, T. (2008). *Transdisciplinary play-based assessment* (2nd ed.). Paul H. Brookes.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. *Gifted Child Quarterly, 40*(1), 41–50.

Melanson, I. J., & Fahmie, T. A. (2023). Functional analysis of problem behavior: A 40-year review. *Journal of Applied Behavior Analysis, 56*(2), 262–281. https://doi.org/10.1002/jaba.983

O'Neill, R. E., Albin, R. W., Storey, K., Horner, R. H., & Sprague, J. R. (2015). *Functional assessment and program development for problem behavior: A practical handbook* (3rd ed.). Cengage Learning.

Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138–154.

Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted Lebanese students: Is DISCOVER effective? *Journal for the Education of the Gifted, 33*(2), 275–295. https://doi.org/10.1177/016235320903300206

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*(3), 215–232.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge University Press.

Thompson, R. H., & Iwata, B. A. (2007). A comparison of outcomes from descriptive and functional analyses of problem behavior. *Journal of Applied Behavior Analysis, 40*(2), 333–338.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*(2), Article 2.

Chapter 6

# 6 Identification models

The tools are many and the evidence varied, yet one question remains: how is it all brought together into a single decision? Do we look for a gap between intelligence and achievement, wait to see what a student's response to sound instruction reveals, or search for a pattern in the student's abilities? And how are several scores combined to judge that a student is gifted? In this chapter we present the models that take up this question of decision, and then read the Saudi regulations in their light.

---

**Scene from two team meetings**

Two third graders both read very slowly and make many errors. At one school, the team said of the first: “His intelligence is average and his achievement is low. The gap is large, so he has a learning disability.” Of the second, they said: “His intelligence is slightly below average and the gap is small, so he does not have a learning disability. He is just a slow learner.”

At another school, the team said of both students: “We won’t make a judgment now. We’ll teach them reading intensively for eight weeks and measure their progress every week. Whoever still doesn’t improve will need a comprehensive evaluation.”

The students are the same, and so is the need, yet the decision changed with the model each school relies on. That is why understanding identification models is a practical necessity for every team.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Calculate the ability–achievement discrepancy for a given student using both the simple-difference and regression methods, and explain how the standard error of the difference affects the stability of the decision.
2. Apply the common nonresponse criteria (final status, slope and dual discrepancy) to a given set of progress-monitoring data, and state at least one limitation of each.
3. Compare the three versions of the patterns of strengths and weaknesses model in terms of what each requires, and summarize one piece of supporting evidence and one piece of critical evidence.
4. Apply the criteria of the hybrid model to a given case, and match them against the DSM criteria for specific learning disorder.
5. Apply the three combination rules to the scores of given students, and determine whom each rule excludes.
6. Analyze the Saudi regulatory texts on learning disabilities and giftedness using the concepts of this chapter, and identify whom they may miss.
7. Judge whether the evidence is sufficient to identify a learning disability in a new case, justify the judgment, and name the missing evidence.

# Introduction

No single test can detect a learning disability or giftedness; both are judgments built on multiple sources of evidence. A team therefore needs a model that specifies which evidence is gathered, how it is combined, and where the dividing line falls. Choosing a model is a decision with direct consequences for students. As the opening scene shows, the same student may be considered to have a learning disability under one model and not under another, and in giftedness the rule for combining scores changes who is identified as gifted and who is excluded. Understanding the logic and limits of each model is a prerequisite for reading the regulations within which a team works, and for recognizing what those regulations leave to its professional judgment.

We open the chapter by explaining why a decision model is needed (Section 6.1), then present the discrepancy model and the criticisms leveled at it (Section 6.2), followed by response to intervention, the dual-discrepancy criterion and when nonresponse can be determined (Section 6.3). We then take up the patterns of strengths and weaknesses model in its three versions and the debate surrounding it (Section 6.4), the hybrid models with a worked case

(Section 6.5), and multiple-criteria models in giftedness, combination rules, universal screening and local norms (Section 6.6). We close by comparing the models and reading the Saudi regulations in their light (Section 6.7).

## 6.1 Why do we need a model?

Some categories have relatively clear criteria: hearing impairment is measured with audiometric equipment, and autism spectrum disorder has detailed criteria in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**; American Psychiatric Association, 2022). Two categories in particular, however, need a model that specifies how we reach a decision: **learning disabilities**, because no single test detects them, and **giftedness**, because its very definition differs from one system to another. A model answers three questions: What evidence do we gather? How do we combine it into a decision? And where do we set the line?

In this chapter, "learning disabilities" names the educational category that qualifies for services in school systems, and "specific learning disorder" (**Specific learning disorder**) is used when the discussion concerns the DSM criteria. The two concepts are close, each has its own criteria, and we set them side by side in Section 6.5.

## 6.2 The discrepancy model

For decades, the **discrepancy model** (**Discrepancy model**) was the most widely used approach to identifying learning disabilities. Its idea is simple: a student with a learning disability is intellectually capable, but his achievement in a particular area falls well below what his ability would predict. In other words, there is a **discrepancy** between ability and achievement (**Ability-achievement discrepancy**). It is usually calculated as the difference between the standard score on an intelligence test and the standard score on an achievement test, and the difference must reach a set threshold, such as one and a half standard deviations, that is, 22.5 points on a scale with a mean of 100 and a standard deviation of 15.

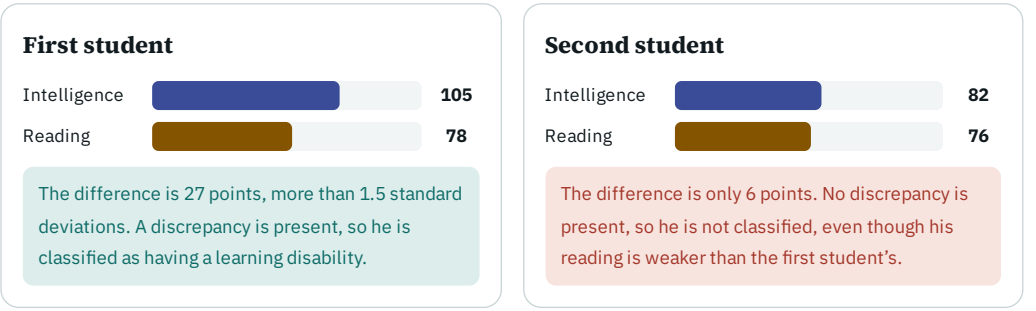



**Figure 6.1:** A hypothetical example of the simple-difference discrepancy model. In each card, the upper bar shows intelligence and the lower bar shows reading, with the score beside each bar. The two students struggle in reading to roughly the same degree, yet the model grants services to one and withholds them from the other.

**Misconception: "He doesn't have a learning disability; he's just a slow learner"**

The first team's remark about the second student in the chapter's opening scene is common in our schools. It sounds like a diagnosis, but it is in fact a popular description with no basis in any classification. "Slow learning" is not a category in the DSM or the International Classification of Diseases, nor among the categories of U.S. law, nor among those adopted by the Special Education Organizational Guide in the Kingdom (Ministry of Education, 1436 AH [Arabic]). Its closest counterpart is "borderline intellectual functioning" (**Borderline intellectual functioning**). Earlier editions of the DSM defined it by a numerical range; the current edition lists it under "other conditions that may be a focus of clinical attention," with no numerical cutoff, and does not regard it as a disorder (American Psychiatric Association, 2022).

The practical danger is that this remark closes the file. The student's intelligence is above the threshold for intellectual disability, so he does not qualify for those services, and the gap between his intelligence and his achievement is small, so under discrepancy logic he does not qualify for learning disability services either. He is left without support, even though his reading weakness is real and he may respond to intensive intervention as well as anyone. The more productive question is "Does he learn when we teach him intensively, and what does he need?" Asking it serves him better than searching for a label for his condition, and it is the question the models in the rest of this chapter set out to answer.

## Why has the discrepancy model been criticized?

- **It does not distinguish who needs different treatment:** A large meta-analysis, that is, a study that pools and analyzes the results of many studies, found that poor readers with a discrepancy and poor readers without one are similar in reading-related skills and need similar instruction (Stuebing et al., 2002). The discrepancy, in other words, does not separate two genuinely different groups.
- **A wait-to-fail model:** In the early grades, reading demands are simple, so the gap does not grow wide enough until third or fourth grade. The student waits years, struggling all the while, before qualifying for services, even though early intervention is far more beneficial.
- **An unstable decision:** Both scores carry measurement error, and the two errors combine in the difference between them. If the reliability of the intelligence test is 0.95 and that of the reading test is 0.90, the standard error of the difference is about $15 \times \sqrt{(2 - 0.95 - 0.90)} \approx 5.8$ points, and the first student's observed difference of 27 is consistent, with 95% confidence, with a true difference of roughly 16 to 38. That range straddles the 22.5 cutoff, so we cannot say with certainty whether the student is above or below it. This is why a student may be classified this year and not next year without anything about him having changed (see the calculation of the standard error of the difference in Chapter 2, 2.5.1).
- **It depends on the tests used:** A different intelligence test or a different achievement test may change the entire decision.
- **It ignores regression toward the mean:** Intelligence and achievement are correlated, but imperfectly, so for a student with an IQ of 130 we expect achievement closer to the mean than 130. The simple difference overlooks this, as the following box shows.

**For the specialist: Regression-based discrepancy**

Rather than comparing achievement with the intelligence score itself, we compare it with the **expected** achievement of a student with that level of intelligence. The expected score is calculated with a regression equation. If both tests have a mean of 100 and a standard deviation of 15, and the correlation coefficient between them is r:

$$\text{Expected achievement} = 100 + r \times (IQ - 100)$$

We then compare actual achievement with expected achievement and judge the difference against the standard error of estimate: $15 \times \sqrt{(1 - r^2)}$

**Worked example: three students, two methods**

1. For illustration, assume that the correlation between the intelligence test and the reading test is r = 0.60. The actual correlation is taken from the test manual or from studies that have linked the two tests.
2. The standard error of estimate is $15 \times \sqrt{(1 - 0.36)} = 15 \times 0.8 = 12$. If we want a cutoff equivalent to 1.5 standard deviations, it falls at $1.5 \times 12 = 18$ points below the expected score.
3. First student (IQ 105, reading 78): the expected score is $100 + 0.6 \times 5 = 103$, and his reading falls 25 points below it, so the discrepancy holds under both methods.
4. A student with an IQ of 130 and reading of 105: the simple difference is 25, so he is classified. But the expected score is $100 + 0.6 \times 30 = 118$, and his reading falls only 13 points below it, so there is no discrepancy under the regression method.
5. A student with an IQ of 85 and reading of 68: the simple difference is 17, so he is not classified. But the expected score is $100 + 0.6 \times (-15) = 91$, and his reading falls 23 points below it, so the discrepancy holds under the regression method.

**Table 6.1: Judgments for three hypothetical students using the simple difference (cutoff 22.5) and regression (cutoff 18 below expected, assuming a correlation of 0.60).**

| Student | Intelligence | Reading | Simple difference | Judgment | Expected | Below expected | Judgment |
|---|---|---|---|---|---|---|---|
| **First (Figure 6.1)** | 105 | 78 | 27 | ✓ | 103 | 25 | ✓ |
| **High IQ** | 130 | 105 | 25 | ✓ | 118 | 13 | ✗ |
| **Slightly low IQ** | 85 | 68 | 17 | ✗ | 91 | 23 | ✓ |

The simple difference thus **over-identifies** students with high intelligence, because it expects their achievement to equal their intelligence, and **under-identifies** students with slightly low intelligence, because it expects lower achievement from them than regression does. This is one face of regression toward the mean (see Chapter 2, 2.5.2). The result bears directly on twice-exceptional learners: a gifted student with average achievement may appear "discrepant" under the simple difference but not under regression. Correcting the formula, however, does not solve the model's other problems; the meta-analysis cited above found no meaningful difference in reading-related skills between discrepant and nondiscrepant poor readers (Stuebing et al., 2002).

For these reasons, the U.S. Individuals with Disabilities Education Act (**IDEA**), as amended in 2004, and its implementing regulations provide that states may not require schools to use a discrepancy, must permit the use of a process based on response to intervention, and may permit other research-based procedures (U.S. Department of Education, 2006, 34 CFR §300.307).

## 6.3 Response to intervention

The **response to intervention** model (**Response to intervention, RTI**) turns the question around. Instead of asking "Is there a gap between the student's ability and achievement?", it asks "Does the student learn when we provide sound, evidence-based instruction?" If the student does not learn despite good instruction from which peers have learned, that is evidence of a difficulty requiring more than ordinary teaching (Fuchs & Fuchs, 2006). The model is usually implemented in three tiers:

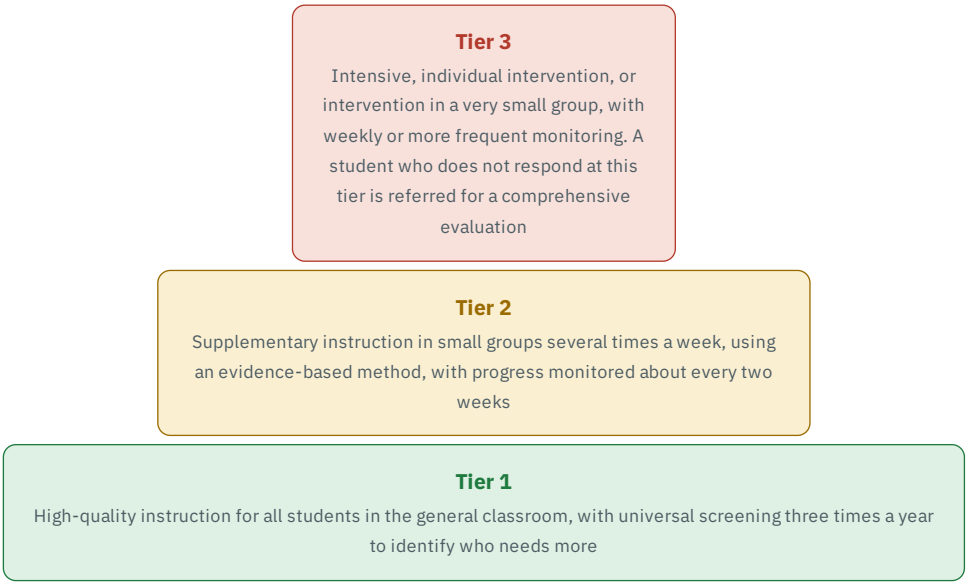



**Figure 6.2:** The tiers of response to intervention, from the broad base (Tier 1) to the narrow peak (Tier 3). Tier 1 is expected to suffice for most students; it is often said that about 80% need nothing more, about 15% need Tier 2, and about 5% need Tier 3. These proportions are guidelines, not fixed rules.

The broader framework that encompasses academic and behavioral domains together is called a **multi-tiered system of supports** (**Multi-tiered system of supports, MTSS**). The idea is the same: support that intensifies with need, and decisions based on repeated data. The best-known source of such data is **curriculum-based measurement**, presented in Chapter 4: brief, equivalent probes administered every week or two, such as the number of words a student reads correctly in one minute, with results graphed to show whether the student is moving toward the goal.

## The dual-discrepancy criterion

How do we know that a student "has not responded"? One of the best-known criteria is the **dual discrepancy** (**Dual discrepancy**), which Lynn Fuchs proposed as a basis for response decisions (Fuchs, 2003): the student is clearly below peers in the **level** of performance and also below them in the **rate of growth** despite intervention. A low-performing student who grows at the same rate as peers, or faster, is responding to intervention and simply needs time. A low-performing student who grows slowly, however, is falling further behind peers and needs more.

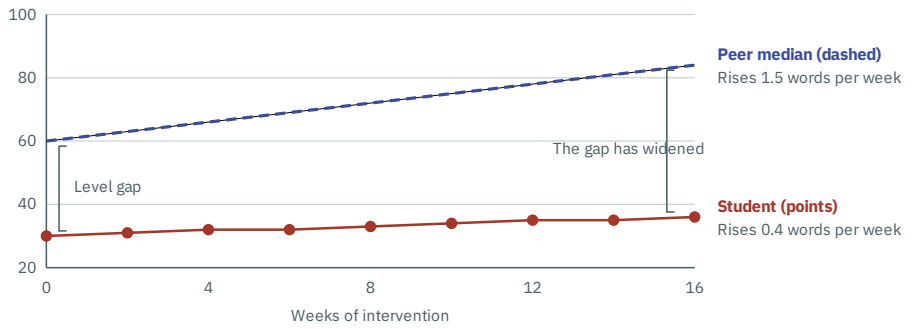



**Figure 6.3:** A hypothetical example of dual discrepancy in reading fluency (words correct per minute). The student started clearly below peers and grew more slowly than they did despite intervention, so the gap widened. This pattern calls for a comprehensive evaluation.

## When do we say a student has not responded?

"Nonresponse" is a judgment built on a criterion and a threshold set in advance, not on a general impression. Table 6.2 presents the best-known criteria in the literature, with examples of thresholds that have been used in research and practice and the limitations of each. The thresholds here are illustrative and vary from study to study. Some are expressed as percentile ranks, that is, the percentage of peers in the comparison sample (the norm sample or the students in the grade) who scored below the student; a rank of 25 means that the student outperformed 25% of them.

**Table 6.2: Common nonresponse criteria, example thresholds and their limitations.**

| Criterion | What it measures | Example thresholds used | Limitations |
|---|---|---|---|
| **Normative final status** | The student's level on a standardized test after intervention ends | Remaining below roughly the 25th percentile (standard score 90) after intervention | Ignores the amount of progress; a student who started very low and made substantial gains may be judged a nonresponder |
| **Benchmark** | Whether the student reaches a level set for the grade and term | The number of words correct per minute set for the end of the term in that grade | Relies on norms that may not exist in Arabic, and is affected by the difficulty of the passage chosen for measurement |

| Criterion | What it measures | Example thresholds used | Limitations |
|---|---|---|---|
| **Slope** | Weekly rate of growth during intervention | A slope below the 25th percentile of peer slopes, or below the slope needed to reach the goal on time | A slope calculated from only a few weeks is unreliable and is affected by variation in passage difficulty (Ardoin et al., 2013) |
| **Dual discrepancy** | Level and slope together | Both level and slope fall one standard deviation or more below the peer mean | Inherits the unreliability of the slope and is affected by the level of the reference class; the same student may be judged a responder in a weak class and a nonresponder in a strong one |

The deeper problem is that these criteria do not identify the same students. When different criteria were applied to the same sample of poor readers, agreement among them on who "had not responded" was generally weak (Barth et al., 2008). The decision thus depends as much on the criterion the team chose as on the student. Added to this is what Ardoin and colleagues found in their review of curriculum-based measurement decision rules: most of these rules rest more on expert opinion than on research, slopes calculated over short periods are unstable, and they shift with variation in the difficulty of passages that are supposed to be equivalent (Ardoin et al., 2013). Three practical conclusions follow: the team should set its criterion and threshold before collecting data, not after; it should collect data over a sufficient period using passages that are truly equivalent; and it should treat the slope as it would any score, that is, as a range of plausible values.

⚠ **Response to intervention does not delay evaluation**

The model can be misunderstood and turned into a pretext for waiting: "We won't evaluate him until he has completed Tiers 2 and 3." The Office of Special Education Programs in the U.S. Department of Education has cautioned that an RTI process may not be used to delay or deny a comprehensive evaluation when the family requests one or the school suspects a disability (U.S. Department of Education, Office of Special Education Programs, 2011). The data gathered across the tiers of support enter the evaluation as evidence; they do not replace it. This holds even more clearly for a student who shows signs of intellectual or sensory disability or a developmental disorder; such a student should not be made to wait until nonresponse has been "proven."

**For the teacher: What should you document during the tiers of support?**

The value of your data for a later decision depends on how well they are documented. For each student, record the program or method used; the number of sessions planned and actually delivered, and the length of each; the student's absences; the result of each measurement, with its date, on the same passage or an equivalent one; and any change in method, with its date. Ask a colleague or supervisor to observe at least two sessions using a checklist of the program's steps; this is **treatment fidelity** (**Treatment fidelity**). If the student does not progress, the team's first question will be: was the intervention delivered as planned? If the answer is no, nonresponse says more about the intervention than about the student.

## Strengths and limitations of the model

**Table 6.3: Strengths and limitations of the response to intervention model.**

| Strengths | Limitations |
|---|---|
| **Provides help early, without waiting for failure or a diagnosis** | Requires high-quality instruction, evidence-based interventions and careful monitoring, which are unevenly available across schools |
| **Rules out poor instruction as an explanation for the problem before the student is diagnosed** | If used on its own, it may delay the evaluation of a student who needs a comprehensive evaluation immediately, such as one suspected of having an intellectual disability |
| **Yields direct instructional data that are useful for planning** | Does not explain why the student failed to respond, nor distinguish among the causes of nonresponse |
| **Reduces misdiagnosis of students who simply need better instruction** | May miss twice-exceptional learners, that is, gifted students with a learning disability, because their performance may not fall below that of their peers |
| **Rests on repeated data rather than a single measurement** | Nonresponse criteria often disagree with one another, and the reliability of slopes over short periods is limited |

Criticism of the model also comes from the opposite direction. Reynolds and Shaywitz argued that response to intervention has not yet proven its validity as a diagnostic tool, that relying on it alone turns "wait to fail" into "watch the student fail," that it leaves the question of cause unanswered, and that it may overlook the high-ability student who struggles relative

to his own ability rather than relative to his peers (Reynolds & Shaywitz, 2009). This criticism does not cancel the model's value for early intervention, but it reminds us that nonresponse is one piece of evidence, to be combined with others in a comprehensive evaluation.

## 6.4 Patterns of strengths and weaknesses

The **patterns of strengths and weaknesses** model (**Patterns of strengths and weaknesses, PSW**) seeks to restore to diagnosis a question that response to intervention leaves aside: **why** is the student struggling? It holds that a learning disability is linked to a weakness in a specific cognitive process, such as phonological processing, the ability to perceive, remember and manipulate speech sounds, or working memory, while other abilities remain intact. It therefore looks for a pattern with three corners (Flanagan et al., 2018):

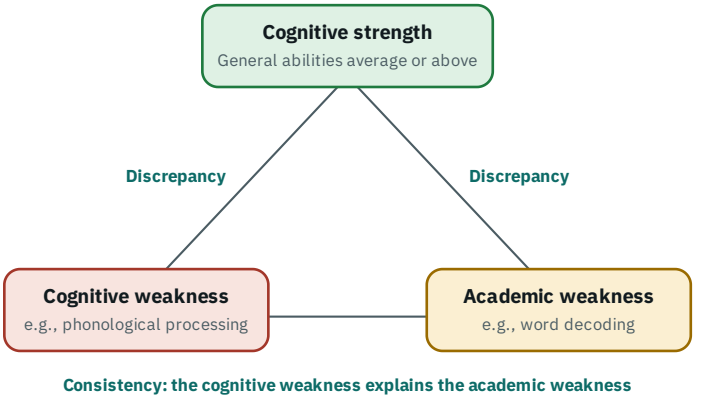



**Figure 6.4:** The pattern sought by the patterns of strengths and weaknesses model: a discrepancy between the cognitive strength and the two weaknesses, and consistency between the cognitive weakness and the academic weakness theoretically linked to it.

Under this model, a student is classified as having a learning disability in reading, for example, if his general abilities are good, he has a weakness in phonological processing, and his decoding achievement is low, because decoding depends on phonological processing. If, on the other hand, he has a weakness in processing speed and a weakness in reading comprehension, the theoretical link between the two is weaker and the pattern less convincing.

## Three versions of the model

The model has three well-known versions that share the logic of the triangle but differ in their theoretical framework and in how "strength" and "weakness" are established. Table 6.4 compares them:

**Table 6.4: The three versions of the patterns of strengths and weaknesses model.**

| Format | Developers | Theoretical framework | How the pattern is established |
|---|---|---|---|
| **Dual discrepancy/consistency (DD/C)** | Flanagan and colleagues (Flanagan et al., 2018) | Cattell-Horn-Carroll theory of abilities and cross-battery assessment | A weakness in a cognitive ability and an academic weakness, both below average, with general ability "average or above" estimated from the intact abilities, and a theoretical consistency between the two weaknesses |
| **Concordance-discordance (C-DM)** | Hale and Fiorello (Hale & Fiorello, 2004) | School neuropsychology, with no commitment to a particular battery | Significance tests of differences based on the standard error of the difference: a significant discordance between the cognitive strength and each of the two weaknesses, and concordance between the cognitive weakness and the academic weakness, that is, a nonsignificant difference between them |
| **Discrepancy/consistency (D/CM)** | Naglieri (Naglieri, 1999) | PASS theory of four processes: planning, attention, simultaneous processing and successive processing | A significant discrepancy between the highest and lowest processes, with the weak process below average; consistency between the weak process and the weak achievement; and a discrepancy between the strong process and the weak achievement |

### The debate over the model

Proponents argue that the model is consistent with the original concept of learning disability as a disorder in basic psychological processes, and that it offers an explanation of the problem that may guide intervention. Studies that have examined it, however, have found important problems. In a simulation study of the three versions, they were good at ruling out students without a disability, that is, highly specific, but weak at identifying students who actually had one; many of those they identified did not carry the "true" pattern on which the simulation was built, and they identified only small proportions of students (Stuebing et al., 2012). An applied study found that the model's methods often disagree about who has a disability and who does not, and that their identification rates are low and inconsistent (Miciak et al., 2014). Nor has it yet been shown that tailoring intervention to a cognitive profile improves outcomes more than direct academic intervention. Part of these problems stems from what we saw in the discrepancy model: the pattern rests on differences between scores, and difference scores are less reliable than the scores themselves.

The model's developers respond that these studies treat it as a mechanical algorithm, whereas, as proposed, it is a framework for organizing evidence within a comprehensive evaluation, one that draws data from multiple sources and methods, including response-to-intervention data, and that is judged through clinical reasoning rather than numerical comparisons alone (Flanagan et al., 2018). They hold that understanding the cognitive weakness helps in choosing accommodations and explaining nonresponse, even if it has not yet been shown to improve the design of the intervention itself. The model therefore remains the subject of genuine scientific debate to this day.

## 6.5 Hybrid models

After two decades of debate, many researchers lean toward a **hybrid model** (**Hybrid model**) that combines the best of the earlier models and drops what has proven weak. One of its best-known versions is the model proposed by Fletcher and colleagues (Fletcher et al., 2019), which rests on three criteria that must all be met:

**Inadequate response**

The student has not improved sufficiently despite good, intensive instruction, with documented progress-monitoring data.

**Low achievement**

Performance in a specific academic area is clearly low on a standardized achievement measure.

**Exclusion**

The weakness is not explained by intellectual or sensory disability, emotional disturbance, lack of opportunity to learn, or limited proficiency in the language of instruction.

Note that this model requires neither an intelligence test to calculate a discrepancy nor an analysis of cognitive processes. It may, however, use an intelligence test for another purpose: to rule out intellectual disability and to gain a fuller understanding of the student (Chapter 8 examines the application of this model to learning disabilities in detail).

The hybrid model is very close to the DSM criteria for specific learning disorder, which makes it a bridge between medical diagnosis and educational eligibility. The DSM requires that the difficulties persist for at least six months despite interventions targeting them (Criterion A), which corresponds to nonresponse. It requires that academic skills be substantially and quantifiably below those expected for the student's age, as confirmed by standardized achievement measures and a comprehensive clinical assessment (Criterion B), and for this it suggests a threshold of about 1.5 standard deviations below the mean, with leniency down to one standard deviation when other evidence supports it. It requires that the difficulties begin during the school years (Criterion C). And it requires that they not be better explained by intellectual disability, uncorrected visual or auditory acuity, another mental or neurological disorder, psychosocial adversity, limited proficiency in the language of instruction, or inadequate instruction (Criterion D) (American Psychiatric Association, 2022). Note that the DSM compares achievement with what is expected for age; it does not require a discrepancy from intelligence.

## A worked case using the hybrid model

### Nayef, 9 years 8 months, fourth grade

All three criteria met

**Background:** Since first grade, Nayef's teacher had noticed that he sounded out words slowly, guessed words from their first letters, and avoided reading aloud. He understands what is read to him well and takes part in oral discussion. In the universal screening at the start of fourth grade, his reading fluency was at the bottom of his grade, so the school began tiered support with him, with weak decoding as the initial hypothesis.

**Step one: tier data.** The team set its decision criterion in advance: dual discrepancy, meaning that both his level and his slope fall one standard deviation or more below his grade's mean after two tiers of support.

**Table 6.5: Nayef's curriculum-based measurement data across the three tiers (words correct per minute; hypothetical figures).**

| Level | Duration and support provided | Start | End | Weekly slope | Comparison with grade |
|---|---|---|---|---|---|
| **Tier 1** | Regular classroom instruction; screening in week 1 and a confirmatory measurement two weeks later | 38 | 40 | — | Grade median 92; Nayef ranked sixth from the bottom among the school's 100 fourth graders |
| **Tier 2** | 12 weeks, 4 sessions a week of 30 minutes each, in a group of 4 students, using an explicit decoding program; weekly measurement | 40 | 46 | 0.5 | Grade slope over the same period about 1.0 |
| **Tier 3** | 14 weeks, a daily session of 30–40 minutes, individually or with one other student, using an intensive decoding and fluency program; weekly measurement | 46 | 55 | 0.64 | Grade median at the end of the period 120, SD about 25; grade slope 1.0, SD about 0.3 |

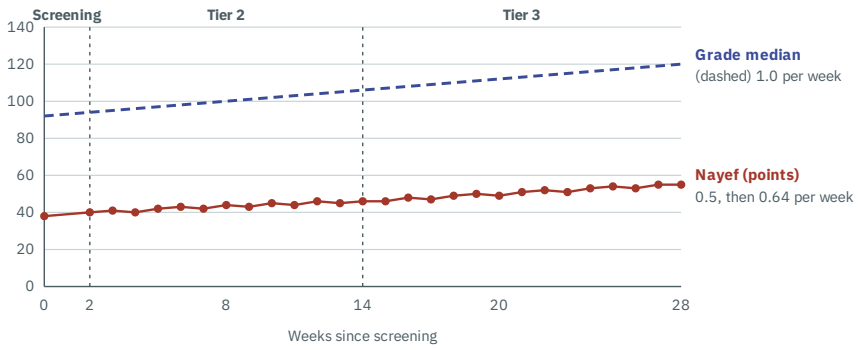



**Figure 6.5:** Nayef's reading fluency across the three tiers (hypothetical data). His slope rose slightly in Tier 3 but remained below his grade's slope, so the gap widened instead of narrowing.

**Reading the data:** Nayef's final level is 55, against 120 for his grade, or about (55 − 120) ÷ 25 ≈ −2.6 standard deviations. His Tier 3 slope is 0.64, against 1.0, or about (0.64 − 1.0) ÷ 0.3 ≈ −1.2 standard deviations. The dual discrepancy is therefore present. Before treating it as evidence about the student, the team checked treatment fidelity: 86% of the planned sessions were delivered, Nayef missed 4% of them, and two observations by the supervisor using the checklist showed 90% adherence to the program's steps. The intervention was thus delivered largely as planned. And because a slope calculated from 14 weeks remains an imperfect estimate, note that the judgment does not hinge on its exact value: even if his true slope were two-tenths higher, it would still fall below the grade's slope, and his level would still be far from the grade's.

**Step two: standardized achievement and exclusion.** With the family's written consent, a comprehensive evaluation was conducted:

**Table 6.6: Results of Nayef's comprehensive evaluation (hypothetical standard scores with a mean of 100 and SD of 15; 95% confidence intervals).**

| Source | Result | What it means |
|---|---|---|
| **Standardized word-reading test (reliability 0.95)** | 76 (69–83) | About 1.6 standard deviations below the mean, with the entire confidence interval below 85 |
| **Standardized reading-comprehension test** | 84 (77–91) | Higher than word reading, and affected by weak decoding in longer texts |
| **Listening comprehension** | 101 | He understands language when it is read to him, so the problem lies in decoding, not in language |
| **Individual intelligence test, for exclusion rather than to calculate a discrepancy** | Full Scale IQ 97 (91–103) | Does not indicate intellectual disability |
| **Vision and hearing screening this year** | Both normal | Sensory problems do not explain his difficulty |
| **School record and family interview** | Arabic is the home language; regular attendance since first grade; reading difficulty noticed since first grade | Not explained by weakness in the language of instruction or lack of opportunity to learn |

| Source | Result | What it means |
|---|---|---|
| **Behavior rating scale completed by the teacher and the mother** | Within the normal range, with clear avoidance of reading tasks | No emotional disturbance explains the picture; the avoidance is most likely a result of the difficulty, not its cause |

### Step three: matching the evidence to the criteria.

**Table 6.7: Nayef's evidence against the criteria of the hybrid model, the DSM and the Saudi regulatory text.**

| Issue | Hybrid model | DSM (DSM-5-TR) | Saudi Regulations |
|---|---|---|---|
| — | ✓ Dual discrepancy after two tiers, with acceptable treatment fidelity | ✓ Criterion A: 26 weeks, about six months of targeted intervention | ✓ Evidence that regular services are "of limited effectiveness" documented with data |
| **Achievement** | ✓ 76 in word reading | ✓ Criterion B: 76, below the 1.5 SD cutoff (78) | ? A "clear discrepancy" between ability (97) and achievement (76), and the text does not specify its size (see The lesson) |
| **Start** | — | ✓ Criterion C: since first grade | — |
| **Exclusion** | ✓ | ✓ Criterion D | ✓ No mental or sensory impairment and no behavioral disorder |

| Nonresponse ✓ | Achievement ✓ | Exclusion ✓ |
|---|---|---|
| Documented dual discrepancy | 76 (69–83) | No other explanation |

**Decision:** Nayef is identified as a student with a learning disability in reading, specifically in decoding and fluency, and he meets the criteria for specific learning disorder with impairment in reading. Continued intensive individual intervention is recommended within an individualized education program, with measurable goals in decoding and fluency, along with accommodations in other subjects, such as reading questions aloud to him and allowing extra time, so that his achievement in science, for example, does not measure his reading instead of his understanding. His slope should be monitored every two weeks and the plan reviewed after one term.

**The lesson**

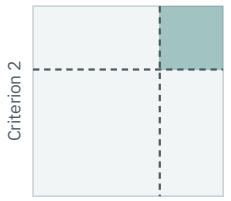
Had the team relied on the simple difference with a cutoff of 1.5 standard deviations, Nayef would not have been identified, because the difference between his intelligence (97) and his reading (76) is 21 points, below 22.5. Under regression, however, his expected achievement is about 98, and his reading falls about 22 points below it, exceeding the cutoff of 18. The judgment about the same student changed with the formula, even though his need did not. This is precisely what the hybrid model avoids: it bases the decision on what no one disputes, namely clearly low achievement, documented nonresponse, and the exclusion of other explanations.

**What evidence could change the decision?** If a retest of word reading with another standardized instrument came out close to average, or if his slope accelerated clearly, approaching his grade's slope, when measured with passages of established equivalence, the team would need to revisit the identification. And if a weakness in listening comprehension or in hearing emerged on re-examination, another explanation for his difficulty would have to be considered.

## 6.6 Multiple-criteria models in giftedness

In most systems it is no longer acceptable to identify a gifted student by a single intelligence score. In contemporary theories, giftedness encompasses intellectual ability, creativity, motivation and achievement in a particular domain, and it interacts with the environment and with opportunity (Gagné, 2009; Renzulli, 1978). Identifying gifted students therefore draws on multiple sources of evidence: an ability test, an achievement test, a teacher rating scale, a portfolio and perhaps a creativity test.

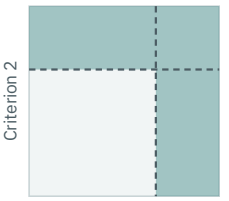
Gathering multiple sources of evidence, however, raises a question that is often overlooked: **how do we combine them into a decision?** An important study showed that the method of combination, known as the **combination rule** (**Combination rule**), by itself changes how many students are identified as gifted, and which students they are, even when the instruments and scores are identical (McBee et al., 2014). There are three main rules:




**The "and" rule**

Conjunctive rule

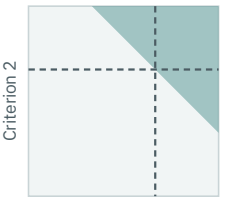
The student must exceed the cutoff on every criterion. It identifies few students and excludes anyone with a weakness in a single area, such as a gifted student with a learning disability.




**The "or" rule**

Disjunctive rule

Exceeding the cutoff on one criterion is enough. It identifies more students and opens the door to diverse profiles of giftedness, but it is sensitive to measurement error on each criterion.




**The averaging rule**

Compensatory rule

Scores are combined into an average that is compared with a cutoff. Strength in one area offsets weakness in another; this rule sits between the other two.

**Figure 6.6:** Three rules for combining two criteria. Each square represents all students, the shaded area represents those identified as gifted under each rule, and the two dashed lines mark the cutoff on each criterion.

## The three rules in numbers

Let us apply the rules to seven hypothetical students, each with two standard scores, one on an ability test and one on an achievement test, with a cutoff of 120 on each. Under the "and" rule both scores must reach 120; under the "or" rule one is enough; under the averaging rule their mean must reach 120.

**Table 6.8: Applying the three combination rules to seven hypothetical students (cutoff 120).**

| Student | Ability | Achievement | Mean | "And" rule | "Or" rule | Averaging rule |
|---|---|---|---|---|---|---|
| **A** | 128 | 126 | 127 | ✓ | ✓ | ✓ |
| **B** | 135 | 108 | 121.5 | ✗ | ✓ | ✓ |
| **C** | 121 | 119 | 120 | ✗ | ✓ | ✓ |
| **D** | 118 | 117 | 117.5 | ✗ | ✗ | ✗ |
| **E** | 140 | 95 | 117.5 | ✗ | ✓ | ✗ |
| **F** | 110 | 131 | 120.5 | ✗ | ✓ | ✓ |
| **G** | 119 | 122 | 120.5 | ✗ | ✓ | ✓ |

| Student | Ability | Achievement | Mean | “And” rule | “Or” rule | Averaging rule |
|---|---|---|---|---|---|---|
| **Number identified** | — | — | — | **1** | **6** | **5** |

The scores are the same, yet the number identified ranges from one student to six. Three observations in the table deserve attention. First, Student E has an ability score of 140 and an achievement score of 95, a pattern that may point to twice-exceptionality or to limited opportunities to learn; this student is identified only under the “or” rule, because the averaging rule lets the weakness consume the strength. Second, Students C and D are very close: the difference between 121 and 118 is too small to be trusted, since the standard error of a test with a reliability of 0.95 is about 3.4 points. Even so, one of them is identified under two rules and the other under none. Third, each rule errs in its own way. The “or” rule gives measurement error on every criterion a chance to lift a student above the cutoff, increasing false identifications, while the “and” rule gives it a chance to push a student below the cutoff on any criterion, increasing false exclusions (McBee et al., 2014). Choosing a rule, technical as it may appear, is therefore a decision about which of the two errors a program is willing to accept, and it should follow from the program's definition of giftedness and the services it provides.

> ⚠ **Point matrices**
>
> Some programs use a matrix that assigns points for each criterion and then adds them up. The method looks fair and comprehensive, but it merges scores that differ in nature and reliability into a single number, and the meaning of each is lost. A student may accumulate high points from criteria of weak reliability, while another who is truly exceptional on one strong criterion is excluded. It is better to specify the combination rule explicitly, based on the program's definition of giftedness.

## Universal screening and local norms

One of the greatest sources of inequity in gifted identification is reliance on nomination alone instead of **universal screening** (**Universal screening**), which is applied to all students. A student whom the teacher does not nominate, because the student is quiet, speaks a different home language, or has a difficulty that masks ability, never enters the assessment at all. A study in a large U.S. school district found that adding universal screening of all second graders markedly increased the representation of students from low-income and minority families in gifted programs (Card & Giuliano, 2016).

Likewise, comparing a student with national norms can exclude an entire school's students in a less advantaged setting, despite their clear superiority over their peers there. Researchers therefore propose using **local norms** (**Local norms**), which identify the highest-performing students in each school relative to their schoolmates, because these are the students who need services beyond what their classroom offers (Peters et al., 2019). (Chapter 12 addresses this in detail.)

## 6.7 Comparing the models and reading the Saudi regulations

**Table 6.9: Comparison of identification models by their central question, their evidence, and what they may miss.**

| Model | Central question | Core evidence | What it may miss |
|---|---|---|---|
| **Discrepancy** | Is achievement lower than ability predicts? | An intelligence test and an achievement test | Students with slightly low ability, and young children before the gap widens; the simple-difference method over-identifies students with high ability |
| — | Does the student learn when taught well? | Frequent progress-monitoring data across tiers of support | The reasons for nonresponse, and twice-exceptional learners (gifted students with learning disabilities); the judgment shifts with the nonresponse criterion chosen |
| — | Does a specific cognitive weakness explain the academic weakness? | A comprehensive cognitive battery and achievement tests | Many students, because its methods have low sensitivity and agree poorly with one another |
| **Hybrid** | Is achievement low despite good intervention, with no other explanation? | Response data, an achievement test, and an assessment for exclusionary factors | Depends on the quality of the intervention the school provides and the fidelity of its implementation |
| **Multiple criteria for giftedness** | Do indicators of giftedness appear across multiple sources? | Ability, achievement, ratings, and performance | Depends on the combination rule and on who reaches assessment |

**In the Saudi context: Which model do the regulations adopt?**

**Learning disabilities.** The Regulations for Special Education Institutes and Programs set three conditions for admitting a student to learning disabilities programs. The first is that "the student shows a clear discrepancy between the level of his abilities and the level of his academic achievement" in one area, such as reading, writing, counting, or mathematical reasoning, "or has a disorder in one of the psychological processes, such as memory, attention, thinking, and perception." The second is that the difficulties do not result from mental impairment, a behavioral disorder, or sensory causes. The third is "that it be established that regular educational services are unsuitable or of limited effectiveness" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 7). The Special Education Organizational Guide restates these conditions in nearly identical wording and makes no mention of response to intervention (Ministry of Education, 1436 AH [Arabic]). The model the texts prescribe is thus **discrepancy, exclusion, and the ineffectiveness of regular instruction**, with no numerical size for the discrepancy and no specified method for establishing ineffectiveness.

**Giftedness.** The published texts contain three different rules, which Table 6.10 presents through the concepts of this chapter:

**Table 6.10: Identification rules in published Saudi texts, read through the concepts of this chapter.**

| Text | What it requires | The rule in this chapter's terms |
|---|---|---|
| **The Regulations (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 8)** | High academic achievement over two semesters with an IQ of at least 115; or an IQ of at least 125 on the Wechsler or 127 on the Stanford-Binet; or outstanding skills and abilities attested by teachers and peers; or creative or innovative thinking | An "or" rule across four pathways, the first of which is a compound pathway under an "and" rule |
| **Guide to Gifted Classes (Ministry of Education, n.d. [Arabic])** | Students are nominated from among those classified in the national program; if the number is not filled, students must pass two or more of four criteria (the cognitive abilities test, personal characteristics, academic achievement, and creative product), one of which must be the cognitive abilities test with a score of at least 120 | A "two of four" rule with a mandatory ability condition, making it in effect an "and" rule between ability and any other criterion |

| Text | What it requires | The rule in this chapter's terms |
|---|---|---|
| **National Program for Gifted Identification (King Abdulaziz and His Companions Foundation for Giftedness and Creativity [Mawhiba], n.d. [Arabic])** | A score on the Mawhiba Multiple Cognitive Abilities Test, through voluntary registration initiated by the family, with a fee except for subsidized students | A single criterion with a cut-off score; voluntary registration is not the same as universal screening |

**A critical reading with the chapter's tools.** In learning disabilities, the absence of a numerical size for the discrepancy is a double-edged sword. It allows two teams to reach two different judgments about the same student, as in the chapter's opening scene, but it also frees the team from an arbitrary cutoff and allows it to estimate the discrepancy through regression and to take the standard error of the difference into account. The phrase "or has a disorder in one of the psychological processes" opens a door close to the logic of patterns of strengths and weaknesses, yet it does not require consistency between the cognitive weakness and the academic weakness, so the critique of that model applies to it with even greater force. The condition that regular services be shown to be ineffective is, in essence, the first criterion of the hybrid model. Two recent studies, however, indicate that practices resembling response to intervention exist in Saudi schools but are mostly informal and lack systematic progress monitoring, and that the absence of formal adoption of the model is among the obstacles to its implementation (Almulla & Mahama, 2026; Altheiabi et al., 2025). A team can therefore read the text in a hybrid way without departing from it, as in Nayef's case: it documents ineffectiveness with monitoring data and a criterion set in advance (Table 6.2), establishes low achievement with a standardized measure and its confidence interval, and satisfies the exclusion requirement.

In giftedness, the Guide to Gifted Classes makes the cognitive abilities score a mandatory condition at 120. This may exclude the twice-exceptional student whose difficulty lowers the ability score, as well as the student whose score reflects language more than ability, and it gives a student who scores 118 a different fate from one who scores 120, although the difference between them lies within measurement error. The different cutoffs in the Regulations for the Wechsler (125) and the Stanford-Binet (127) are consistent with the different standard deviations of the two scales in their editions at the time, so both cutoffs

fall at roughly the same distance from the mean. This is a sound application of the idea that a score can be understood only in relation to its scale. The national program, for its part, administers a standardized instrument to large numbers of students, but it depends on family registration, so a student's access to assessment hinges on the family's awareness and means; this is precisely the problem universal screening addresses. We found no officially published score scale for the test, and a recent peer-reviewed study reports an operational cut-off score of 1540 (Almulhim et al., 2026). Nor did we find a published regulatory text specifying how the four criteria are to be weighed when the scores diverge, or a text that permits or prohibits combining a gifted program with learning disabilities services.

*A question for reflection:* If the rule of the Guide to Gifted Classes were applied to Student E in Table 6.8, and the student passed the personal characteristics criterion, would the student be admitted? And if the ability score were 118 rather than 140, what would change? What would you propose as a rule that does justice to both students without opening the door to everyone who comes close to the cutoff?

## Conclusion

We conclude in this chapter that every identification model accepts some kind of error in exchange for what it offers. The discrepancy model is simple, but it delays help and fluctuates with measurement error. Response to intervention offers help early and rules out poor instruction as an explanation, but it does not explain why a student fails to respond. Patterns of strengths and weaknesses are high in specificity and low in sensitivity, and hybrid models combine nonresponse, low achievement, and exclusion. In giftedness, the combination rule determines who is identified and who is left out. It follows that a team should recognize the kind of error accepted by the model it works with, should not let the label “slow learner” become a verdict that closes the question of what the student needs, and should not allow response to intervention to become a reason for delaying assessment.

Some questions remain unsettled by research, such as the value of patterns of strengths and weaknesses for guiding intervention and the weak agreement among nonresponse criteria. Also open in the Saudi context is the fact that, for learning disabilities, the regulations prescribe discrepancy, exclusion, and the ineffectiveness of regular instruction without a numerical cutoff, leaving the cutoff to the team's discretion, and that, for giftedness, they use different combination rules, the most consequential being the cognitive abilities requirement of 120 for gifted classes.

## 6.8 Summary, key terms, and activities

### Chapter summary

- Learning disabilities and giftedness require a model that specifies the evidence, how it is combined, and where the cutoff lies, because no single test reveals either one.
- The discrepancy model is simple, but it does not distinguish groups that genuinely differ, it delays help, and it fluctuates with measurement error, because the standard error of the difference is larger than the error of each score. The simple-difference method over-identifies students with high ability and under-identifies those with slightly low ability; regression corrects this without solving the model's other problems.
- “Slow learner” is not a diagnostic category, and settling for it denies the student services without answering what the student needs.
- Response to intervention offers help early and rules out poor instruction as an explanation, but it does not explain why a student fails to respond; its criteria agree poorly, and it must not delay assessment.
- Patterns of strengths and weaknesses comprise three models that seek a cognitive explanation; they are high in specificity and low in sensitivity, and their value for intervention has yet to be demonstrated.
- Hybrid models combine nonresponse, low achievement, and exclusion, and they come close to the DSM criteria for specific learning disorder.
- In giftedness, the combination rule changes who is identified, and each rule accepts a particular kind of error; universal screening and local norms widen access.
- For learning disabilities, the Saudi regulations prescribe discrepancy, exclusion, and the ineffectiveness of regular instruction without a numerical cutoff; for giftedness, they use different combination rules, the most consequential being the cognitive abilities requirement of 120 for gifted classes.

## Glossary of terms

| | |
|---|---|
| **Discrepancy model** | Identifying a learning disability by a large gap between the ability score and the achievement score. |
| **Regression-based discrepancy** | Comparing actual achievement with the achievement predicted from the ability score by a regression equation, rather than with the ability score itself. |
| **Standard error of estimate** | The spread of actual scores around the scores predicted by a regression equation, used to judge how far achievement falls from what is expected. |
| **Borderline intellectual functioning** | Below-average intellectual functioning that lies above the range of intellectual disability; the DSM lists it as a condition that may be a focus of clinical attention rather than a disorder, and “slow learner” has no corresponding category in current classifications. |
| **Response to intervention (RTI)** | A model that provides tiered support and monitors the student's progress, using nonresponse as evidence of the need for a comprehensive evaluation. |
| **Multi-tiered system of supports (MTSS)** | A schoolwide framework for providing academic and behavioral support at increasing levels of intensity. |
| **Dual discrepancy** | A student's falling below peers in both level of performance and rate of growth. |
| **Nonresponse criterion** | The rule a team sets in advance for judging that a student has not responded to intervention, such as normative final status, slope, or dual discrepancy, together with a specified threshold. |
| **Treatment fidelity** | The extent to which an intervention is delivered as designed: the number and length of sessions and adherence to its steps. |
| **Patterns of strengths and weaknesses (PSW)** | A model that looks for a specific cognitive weakness explaining an academic weakness alongside good general abilities; it has three well-known versions. |

| | |
|---|---|
| **Hybrid model** | A model for identifying learning disabilities that combines inadequate response to intervention, low achievement, and the exclusion of other explanations. |
| **Meta-analysis** | A statistical method that pools the results of multiple studies on a single topic and analyzes them together to reach a more general conclusion. |
| **Combination rule** | The way scores on multiple criteria are combined into a single decision, including the "and" rule, the "or" rule, and the averaging rule. |
| **Universal screening** | Assessing all students in a grade or stage, rather than relying on nomination or family registration. |
| **Local norms** | Norms that compare a student with schoolmates in the same school or district rather than with a national sample. |

## Review questions

1. Calculate the discrepancy for a student with an IQ of 112 and a mathematics achievement score of 90 using the simple-difference method, and decide whether it reaches a cutoff of 1.5 standard deviations. Then calculate it using the regression method, assuming a correlation of 0.60 and a cutoff of 18 points below the predicted score. Does the judgment change? And what does it mean that the student is close to the cutoff under both methods?
2. Why is the discrepancy model described as a "wait-to-fail" model? And why does regression-based discrepancy not solve this problem?
3. True or false? Explain: "A student described as a slow learner does not need assessment, because the weakness is expected given the student's intelligence."
4. A student started below peers but is growing faster than they are. Does the student show a dual discrepancy? What do you decide?
5. A second-grade girl remained at the 20th percentile on a standardized test after Tier 2, but her slope equals that of her peers. Judge her response using the normative final status criterion, then the slope criterion, then dual discrepancy. What do you conclude from the differing judgments?
6. A family requested an evaluation of their son because they suspected dyslexia, and the school replied that he must first complete Tier 3 support. What is the position of the U.S.

Office of Special Education Programs memorandum on this? And how can the school reconcile the family's right with the continuation of support?

7. How might the response to intervention model miss a gifted student with a learning disability? And how might the patterns of strengths and weaknesses model miss the same student?
8. What does each of the three patterns of strengths and weaknesses models require? Why might these models be high in specificity and low in sensitivity? And what does this mean for a student who meets none of their patterns?
9. A gifted program uses the “and” rule between an intelligence test and an achievement test. Which students might it unfairly exclude? What is the alternative? Draw on Table 6.8.
10. A student scored 118 on the cognitive abilities test and passed the achievement, personal characteristics, and creative product criteria. What does the rule of the Guide to Gifted Classes decide for this student? And what does the standard error say about the difference between 118 and 120?
11. Why might local norms be fairer than national norms in determining who needs gifted services?

## Case for application: Lujain, age 13, Grade 7

**The story:** Lujain moved with her family from another region in the middle of sixth grade; this is her third school since fourth grade. Her teacher says she “understands the explanation but falls apart in arithmetic.” She counts on her fingers to add small numbers and takes a long time on multiplication problems. Her grades in other subjects are average. The school placed her in Tier 2 support in mathematics and then, at her family's request, referred her to the assessment team.

**Table 6.11: Evidence available on Lujain at referral (hypothetical data).**

| Source | Result |
|---|---|
| **Tier 2 in mathematics** | 10 weeks, three sessions a week; 55% of planned sessions were delivered because of scheduling conflicts; no session was observed with a checklist |
| **Curriculum-based measurement in computation (correct digits in two minutes)** | From 18 to 22, a slope of 0.4 per week; class slope about 0.8 |
| **Standardized test of arithmetic operations** | 81 (75–87) |

| Source | Result |
|---|---|
| **Standardized test of mathematical reasoning** | 88 (81–95) |
| **Individual intelligence test** | Full Scale IQ 92 (86–98) |
| **Vision and hearing screening** | Last recorded screening in first grade |
| **School record** | Absent about 20% of school days last semester because of the move; no data from the two previous schools |

**Tasks:** (a) Apply the three criteria of the hybrid model to the evidence, and for each criterion state whether it is met, not met, or unresolved, with your reasoning. (b) Match the evidence against the DSM criteria for specific learning disorder, especially Criteria A and D. (c) Does the evidence meet the conditions of the Saudi regulations? (d) What is your decision now? What three pieces of evidence would you request before the final decision, and how would you gather them without delaying Lujain's access to support?

**Table 6.12: Mapping chapter objectives to questions and activities.**

| Objective | Questions | Activities |
|---|---|---|
| **1 Calculating discrepancy by both methods and the effect of the standard error of the difference** | 1, 2, 3 | — |
| **2 Applying nonresponse criteria and their limits** | 4, 5, 6, 7 | 2 |
| **3 Comparing the patterns of strengths and weaknesses models** | 7, 8 | 3 |
| **4 Applying the hybrid model and matching it to the DSM** | Lujain's case (a, b) | 2 |
| **5 Applying combination rules and determining whom each rule excludes** | 9, 11 | 1 |
| **6 Analyzing Saudi regulatory texts** | 10, Lujain's case (c) | 4 |
| **7 Judging the sufficiency of evidence in a new case** | Lujain's case (d) | 4 |

## Learning activities

For any activity that uses real student data: it is conducted with the consent of the family and the school administration, without any data that identify the student, and under professional supervision; no decision is based on its results, and they are communicated only through official channels. All the activities can be carried out with hypothetical data.

1. **Simulating combination rules** (about one hour): In a spreadsheet, create scores for twenty hypothetical students on two criteria and identify the gifted under each of the three rules. Then add a small random error (between −4 and +4) to each score and repeat the identification. *Product:* a table showing how many students each rule identified before and after the error, and a paragraph explaining which rules were most affected by the error and whose status changed.
2. **Designing tiers of support** (about two hours): Design a response to intervention plan in reading for first grade in an elementary school you know or a hypothetical one. What happens at each tier? Who delivers it? How often is progress measured? What nonresponse criterion and threshold will you use, chosen from Table 6.2? How will implementation fidelity be documented? And when is a case referred for comprehensive evaluation without waiting? *Product:* a two-page outline.
3. **Debate** (one class session): The group splits into two teams, one defending the response to intervention model and the other the patterns of strengths and weaknesses model, drawing on the evidence in this chapter, including Table 6.4 and the critique by Reynolds and Shaywitz. *Product:* a one-page paper from each team summarizing its two strongest arguments and the other team's strongest argument.
4. **From text to procedure** (desk activity, about two hours): Read the text of Article 7 of the Regulations as presented in this chapter, and write a proposed procedure for a school team specifying how a “clear discrepancy” is to be estimated, how regular services are to be shown to be “of limited effectiveness,” the minimum duration of support and the data it requires, and how the exclusion requirement is to be met. Then apply your procedure to Lujain's case. *Product:* a one-page procedure and a paragraph on the result of applying it.

## Chapter references

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Almulhim, N., Aljughaiman, A. M., & Alfaiz, F. S. (2026). Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis. *Journal of Intelligence, 14*(8), Article 166. https://doi.org/10.3390/jintelligence14080166

Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications, 13*(1), Article 974. https://doi.org/10.1057/s41599-026-07376-9

Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties, 30*(1), 75–92. https://doi.org/10.1080/19404158.2025.2473425

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18. https://doi.org/10.1016/j.jsp.2012.09.004

Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences, 18*(3), 296–307.

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678–13683.

Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., McDonough, E. M., & Ortiz, S. O. (2018). Dual discrepancy/consistency operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed.). Guilford.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). Guilford.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99.

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice, 18*(3), 172–186.

Gagné, F. (2009). Building gifts into talents: Detailed overview of the DMGT 2.0. In B. MacFarlane & T. Stambaugh (Eds.), *Leading change in gifted education: The festschrift of Dr. Joyce VanTassel-Baska* (pp. 61–80). Prufrock Press.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly, 58*(1), 69–89.

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29*(1), 21–37.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. Wiley.

Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open, 5*(2).

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180–184.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly, 24*(2), 130–145.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41*(1), 3–22.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39*(2), 469–518.

U.S. Department of Education. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final rule (34 CFR Parts 300 and 301). *Federal Register, 71*(156).

U.S. Department of Education, Office of Special Education Programs. (2011). *A response to intervention (RTI) process cannot be used to delay-deny an evaluation for eligibility under IDEA* (OSEP Memo 11-07).

Part Three

# Categories

From indicators to judgment


Chapter 7: Intellectual disability

Chapter 8: Learning disabilities

Chapter 9: Autism spectrum disorder

Chapter 10: Attention-deficit/hyperactivity disorder and behavioral and emotional disorders

Chapter 11: Sensory and multiple disabilities and communication disorders

Chapter 12: Giftedness

Chapter 13: Twice-exceptional learners

# 7 Chapter 7 Intellectual disability

An IQ score cannot carry the weight of diagnosis on its own. Intellectual disability is a judgment that rests on three pillars: a clear deficit in intellectual ability, deficits in the skills of everyday life, and onset of both during the developmental period. In this chapter we trace how each pillar is measured and how the evidence for all three is brought together, and we explain why the number 70 is not a hard dividing line.

**Scene from the first grade**

Halfway through the school year, Nawaf still cannot count to ten without a mistake, and he does not recognize most of the letters. He speaks in short sentences of two or three words and needs help opening his lunch box and closing his backpack. At recess he plays with the kindergarten children rather than with his classmates. When the teacher explains the rules of a new game, he looks at her with a smile and then does whatever the children around him are doing.

His mother says he “was a little late with everything”: he walked at eighteen months and began talking after the age of three. She asks the teacher, “Will he catch up with his classmates if we work hard with him?”

The mother's question is a fair one, but answering it requires more than the teacher's observations. Nawaf's delay may reflect an intellectual disability, but it may equally be a language delay, a hearing impairment, or the effect of limited opportunities. In this chapter we set out the path by which the team distinguishes among these possibilities.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Determine, for a presented case, the extent to which the three pillars of diagnosis are met, and classify examples of adaptive deficits into the conceptual, social and practical domains.
2. Determine the severity level for a presented case from its adaptive behavior data, and compare this with the approach of the International Classification of Diseases (ICD-11).
3. Select a screening tool, an intelligence test or an adaptive behavior scale for a presented case, and justify the choice in terms of the student's age, the test's floor, the recency of its norms and the status of its Arabic version.
4. Apply the eligibility criteria of the Saudi regulations to a presented case, and show where they agree with the DSM and where they depart from it.
5. Judge whether the evidence suffices for a diagnosis in a new case with scores near the thresholds, taking measurement error and the age of the norms into account, and give reasons for the judgment.
6. Distinguish intellectual disability from similar and co-occurring conditions, and determine when to recommend medical and genetic referral.

# Introduction

Intellectual disability is among the categories most prone to having the judgment reduced to a single number, the IQ score, even though a sound diagnosis requires three criteria to be met together: a clear deficit in intellectual ability, deficits in adaptive behavior, and onset during the developmental period. The judgment has far-reaching consequences for a student's educational and social path, and errors are costly in both directions: the label may be attached to a student whose performance is better explained by weak hearing, language or opportunities, or it may be withheld from a student who needs the support it opens up. Sound judgment also requires that the IQ threshold be read in light of the confidence interval, as an approximate and flexible range.

We open the chapter by defining intellectual disability and its causes (Section 7.1), then present the three diagnostic criteria (Section 7.2), severity levels and support needs (Section 7.3), and early warning signs (Section 7.4). We then turn to the decision pathway and the place of genetic testing within it (Section 7.5), instrument profiles and the Arabic tools available (Section 7.6), and apply all of this to four cases (Section 7.7). From there we move to differential diagnosis (Section 7.8) and co-occurrence (Section 7.9), and we close with eligibility criteria in the Saudi context (Section 7.10).

## 7.1 What is intellectual disability?

Intellectual disability is a developmental condition marked by a clear deficit in intellectual ability, that is, in reasoning, problem solving, planning, abstract thinking and learning from experience, accompanied by a clear deficit in adaptive behavior, that is, in the skills a person needs in everyday life, with onset during the developmental period. A broad review estimates its prevalence at about 1% of the population (Maulik et al., 2011).

Its name has changed over the decades. It was once called "mental retardation," a term later replaced by "intellectual disability" in most manuals and legislation, both because the older term had become stigmatizing and because the newer one fits a view of disability as an interaction between the individual and the environment. The eleventh revision of the International Classification of Diseases (**ICD-11**) calls it "disorders of intellectual development" (**Disorders of intellectual development**) (World Health Organization [WHO], 2024). In Saudi documents, the Regulations for Special Education Institutes and Programs use the term "mental retardation" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), while the Special Education Organizational Guide uses "mental impairment" (Ministry of Education, 1436 AH [Arabic]). A teacher may therefore come across any of these names in student files; they all refer to the same thing.

## What causes it?

Intellectual disability has many causes. The cause used to remain unknown in a large share of cases, especially mild ones, but that share has shrunk as modern genetic testing has spread (see 7.5). Known causes include:

- **Genetic and chromosomal causes:** such as Down syndrome, fragile X syndrome (**Fragile X**), and recessive genetic disorders, including metabolic disorders such as phenylketonuria (**PKU**), whose effects can be prevented through early newborn screening and diet.
- **Prenatal causes:** such as fetal exposure to alcohol, to certain infections, or to severe malnutrition.
- **Perinatal and postnatal causes:** such as severe oxygen deprivation, extreme prematurity, brain infections, head injuries, lead poisoning and severe environmental deprivation.

**Common misconceptions**

- **"Intellectual disability means the child cannot learn."** Most children with intellectual disability, the majority of whom are at the mild level, learn to read, write and use functional arithmetic, and acquire vocational skills, but at a slower pace and with instruction that is more explicit and more repetitive.
- **"The IQ score determines everything."** Two children with the same IQ score may differ widely in their independence and social skills. That is why severity is determined by adaptive behavior, as we will see.
- **"Intellectual disability does not change."** Intellectual ability is relatively stable, but functioning improves considerably with appropriate support, and a person may need less support over time.
- **"Knowing the cause is of no use to the school."** The cause does not change the diagnostic criteria, but it may reveal health risks that require monitoring, indicate the likelihood of recurrence in the family, and in some metabolic disorders point to a specific treatment.

## 7.2 Diagnostic criteria

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**), and the American Association on Intellectual and Developmental Disabilities (**AAIDD**), in the twelfth edition of its manual, agree on three criteria that must all be met (American Psychiatric Association [APA], 2022; Schalock et al., 2021):

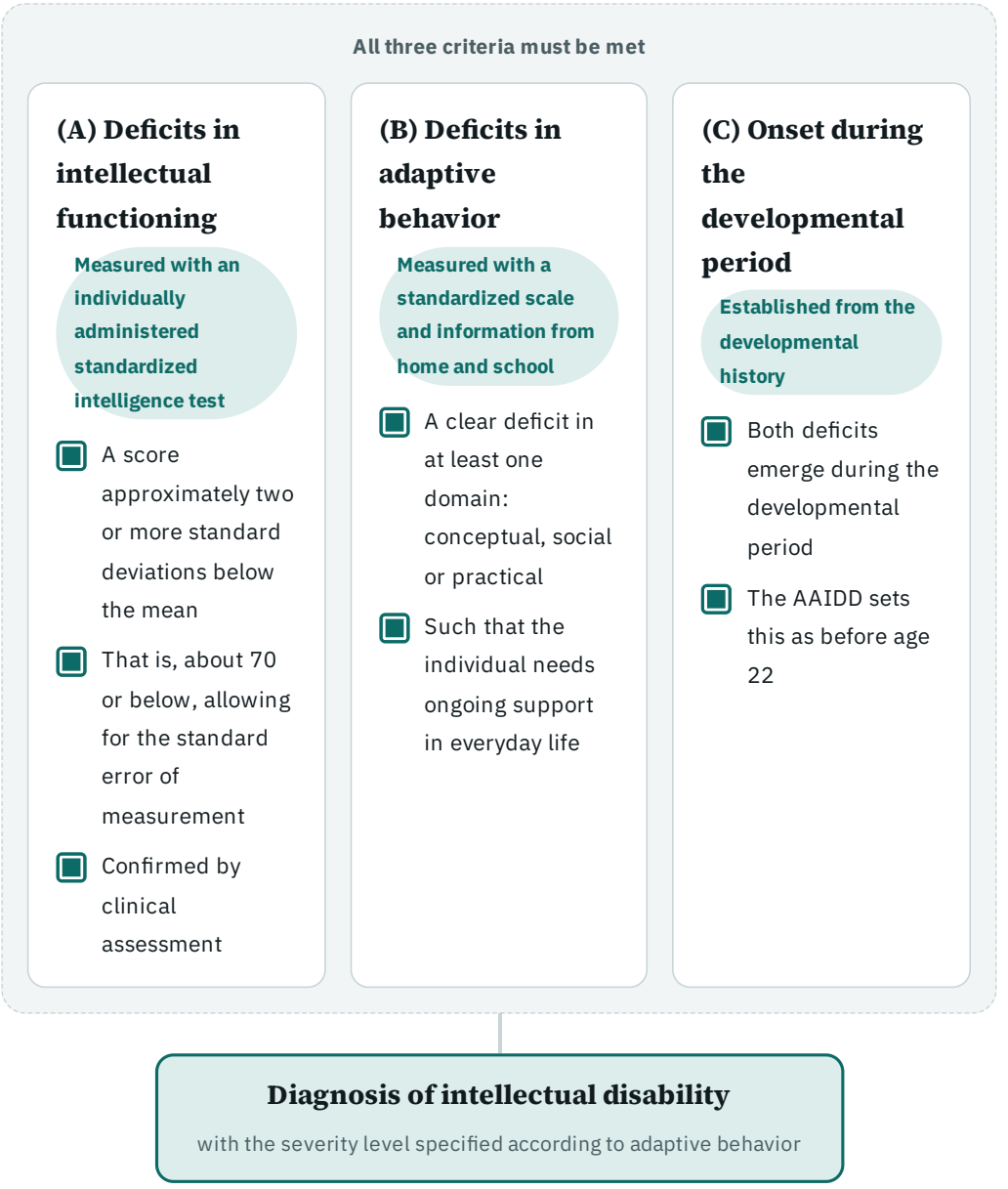



**Figure 7.1:** The three criteria for diagnosing intellectual disability. If any criterion is absent, the diagnosis cannot be made, whatever the IQ score.

## Criterion A: Intellectual ability

This criterion is measured with an individually administered standardized intelligence test of high reliability (Chapter 3 presents these tests in detail). A deficit means that the **Full Scale IQ** (**Full Scale IQ, FSIQ**) falls about two standard deviations below the mean, that is, around 70 on a test with a mean of 100 and a standard deviation of 15. Yet both manuals stress that this number is **not a hard cutoff**. Every score carries some measurement error and is therefore read within the confidence interval that surrounds it (see 2.5). A score of 72, for example, may have a confidence interval extending from 65 to 79. For this reason, a person whose score falls between 70 and 75 may be diagnosed if the adaptive deficits are very clear.

The test must also suit the student: in its language, in the recency of its norms, and in how it accommodates any other disabilities. A low score on a verbal test taken by a child who has not mastered the language of the test does not measure intellectual ability. Outdated norms in particular inflate scores near the threshold, a point we return to in Maryam's case (7.7).

## Criterion B: Adaptive behavior

This is the criterion that keeps diagnosis from turning into a verdict on a number. The question it asks is whether the student can manage everyday life as the community expects of someone that age. It is measured in three domains: conceptual (such as language, reading, arithmetic, money and time), social (such as getting along with others and understanding the rules of interaction), and practical (such as self-care, getting around and safety):

**Table 7.1: Examples of adaptive behavior deficits in the three domains.**

| Area | Examples of deficits in an elementary school student |
|---|---|
| **Conceptual** | Does not know the value of simple coins and notes, cannot tell time, cannot follow three-step instructions, and does not plan for what is needed in the next lesson. |
| **Social** | Does not understand the rules of group games, believes everything he is told and is therefore easily exploited, does not read classmates' feelings, and behaves with strangers as he does with family. |
| **Practical** | Needs help with dressing or personal hygiene, cannot find the way to the classroom without someone accompanying him, and does not behave safely when crossing the street. |

A clear deficit in one of the three domains is enough, because a deficit in any of them hinders independent living. Adaptive behavior is measured with a standardized scale, and information should preferably be gathered from more than one source, such as the family and the teacher.

## Criterion C: Onset during the developmental period

This criterion distinguishes intellectual disability from a decline in abilities after development is complete, as in brain injury in adults or dementia. What is required is that the deficits appear during childhood or adolescence, not that they were diagnosed at the time. An adult who was never assessed as a child can be diagnosed if the developmental history shows that the deficits were present from childhood. In its twelfth edition, the AAIDD raised the age of onset from 18 to before 22, on the grounds that brain development continues into the early twenties (Schalock et al., 2021). The DSM, for its part, simply states “during the developmental period” without specifying an age (APA, 2022).

**Global developmental delay**

In the first five years of life, intelligence is hard to measure reliably. The DSM therefore provides a provisional diagnosis, **global developmental delay** (**Global developmental delay**), for a child under five who shows delays in two or more developmental domains and cannot be fully assessed. The child is reassessed later (Shevell et al., 2003). Some of these children outgrow their delay, in which case the later diagnosis is something other than intellectual disability.

## 7.3 Severity levels and support needs

One of the most notable changes introduced in the fifth edition of the DSM is that severity is now determined by adaptive behavior. Earlier manuals divided severity according to IQ ranges; today the question is how much support the person needs in everyday life. The DSM uses the phrase "severity levels" for intellectual disability and "support levels" for autism spectrum disorder. The former describes the depth of adaptive deficits in each domain; the latter describes how much support the symptoms call for.

**Table 7.2: Severity levels of intellectual disability by adaptive behavior domain (adapted from APA, 2022).**

| Level | Conceptual domain | Social domain | Practical domain |
|---|---|---|---|
| **Mild** | Difficulty learning reading, writing, arithmetic, time and money; needs support to meet age-related expectations | Less mature than peers in social interaction; finds subtle social situations hard to grasp; may be exploited | Manages self-care; needs support with complex tasks such as shopping, transportation and managing money |

| Level | Conceptual domain | Social domain | Practical domain |
|---|---|---|---|
| **Average** | Academic skills progress slowly and remain at a basic level that requires support for everyday use | Spoken language is much simpler than that of peers; may form relationships but needs support in reading social situations | Learns self-care after extended teaching; needs ongoing support with most independent tasks |
| **Severe** | Limited understanding of written language, numbers and time; needs substantial support with problem solving | Language is limited in vocabulary and grammar and centers on the here and now | Needs support with all activities of daily living and constant supervision |
| **Profound** | Deals with physical objects more than symbols; co-occurring sensory and motor problems may hinder performance | Very limited understanding of symbolic communication; usually expresses himself or herself through nonverbal means | Depends on others for all aspects of physical care, health and safety |

The mild level accounts for the majority of cases. The AAIDD calls for moving beyond classification by severity to determining **support needs intensity** in each area of life, using tools designed for this purpose, such as the Supports Intensity Scale–Children's Version (**SIS-C**) (Thompson et al., 2016). The more important educational question goes beyond “What is his level?” to “What support does he need, where, and how much?”

**Comparison with ICD-11**

The ICD agrees with the DSM on the three pillars but differs from it in three respects that bear on decisions (WHO, 2024):

- **Severity on both dimensions:** severity is determined by intellectual ability and adaptive behavior together. The mild level corresponds to performance roughly two to three standard deviations below the mean, and the moderate level to three to four. Severe and profound both lie more than four standard deviations below the mean and are distinguished by adaptive behavior alone, because standardized tests do not discriminate within this range.
- **Behavioral indicators:** when suitable standardized tests are unavailable or cannot be administered, clinical judgment draws on detailed behavioral indicators for each level and age, set out in the ICD's descriptive guidelines.
- **The "provisional" category (Provisional):** used for a child under four, or when a valid assessment is not possible because of a sensory or motor impairment or severe behavior. It corresponds to global developmental delay in the DSM, though with a different age limit.

## 7.4 Early warning signs

Severe intellectual disability usually becomes apparent in the first year of life and is detected early, especially when it is associated with a known syndrome. Mild intellectual disability, by contrast, may go undetected until school entry, when difficulty with academic learning emerges. Indicators that call for a developmental evaluation include:

- A clear delay spanning several developmental domains at once: motor skills, language, play and self-care.

- Play resembling that of much younger children, such as exploring objects by mouthing or throwing them at an age when pretend play is expected.
- Difficulty learning new skills, needing a great deal of repetition to learn a simple skill, and forgetting it quickly.
- At school: difficulty that extends to nearly all subjects, together with difficulties in independence and social skills.
- **Loss of previously acquired skills:** this sign is different from delay and calls for urgent medical referral before any educational evaluation.

A young child does not need a diagnosis in order to benefit from services. In the Kingdom, the Special Education Organizational Guide defines those eligible for early intervention services as children no older than six who show a delay "of two standard deviations below the mean in two or more domains" or of 25%, and it adds children "at risk" who have been exposed to at least three environmental risk factors (Ministry of Education, 1436 AH [Arabic]). The Law on the Rights of Persons with Disabilities provides for designing and implementing early intervention programs and for making modern assessment and diagnostic tools available (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic], Art. 8). A developmental screening tool in early childhood is thus a gateway to services before it is a gateway to a label.

**For the teacher: a difference worth noticing**

A student with learning disabilities usually struggles in one specific area and seems bright in others: he understands a story when it is read to him, solves problems orally, and behaves maturely with classmates. A student suspected of having an intellectual disability, by contrast, shows difficulties across several areas at once: academic, social and practical. This difference alone is not enough for a diagnosis, but it can guide your observations.

# 7.5 The decision pathway and genetic evaluation

**1 Initial indicators**

Delays in multiple developmental domains, or pervasive academic difficulty together with limited independence.

**2 Hearing and vision screening: a prerequisite for any tool**

No developmental scale or intelligence test is administered until hearing and vision have been confirmed as intact or corrected, because impairments in either lower scores and can mimic intellectual disability.

**3 Developmental screening and history**

A developmental screening tool for young children, such as the Ages and Stages Questionnaires (**ASQ-3**), and a review of the medical, developmental and family history and of newborn screening results.

Delay in more than one domain: referral

**4 Comprehensive evaluation**

An individual intelligence test suited to the student's age and language, an adaptive behavior scale from two informants, a developmental history interview, a language evaluation, and medical referral for first-tier genetic testing.

**5 Matching against the criteria**

Is intellectual ability clearly low, taking the confidence interval and the age of the norms into account? Is adaptive behavior low in at least one domain? Did onset occur during the developmental period? Have other explanations been ruled out?

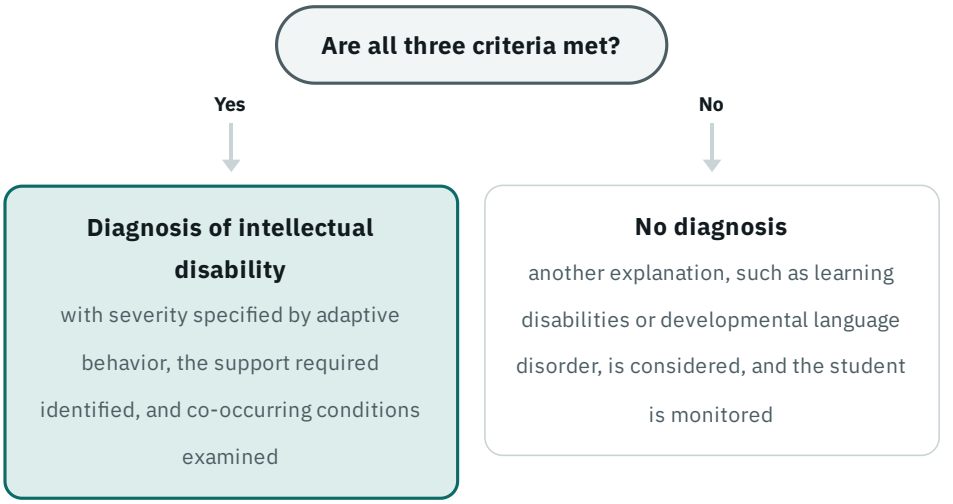



**Figure 7.2:** The decision pathway for diagnosing intellectual disability. Sensory screening precedes every tool. A student under five may receive a provisional diagnosis of global developmental delay and is then reassessed.

## Genetic testing: a recommended first-tier referral

"Medical and genetic evaluation as needed" was once a common phrase in diagnostic reports. The current recommendation, however, is that genetic testing be offered to every child with unexplained intellectual disability or global developmental delay. A consensus statement identified **chromosomal microarray** (**Chromosomal microarray**) as a first-tier test for individuals with developmental disabilities or congenital anomalies, in place of conventional karyotyping (Miller et al., 2010). The American College of Medical Genetics and Genomics later recommended that **exome sequencing** (**Exome sequencing**) or genome sequencing be considered as a first- or second-tier test for children with intellectual disability or developmental delay (Manickam et al., 2021).

The decision rests with the developmental pediatrician or the geneticist. The educational team's role is nonetheless clear: to recommend the referral in its report as a step endorsed by medical guidelines, and to document whatever in the family and developmental history supports it, such as parental consanguinity, relatives with similar conditions, and any loss of acquired

skills. The test result does not change the diagnostic criteria; intellectual disability is diagnosed on the three criteria even when its cause is unknown. But the result may reveal health risks that need monitoring, and it answers a question that weighs heavily on families: will this happen again in our other children?

**In the Saudi context: consanguineous marriage and national screening**

Consanguineous marriage is common in the Kingdom. A study of 3,212 Saudi families found that 57.7% were based on consanguineous marriage, 28.4% of them between first cousins, with variation across regions (el-Hazmi et al., 1995). A later study of a national sample of 11,554 mothers found a similar rate of 56% (El Mouzan et al., 2008). Consanguinity raises the likelihood of recessive genetic disorders, many of which cause intellectual disability.

Three relevant programs operate in the Kingdom. The **National Newborn Screening Program** began in 1426 AH (2005 CE), covering sixteen metabolic and endocrine diseases, later expanded to seventeen. Its early years detected roughly one positive case per thousand births, and the program's manual attributed this high rate to the fact that these diseases are recessive genetic conditions whose frequency rises with consanguineous marriage (Ministry of Health, 2016 [Arabic]). **Premarital screening** has been mandatory since 2004, but it is limited to sickle cell disease and beta thalassemia, and acting on its results remains voluntary (Memish & Saeedi, 2011). The Law on the Rights of Persons with Disabilities guarantees screening and diagnostic services and programs for the early detection of disability (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic], Art. 9).

The lesson for the evaluator is that premarital screening does not cover the hundreds of recessive disorders that can cause intellectual disability, and that a normal newborn screen does not rule out a genetic cause. Parental consanguinity is information that strengthens the case for genetic referral; the answer comes from the test itself.

**A question for reflection:** In a developmental history interview, how can you ask the family about parental consanguinity and about similar conditions among relatives without the question being heard as blame?

## 7.6 Instrument profiles

Here we focus on what is specific to using intelligence tests and adaptive behavior scales in intellectual disability (Chapter 3 presents them in general terms), and we add a developmental screening tool for young children. Each profile follows the same template: the idea behind the tool, what it actually looks like, a scene from administration, a table of its technical information, how to interpret its results, and its limitations. The psychometric figures in the tables are approximations drawn from the publishers' manuals, so readers should consult the manual itself before any professional use.

**Developmental screening**

Identifies young children who need a comprehensive developmental evaluation.

**ASQ-3**

**Intellectual ability**

An individual intelligence test whose range extends to very low scores.

**SB5 · WISC-V**

**Adaptive behavior**

Standardized scales completed by the family and the teacher.

**Vineland-3 · ABAS-3 · DABS**

### ASQ-3 Ages and Stages Questionnaires, Third Edition

Developmental screening

### The idea

A screening tool completed by parents, designed to be administered periodically during health care visits or in early childhood programs to identify children whose development may be delayed in one or more domains (Squires & Bricker, 2009).

### What it looks like

Twenty-one questionnaires, each for a specific age interval. The questionnaires cover intervals from 2 to 60 months, but their range of use is somewhat wider, from 1 to 66 months, because each questionnaire has an age window within which it can be given before and after its target age. Each questionnaire contains thirty questions about behaviors the child can perform, six in each of five domains: communication, gross motor, fine motor, problem solving and personal-social. Parents answer each question with "yes," "sometimes" or "not yet," and many questions ask them to try the skill with the child before answering.

### A scene from administration

At a primary health care center, the nurse gives the mother of a 30-month-old the questionnaire for the appropriate interval and asks her to try some of the tasks with him at home during the week, using simple household items, before answering. When the mother returns the questionnaire, the nurse totals the score for each domain and compares it with that domain's two cutoffs. The communication score falls below the lower cutoff, the problem-solving score lies in the monitoring zone, and the other domains are on track. The nurse avoids describing the child as "delayed" and explains to the mother that the questionnaire calls for a closer look: she refers the child first for a hearing test, then for a developmental and language evaluation, and gives the mother home activities for the domain that needs monitoring.

**Table 7.3: ASQ-3 at a glance.**

| | |
|---|---|
| **Purpose** | General developmental screening that precedes diagnosis and points toward it |
| **Ages** | Questionnaire intervals from 2 to 60 months; usable from 1 to 66 months |
| **Edition and year** | Third, 2009 |
| **Publisher** | Paul H. Brookes |
| **Qualification level** | Completed by parents; scored and interpreted by a professional trained in early childhood or health care; no psychological assessment license required |
| **Time** | 10 to 15 minutes to complete, and a few minutes to score |
| **Score types** | A raw score for each domain, from 0 to 60 |
| **Cutoff scores** | A lower cutoff for each domain set at two standard deviations below the norm sample mean, and a monitoring zone between one and two standard deviations |
| **Reliability** | The manual reports test–retest reliability and interrater agreement of approximately .90 |
| **Classification accuracy** | The manual reports sensitivity and specificity averaging about .85, varying across age intervals |
| **Year of norm collection** | In the years before publication, with a US sample of more than fifteen thousand children |
| **Arabic version** | We know of no published standardized Arabic version at the time of writing. Translation alone does not carry US cutoff scores over to children in a different setting |

### How are the results interpreted?

Each domain score is compared with two cutoffs: a score above the upper cutoff indicates development on track, a score between the two indicates a need for monitoring and home activities, and a score below the lower cutoff indicates a need for further evaluation. The questionnaire yields no total score; it is the pattern across domains that is interpreted.

### Limitations

The tool depends on the accuracy of the parents' observations and their understanding of the questions. A delay in a single domain does not mean intellectual disability; it may reflect only a language or motor delay, or an undetected hearing loss. Tool website

## SB5 Stanford-Binet Intelligence Scales, Fifth Edition

Intellectual ability

### The idea: why is it preferred for this population?

The Stanford-Binet extends down to very young ages, starting at two years, and contains many easy items at its lower end. This matters in intellectual disability, because a test whose items start out difficult produces what is called a **floor effect** (**Floor effect**): the student fails every available item, so the examiner cannot tell where the student's ability actually lies, and the subtest scores all look alike (see 2.5.3).

### What it looks like

The examiner begins with two short subtests called **routing subtests** (**Routing subtests**), one verbal and one nonverbal. Their results determine the level at which the examiner starts the remaining subtests. The scale then measures five factors, fluid reasoning, knowledge, quantitative reasoning,

visual-spatial processing and working memory, each in a verbal and a nonverbal domain, for a total of ten subtests. The materials consist of easel-mounted item books, blocks, toy pieces and pictures (Roid, 2003).

### A scene from administration

The specialist sits with Nawaf in a quiet room one morning, having first confirmed the results of his hearing and vision screening. She begins with the nonverbal routing subtest. Nawaf answers the first items, then starts making errors, and she stops at the discontinue rule specified in the manual. His scores on the two routing subtests set the starting level for the remaining subtests, so he begins at the earliest levels, usually given to children younger than he is. This is what allows the scale to map his profile rather than hit its floor. The specialist notices that he loses attention after about twenty minutes, so she gives him a short break and records this in her behavioral observations, since it is part of interpreting the score.

**Table 7.4: SB5 at a glance.**

| | |
|---|---|
| **Purpose** | Individual assessment of intellectual ability for diagnosis and planning |
| **Ages** | 2 to 85+ years |
| **Edition and year** | Fifth, 2003 |
| **Publisher** | Riverside at the time of publication; check the current distributor |
| **Qualification level** | Publishers restrict it to those with advanced training in administering and interpreting individual intelligence tests; check the publisher's current designation for the level |
| **Time** | About an hour for the full battery, less for the abbreviated form |

| | |
|---|---|
| **Score types** | Full Scale IQ, Verbal and Nonverbal IQs, and five factor indexes (mean 100, SD 15), with the Full Scale IQ ranging from 40 to 160; scaled scores for the ten subtests (mean 10, SD 3) |
| **Cutoff scores** | The scale itself has no diagnostic cutoff; decisions rest on DSM criteria and local regulations |
| **Reliability** | The manual reports an internal consistency coefficient of about .98 for the Full Scale IQ and a standard error of measurement of about two and a half points |
| **Classification accuracy** | The manual presents studies of special groups, including intellectual disability, but does not report sensitivity and specificity at a particular cutoff, as screening tools do |
| **Year of norm collection** | Around 2001–2002, with a US sample of 4,800 individuals |
| **Arabic version** | At least two Gulf versions exist: one prepared by Mahmoud Abu Al-Nil and colleagues, and one prepared by Safwat Farag with Arab norms; a study in the Northern Borders region compared the two (Mohammed, 2016 [Arabic]). We found no published data on their norm samples or the year those norms were collected |

### How are the results interpreted?

The Full Scale IQ is read first, together with its confidence interval, and then the agreement between the Verbal and Nonverbal IQs is examined. Close agreement points toward a general intellectual deficit, whereas a large difference favoring the nonverbal side calls for checking language and hearing before reaching a judgment. If most scaled scores sit at the very bottom of their range, the overall number confirms the deficit but cannot distinguish among its levels, and severity is then based on adaptive behavior (see Al-Anoud's case in 7.7).

### Its limits

Its original norms are more than twenty years old, and this matters more near the threshold of 70 than anywhere else (see the box after Maryam's case). The years in which the Arabic versions were standardized are not precisely known.

The Wechsler Intelligence Scale for Children, Fifth Edition (**WISC-V**), is also used with this population, for ages 6 years to 16 years 11 months; its US norms were collected around 2013–2014 (Wechsler, 2014). Its lowest Full Scale IQ is also 40, but its floor is higher on some subtests, so it may be less precise with moderate and severe intellectual disability. We found no published Saudi standardization of its fourth or fifth editions; the closest we found is a Saudi standardization of its older revised edition (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024).

## Vineland-3 Vineland Adaptive Behavior Scales, Third Edition

Adaptive behavior

### The idea

The Vineland Adaptive Behavior Scales, Third Edition (Vineland-3; Sparrow et al., 2016), measure what a person typically does in everyday life, not what he or she can do in a testing situation. The source is therefore always someone who knows the person well, and the question is: does the person perform this skill independently and regularly when it is needed?

### What it looks like

The scales come in three forms: a semi-structured interview with a parent or caregiver, a form completed by a parent, and a form completed by a teacher. Each has a comprehensive version and a shorter domain-level version. Items are distributed across three main domains, **Communication**, **Daily Living Skills** and **Socialization**, along with an optional Motor Skills domain

administered to young children up to about age nine and an optional Maladaptive Behavior scale. Each item is scored 2 if the person usually performs the behavior, 1 if the person performs it sometimes or partially, and 0 if the person does not perform it.

### A scene from administration

The specialist sits with the mother of a nine-year-old girl. Rather than reading the items to her one by one, he opens with a broad question: “Tell me about your daughter's morning on a school day, from the time she wakes up until she leaves the house.” From the mother's description he scores many items without asking about them directly. When she says her daughter “gets dressed on her own,” he asks, “Does she choose clothes that suit the weather? Does she need someone to remind her?” because the answer separates a 2 from a 1. He follows the start and discontinue rules in the manual, so he does not ask about skills far beyond the girl's level. The interview takes about an hour. At the end he asks, “Is there a skill she can do but doesn't get the chance to?” and notes the answer next to the score.

**Table 7.5: Vineland-3 at a glance.**

| | |
|---|---|
| **Purpose** | Measuring adaptive behavior for diagnosis and educational planning |
| **Ages** | Birth to 90 years for the interview and parent form; 3 to 21 years for the teacher form |
| **Edition and year** | Third, 2016 |
| **Publisher** | Pearson |
| **Qualification level** | The interview requires an examiner trained in semi-structured interviewing; check the qualification level specified by the publisher |
| **Time** | Between half an hour and an hour for the comprehensive form, less for the abbreviated form |

| | |
|---|---|
| **Score types** | An Adaptive Behavior Composite and domain scores (mean 100, SD 15); subdomain scores (mean 15, SD 3) |
| **Cutoff scores** | The scale itself has no cutoff; the profile is compared against the two-standard-deviation criterion, allowing for the margin of error |
| **Reliability** | The manual reports internal consistency coefficients for the composite in the .90s |
| **Classification accuracy** | The manual presents comparisons with clinical groups, including intellectual disability, without sensitivity and specificity at a particular cutoff |
| **Year of norm collection** | In the years before publication, with a US sample |
| **Arabic version** | An Arabic database lists a title, "Vineland Adaptive Behavior Scale: Saudi Norms," but we were unable to verify its author, year or sample, or the edition on which it was based. We know of no published Arabic standardization of the third edition |

### How are the results interpreted?

The Vineland domains **do not map directly** onto the three DSM domains. Socialization is closest to the social domain, Daily Living Skills is closest to the practical domain, and Communication covers part of the conceptual domain. Money and time skills, which are conceptual in the DSM, fall under Daily Living Skills in the Vineland. Linking the two systems is therefore a clinical judgment that the examiner must make and state explicitly in the report.

### Limitations

A score is only as accurate as the respondent and his or her understanding of the questions. The interview requires a skilled examiner, because leading questions can raise or lower the score. The US norms also assume

opportunities and expectations that may not be available to the child in his or her own setting (see the box below).

## ABAS-3 Adaptive Behavior Assessment System, Third Edition

Adaptive behavior

### The idea

The Adaptive Behavior Assessment System, Third Edition (ABAS-3; Harrison & Oakland, 2015), measures adaptive behavior through forms completed by people who know the child well. Its advantage for diagnosis is that it yields **the same three domains** used by the DSM and the AAIDD: conceptual, social and practical.

### What it looks like

Five forms: for parents (birth to 5 years, and 5 to 21), for teachers or daycare providers (2 to 5, and 5 to 21), and for adults (16 to 89). The respondent rates each skill according to whether the child performs it when it is needed, always, sometimes, or never, or is not able to perform it. Items are grouped into eleven skill areas, including communication, functional academics, self-care, and health and safety, then into the three domains, and finally into a general composite.

### A scene from administration

The specialist sends the parent form to Umm Nawaf with a short note explaining that the questions concern what Nawaf usually does in his day, and that "is not able" differs from "is not asked to." She gives his teacher the teacher form. When both forms come back, she reviews the omitted items, since the manual sets a maximum number for each skill area beyond which the area cannot be scored. She then calls the mother to ask about two items she answered by guessing.

**Table 7.6: ABAS-3 at a glance.**

| | |
|---|---|
| **Purpose** | Measuring adaptive behavior for diagnosis and planning, across the three DSM domains |
| **Ages** | Birth to 89 years, depending on the form |
| **Edition and year** | Third, 2015 |
| **Publisher** | WPS |
| **Qualification level** | Completed by a parent or teacher; scored and interpreted by a qualified professional; check the qualification level specified by the publisher |
| **Time** | About 15 to 20 minutes to complete the form |
| **Score types** | A General Adaptive Composite and scores for the three domains (mean 100, SD 15); scaled scores for the skill areas (mean 10, SD 3) |
| **Cutoff scores** | The scale itself has no cutoff; the profile is compared against the two-standard-deviation criterion, allowing for the margin of error |
| **Reliability** | The manual reports internal consistency coefficients for the general composite in the high .90s |
| **Classification accuracy** | The manual presents comparisons with clinical groups, without sensitivity and specificity at a particular cut-off |
| **Year of norm collection** | In the years before publication, with a US sample |
| **Arabic version** | The teacher form was adapted into Arabic in Oman and studied with 410 children with intellectual disability; the study supported a single general factor and high reliability (Emam et al., 2020). We found no published Saudi standardization |

### How are the results interpreted?

Deficits are considered clear when the general composite, or the score for at least one domain, falls approximately two standard deviations below the mean, again taking measurement error into account (Tassé et al., 2012). For this reason, the student's full profile is read across all domains alongside the total score:

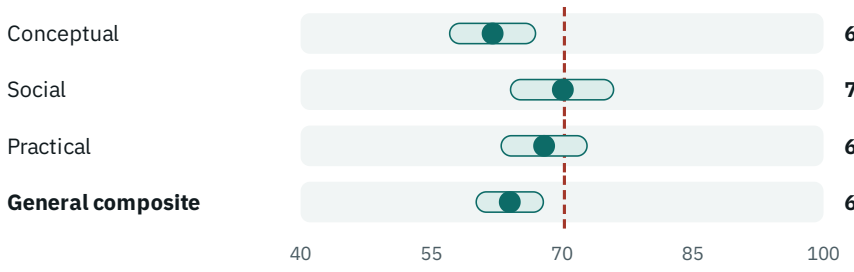



**Figure 7.3:** A hypothetical adaptive behavior profile for Nawaf on the Adaptive Behavior Assessment System, as rated by his mother, with the confidence interval around each score. The dashed line marks a score of 70. The conceptual domain is clearly lower, while the other two domains sit near the cut-off.

### Its limits

Unlike an interview, the form does not let the respondent explain an answer, so guessing becomes common when the respondent does not know how the child behaves in a particular situation. A new teacher knows the student less well than the family does. And no Arabic norms for the general child population are known for the instrument (Table 7.7).

In both cases, it is preferable to combine more than one respondent, such as the mother and the teacher, because a child may do at home what he does not do at school, and vice versa. The discrepancy between the two sources is information in its own right, which the examiner interprets and keeps in the report.

**For the specialist: A scale designed for decisions at the cut-off**

Most adaptive behavior scales are designed to describe performance across its full range, so their items are spread over all levels of difficulty. The **Diagnostic Adaptive Behavior Scale** (**Diagnostic Adaptive Behavior Scale, DABS**), developed by the American Association on Intellectual and Developmental Disabilities for ages 4 to 21, was instead built with item response theory (see 2.5.4) to reach its greatest precision in the critical region around two standard deviations below the mean, that is, where the diagnostic decision is made. It is administered through an interview with someone who knows the individual well, and yields scores for the conceptual, social, and practical domains and a total score. We know of no published standardized Arabic version at the time of writing. Its underlying idea is a general lesson in measurement: an instrument that measures everyone does not necessarily measure most precisely at the point where the decision is made.

**Adaptive behavior and culture**

What is expected of a child in everyday life varies across communities and families. A child in a family where children are spoon-fed until a late age may not eat independently simply because he has never been asked to. The examiner therefore always asks: Has the child had the opportunity to practice this skill? And is he expected to perform it in his environment?

## Available Arabic instruments

Table 7.7 summarizes the Arabic versions of this chapter's instruments that we were able to find at the time of writing. What matters most in it is what we know about their norms: when they were collected, from whom, and with how

many participants.

**Table 7.7: Known Arabic versions of instruments for diagnosing intellectual disability.**

| Instrument | What we found | What we do not know |
|---|---|---|
| **Stanford-Binet, Fifth Edition** | Two Gulf versions (Abu Al-Nil and colleagues; Farag), and a study comparing them in the Northern Borders Region (Mohammed, 2016 [Arabic]) | Size of the normative sample and year of collection |
| **Wechsler Intelligence Scale for Children** | A Saudi standardization of the older revised edition for ages 6–16 (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024) | A published Saudi standardization of the fourth and fifth editions |
| **Wechsler Preschool and Primary Scale of Intelligence, Third Edition** | A Saudi version administered to 788 children aged 4 years to 7 years 3 months in Jeddah between 2018 and 2022 (Ejbara, 2024) | Its generalizability to other regions of the Kingdom |
| **Vineland scales** | A title referring to Saudi norms in an Arabic database | Author, year, sample, and source edition |
| — | The teacher form in Oman, with 410 children with intellectual disability (Emam et al., 2020) | Saudi norms, and norms for the general child population |
| **Ages and Stages Questionnaires; Diagnostic Adaptive Behavior Scale** | Nothing published that we could find | — |

Readers may come across, in some centers, an old Arabic adaptation of the adaptive behavior scale issued by the American Association on Mental Deficiency in the 1970s (**AAMD Adaptive Behavior Scale**). Three questions arise

before relying on it: When were its Arabic norms collected, and on what sample? Do its items fit a child's life today, now that the ways people handle money, get around, and communicate have changed so profoundly? And are its domains read in terms of the three contemporary domains or according to an outdated division? A scale whose norms have aged compares the child with a generation that no longer exists.

## 7.7 When is the judgment made? Four cases

We follow four hypothetical cases, and before examining each in detail we place them on a single map that brings together the two dimensions: intellectual ability and adaptive behavior.

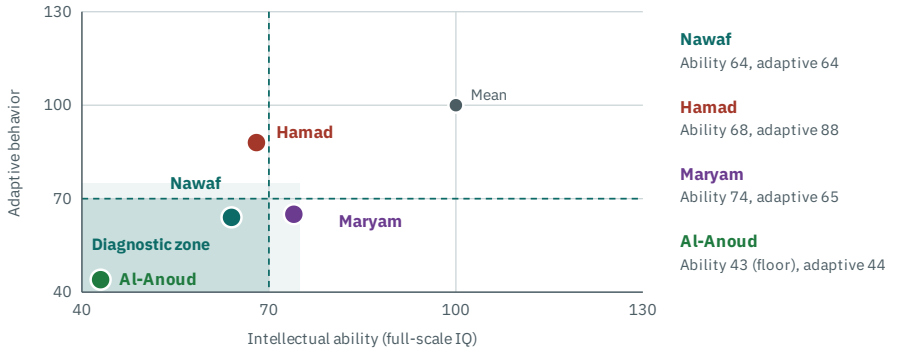



**Figure 7.4:** Where the four cases fall. The dark area is below 70 on both dimensions; the lighter band around it represents the margin of measurement error up to about 75. Nawaf lies inside the zone; Hamad has low ability but adaptive behavior close to average; Maryam is at the edge on ability with clear adaptive deficits; and Al-Anoud is at the bottom of the zone, where the test reaches its floor.

## Case 1: Nawaf, 6 years 8 months

All three criteria are clear

**The story:** He is the student in the opening scene. His teacher documented his difficulties for three weeks and gave him support in a small group, and he progressed very slowly. His developmental history shows delays in walking, talking, and toilet training, with no loss of previously acquired skills. His newborn screening results were normal.

**Table 7.8: Evidence in Nawaf's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision screening (before any instrument)** | Both normal | His senses do not explain his difficulties |
| **SB5** | Full-scale IQ 64 (59–70); verbal and nonverbal scores close to each other | A general intellectual deficit spanning both verbal and nonverbal abilities |
| **ABAS-3, rated by the mother** | General composite 64; conceptual 62 | Clear adaptive deficits, most pronounced in the conceptual domain |
| **Vineland-3, teacher form** | Adaptive Behavior Composite 67; communication 63, daily living skills 70, socialization 69 | The school sees the same picture. The lowest domain is communication, which corresponds, by clinical judgment, to part of the conceptual domain |
| **Language assessment** | Receptive and expressive language clearly below his age, and consistent with his intellectual level | No separate language disorder accounts for the picture on its own |

| A ✓ | B ✓ | C ✓ |
| --- | --- | --- |
| Ability 64 | Adaptive 64 and 67 from two sources | Delay since infancy |

**Decision:** Nawaf is diagnosed with intellectual disability at the mild level, because his adaptive deficits still allow him to manage basic self-care, while he needs support with academic and social skills. He also meets the eligibility requirements of the Special Education Organizational Guide (see 7.10). The team recommends identifying his support needs at school and referring him to a developmental pediatrician or a geneticist for first-tier genetic testing, a standard step in every case of unexplained intellectual disability.

**And the answer to his mother's question:** Nawaf will learn and make progress, and he may learn to read and acquire functional math. But his progress will be slower than that of his classmates, and he needs clearer, more repetitive instruction and goals designed for his level, which will serve him far better than redoubling "effort" in the same curriculum.

> **What evidence could change the decision?** If Nawaf began to lose skills he had previously acquired, or if a reevaluation after years of intensive support showed adaptive behavior approaching the average range, the first scenario would call for an urgent medical referral and the second for a review of the diagnosis and its level.

## Case 2: Hamad, 10 years Adaptive deficits are absent

**The story:** Hamad moved with his family from another country two years ago. His first language is not Arabic, and he had not attended school regularly before the move. His achievement in every subject is very low, so his teacher referred him on suspicion of intellectual disability.

**Table 7.9: Evidence in Hamad's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision screening** | Both normal | His senses do not explain his difficulties |
| **Verbal intelligence test in Arabic** | Full-scale score 68 | Low, but the test was given in a language he has not mastered |
| **Nonverbal intelligence test** | 89 | His visual reasoning is close to average |
| **Vineland-3, interview with the father through an interpreter** | Adaptive Behavior Composite 88 | Takes care of himself, helps in his family's shop, and gets around the neighborhood on his own |
| **Recess observation** | Plays and negotiates fluently with children who speak his language | Age-appropriate social skills in his own language |

| A ? | B ✗ | C |
|---|---|---|
| Test not linguistically valid | Adaptive 88 | Not applicable |

**Decision:** Hamad is not diagnosed with intellectual disability. His low score came from a test in a language he has not mastered, and his adaptive behavior in daily life is close to average. He needs an intensive Arabic-language program and compensation for the schooling he missed, with his progress monitored. If he still does not progress, the possibility of learning disabilities should be reconsidered.

**The lesson**

Criterion B is what saved Hamad from a misdiagnosis. That is precisely why it is required: a low score may be measuring language or opportunity, whereas everyday life reveals the actual capacity to adapt.

**What evidence could change the decision?** If his progress in Arabic and in achievement remained very slight after a sufficient period in the intensive program, or if information from the school and from his mother revealed deficits in daily life that the interview with his father had not uncovered, he would need to be reassessed with nonverbal instruments and multiple sources before intellectual disability could be ruled out for good.

## Case 3: Maryam, 12 years

A score above 70, and diagnosis is possible

**The story:** Maryam is in sixth grade and has received resource room support since second grade as a student with learning disabilities. Yet her progress is weak in every subject; she needs someone to remind her of each step in her day, she does not understand money or appointments, and more than once she has been exploited by classmates who talked her into handing over her pocket money.

**Table 7.10: Evidence in Maryam's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision screening** | Both normal | Her senses do not explain her difficulties |

| Source | Result | What it means |
|---|---|---|
| **WISC-V** | Full-scale IQ 74 (69–80); index scores close together, between 70 and 78 | Generally low ability, with a confidence interval that extends below 70 |
| **ABAS-3, rated by the mother** | General composite 65; conceptual 60, social 68 | Clear adaptive deficits in more than one domain |
| **ABAS-3, rated by the teacher** | General composite 66 | The same picture at school |
| **Developmental history** | Delays in language and self-help skills since early childhood | Onset during the developmental period |
| **Intervention data** | Very limited progress in all areas despite four years of support | The pattern does not resemble that of learning disabilities |

| A ✓ | B ✓ | C ✓ |
|---|---|---|
| 74 within the error margin | Adaptive 65 from two sources | Since early childhood |

**Decision:** The team concludes that Maryam meets the criteria for intellectual disability at the mild level. Her score of 74 lies within the margin of measurement error, her confidence interval extends below 70, and her adaptive deficits are clear from two sources and have persisted since early childhood. Her earlier classification is thereby corrected, so that she receives support matched to her real needs, including direct instruction in self-protection and money management. Medical and genetic referral is recommended, as in Nawaf's case.

Had Maryam scored 74 with adaptive behavior close to average, she would not have been diagnosed. This is what it means to say that the cut-off is a professional judgment that gathers all the evidence around the

number.

> **What evidence could change the decision?** If a careful interview with the mother revealed that her low ratings reflect the limited opportunities Maryam has at home, or if a test with more recent norms yielded a clearly higher score with wide index scatter resembling a learning disability pattern, the team would need to revisit the decision.

**For the specialist: The age of norms near the cut-off**

In many countries, average performance on intelligence tests has risen from one generation to the next, and a meta-analysis estimated this gain at about two to three standard score points per decade (Trahan et al., 2014; see also 2.7). If this holds in the child's environment, each decade that passes after the norms were collected raises the child's score by a similar amount, because the child is being compared with a generation that performed less well than his or her own.

Let us apply this to Maryam. The U.S. norms for the Wechsler scale were collected around 2013–2014, so if Maryam were tested in 2026 the norms would be about twelve years old, implying a possible inflation of between two and a half and three and a half points. Her score of 74 might therefore correspond to about 71 on contemporary norms, which strengthens the decision rather than changing it. If, however, the 74 had been obtained on the Stanford-Binet, whose norms were collected around 2001, the norms would be nearly twenty-five years old and the possible inflation between five and seven points, so the score might correspond to 67 to 69. This difference carries practical weight, since it falls on the very line that separates eligibility under the Saudi regulations (see 7.10).

Even so, the number is not subtracted mechanically. The effect varies from country to country, it has slowed or reversed in some, and we know of no studies estimating it in the Kingdom. The collection years of Arabic standardizations are, for the most part, not precisely known. Accepted practice is for the examiner to state the year of the norms in the report, to take their age into account when interpreting a score near the cut-off, as the AAIDD manual recommends (Schalock et al., 2021), and to weigh this alongside the rest of the evidence.

## Case 4: Al-Anoud, 15 years

**The diagnosis stands; the question is severity and support**

**The story:** Al-Anoud is in ninth grade (third year of intermediate school) in an intellectual disability program attached to a girls' school. She was diagnosed at age seven, when her full-scale score on the Arabic version of the Stanford-Binet was 56, and she was placed in the "educable" category. As she moves to secondary school, the team has requested a reevaluation. Her teacher says her progress has "stalled": she reads about twenty familiar words, including her name and some signs, counts objects up to ten, and does not handle money. She speaks in short sentences that her family understands, dresses and eats on her own, needs reminders about personal hygiene, and does not leave the house alone. Her parents are first cousins, and she has a younger brother who is delayed in walking and talking. The family has never been referred for genetic testing.

**Table 7.11: Evidence in Al-Anoud's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision screening (this year)** | Both normal | Her senses do not explain her difficulties |
| **Stanford-Binet, the same Arabic version** | Full-scale score 43, with eight of the ten subtests at a scaled score of 1 or 2 | A score close to the test's floor: it confirms a substantial intellectual deficit but does not distinguish among its levels |
| **Vineland-3, interview with the mother** | Adaptive Behavior Composite 44; communication 40, daily living skills 52, socialization 42 | Substantial adaptive deficits, mildest in self-care |

| Source | Result | What it means |
|---|---|---|
| **ABAS-3, rated by the teacher** | General composite 46; conceptual 41, social 48, practical 53 | The same picture at school |
| **Developmental and family history** | Delay since infancy, no loss of acquired skills, consanguineous parents, and a developmentally delayed younger brother | Onset during the developmental period, and strong indicators of a genetic cause |
| **Genetic testing (after referral)** | Chromosomal microarray normal; exome sequencing revealed a recessive genetic cause | The cause is known; the family is referred for genetic counseling, and the younger brother is evaluated |

**A ✓**
Below 50 despite the floor

**B ✓**
Adaptive 44 and 46 from two sources

**C ✓**
Since infancy

**Decision:** The diagnosis stands, and its level is specified as **moderate**. Al-Anoud manages basic self-care after extended teaching, needs ongoing support with most independent tasks, and her spoken language is simple, which matches the DSM description of the moderate level (Table 7.2). ICD-11 agrees, since both her ability and her adaptive behavior fall between three and four standard deviations below the mean. The drop in her score from 56 to 43 does not mean that she has regressed. A standard score compares her with same-age peers, and the gap between her and them widens with age even as she continues to learn, while the test's floor compresses low scores together. The team recommends a transition plan for secondary school that focuses on functional academic skills (reading safety signs, using money with a

calculator), supervised community mobility skills, early vocational exploration, and determining the intensity of her support needs in each domain.

**The lesson**

When a test reaches its floor, the IQ stops being informative: it confirms the presence of a deficit but says nothing about its level or about the support required. Both severity and planning then shift to adaptive behavior. This is why ICD-11 provides behavioral indicators for use when tests fall short. The difference between a standard score that declines with age, which is expected, and a loss of acquired skills, which is a medical red flag, is one the team must ask about explicitly.

**What evidence could change the decision?** If it emerged that Al-Anoud had begun to lose skills she had acquired, or if the geneticist determined that the identified cause is a progressive disorder, the question would shift from specifying severity to urgent medical follow-up and a comprehensive reevaluation of her level and needs.

## 7.8 Differential diagnosis

**Table 7.12: Differential diagnosis of intellectual disability.**

| Condition | Similarity | Distinguishing features |
|---|---|---|
| **Learning disabilities** | Low achievement and difficulty at school | Intellectual ability and adaptive behavior in or near the average range; the difficulty is confined to a specific academic area |

| Condition | Similarity | Distinguishing features |
|---|---|---|
| **Borderline intellectual functioning** | Low ability and difficulty learning | The individual does not meet the criteria for intellectual disability yet still needs support. The DSM lists it among "Other Conditions That May Be a Focus of Clinical Attention" without a numerical cut-off. The 71–84 range common in textbooks comes from the fourth edition of the manual (DSM-IV). Distinguishing it from mild intellectual disability requires careful assessment that considers ability and adaptive behavior together |
| — | Language delay and weak performance on verbal tests | Adaptive behavior and nonverbal ability are clearly better than language. A recent international consensus dropped the requirement of a discrepancy between language and nonverbal intelligence for this diagnosis, but it treated language disorder co-occurring with intellectual disability as a separate category (Bishop et al., 2017) |
| **Autism spectrum disorder** | Weak communication and social skills | Social deficits are more severe than would be expected for the intellectual level, together with restricted and repetitive patterns of behavior. The two disorders may co-occur in the same student (see 7.9) |
| **Hearing or visual impairment** | Weak performance on tests and in learning | The picture improves considerably when the student is assessed through a medium suited to his or her senses, which is why sensory screening precedes every instrument |

| Condition | Similarity | Distinguishing features |
|---|---|---|
| **Acquired brain injury** | Intellectual and adaptive deficits | A typical developmental history before the injury, and an uneven profile dominated by weaknesses in attention, memory, and processing speed. If the injury occurred during the developmental period and the criteria are met, intellectual disability may be diagnosed alongside it |
| **Limited opportunity and language difference** | Low scores and weak achievement | Adaptive behavior appropriate to the student's environment, and rapid progress with instruction, as in Hamad's case |
| **Severe environmental deprivation** | Delays across multiple areas | A history of neglect, and marked improvement when care improves |

## 7.9 Co-occurrence

Differential diagnosis asks: which of the two conditions? Co-occurrence asks instead: is another condition present as well? In intellectual disability, co-occurrence is the rule rather than the exception. In a British national sample, 36% of children and adolescents with intellectual disability had a diagnosable psychiatric disorder, compared with 8% of their peers (Emerson & Hatton, 2007). The greatest risk here is what is known as **diagnostic overshadowing** (**Diagnostic overshadowing**): attributing every new symptom to the intellectual disability itself, so that anxiety, depression, pain, or epileptic seizures go unnoticed (Reiss et al., 1982).

**Table 7.13: Common co-occurring conditions in intellectual disability and what their assessment requires.**

| Co-occurring condition | What we know about its frequency | Assessment guidance |
|---|---|---|
| **Autism spectrum disorder** | In U.S. surveillance data for 2020, 37.9% of children with autism for whom ability data were available had intellectual disability (Maenner et al., 2023) | Autism is diagnosed alongside intellectual disability only when social communication is below what would be expected for the general developmental level (APA, 2022). Standardized observation tools, such as the Autism Diagnostic Observation Schedule, become less accurate as mental age drops substantially, so their results are read with caution in severe disability |
| **—** | Higher than in the general child population, with wide variation across studies depending on sample and instrument | Symptoms are compared with what is expected for **mental age**, not chronological age, since the DSM requires them to be excessive relative to it (APA, 2022). Rating scales are normed on the general child population, so their scores rise simply because developmental level is low |
| **Epilepsy** | About one fifth of people with intellectual disability, with the proportion rising with severity (Robertson et al., 2015) | Ask the family about seizures, their type, and medications, because unnoticed seizures and some sedating medications lower attention and test performance. Schedule the assessment away from seizure episodes |
| **Sensory impairments** | More common than in the general child population, especially in severe disability and certain syndromes | Hearing and vision screening is the second step in the pathway and is repeated periodically, because sensory impairments may emerge or change over time |

| Co-occurring condition | What we know about its frequency | Assessment guidance |
|---|---|---|
| **Anxiety and depression** | Higher than among peers | A student with limited language expresses them through behavior: withdrawal, sleep disturbance, or new aggressive behavior. Ask about change from the student's usual baseline, and rely on more than one source of information |
| **Physical disability, such as cerebral palsy** | Frequently co-occurs with intellectual disability | Timed and motor tests lower the score for motor rather than intellectual reasons, so choose tests that do not depend on movement, and report the effect of any accommodation on the validity of the norms |

The practical rule is that after every diagnosis of intellectual disability the team asks two questions: Is there anything in the file that intellectual disability alone does not explain? And was that aspect measured with an instrument suited to the student's developmental level?

# 7.10 In the Saudi context: Eligibility criteria

**In the Saudi context**

Table 7.14 sets the criteria presented in this chapter against those of the U.S. Individuals with Disabilities Education Act (**IDEA**) and of two Saudi documents: the Regulations for Special Education Institutes and Programs (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]) and the Special Education Organizational Guide, the most recent published edition we could find (Ministry of Education, 1436 AH [Arabic]). The Saudi criteria are criteria of **eligibility for services** in education, and their function differs from that of medical diagnostic criteria.

**Table 7.14: Criteria for intellectual disability in the DSM and ICD-11, in U.S. law, and in the Saudi regulations.**

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Terminology** | Intellectual disability; disorders of intellectual development | **Intellectual disability** | "Mental retardation" in the Regulations, and "mental handicap" in the Guide |

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Intellectual ability cut-off** | Approximately two standard deviations below the mean, taking measurement error into account, with no fixed number | Significantly subaverage general intellectual functioning, with no number in the federal text; states set their own criteria | Guide: no higher than 75 on the Wechsler, or 69 on the Stanford-Binet Fifth Edition, or the equivalent. Regulations: 55–75 on the Wechsler (52–73 on the Stanford-Binet) for the "educable" category, and 40–54 (36–51) for the "trainable" category |
| **Adaptive behavior** | Deficit in at least one of three domains, on a standardized measure | Deficits in adaptive behavior, without specifying the number of domains | Deficits in at least two of nine areas, "according to formal or informal adaptive behavior tests" |
| **Onset** | During the developmental period (before 22 according to AAIDD) | Manifested during the developmental period | "During the developmental stage," with no specified age; the admission age in the Regulations is 6 to 15 years |

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Educational impact** | Not required for diagnosis | Required: must adversely affect educational performance | Implicit in the definition of a student with a disability: a deficit "to the extent that requires the provision of special education services" |
| **Severity** | By adaptive behavior (DSM-5-TR), and by both dimensions together (ICD-11) | Not specified | Regulations: by IQ range. We found no severity classification in the eligibility requirement of the Guide |

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Who decides** | A qualified professional under each country's regulations | A team of qualified professionals together with the parents | The assessment and diagnosis team, comprising the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the student counselor, specialists relevant to the category, the physician if available, and the parent or guardian. Instruments are administered and interpreted by "qualified specialists" using instruments "approved by the Ministry of Education" |

Sources: APA (2022); WHO (2024); Schalock et al. (2021); Assistance to States for the Education of Children With Disabilities (2006, §300.8(c)(6)); Ministry of Education [Al-Ma'arif] (1422 AH [Arabic], Articles 1 and 6); Ministry of Education (1436 AH [Arabic]).

This comparison prompts three observations. **The first** is that the two numbers in the Guide are not equivalent. A cut-off of 75 on the Wechsler lies about one and two-thirds standard deviations below the mean, whereas a cut-off of 69 on the Stanford-Binet Fifth Edition lies more than two standard deviations below it. In the older Regulations, the difference between 75 and 73 reflected the different standard deviation of earlier editions of the Stanford-Binet, which was 16 rather than 15. The fifth edition, however, has a standard deviation of 15, like the Wechsler, so this cannot explain the gap between 75 and 69, and we found no rationale for it in the Guide. Maryam, who scored 74 on the Wechsler, meets the intelligence requirement; had she obtained the same 74 on the Stanford-Binet, she would not, even though the two scores denote the same position relative to peers.

**The second** is that the nine adaptive behavior areas in the Guide (communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, and leisure and work skills) are close to the ten areas in the 1992 AAIDD definition (Luckasson et al., 1992), which the association itself has since replaced with the three domains. Requiring two of nine areas differs from requiring one of three domains, and permitting "informal" instruments for this criterion is broader than the DSM's call for a standardized measure. **The third** is that the Guide requires instruments "approved by the Ministry of Education," yet at the time of writing we found no published list of such instruments, nor any published Saudi standardization of the recent editions of the Wechsler scales or of adaptive behavior scales (Table 7.7).

**A question for reflection:** If Maryam's case came before you at your school, how would you write the team's decision so that it respects the eligibility criterion in the Organizational Guide while also reporting the confidence interval and the age of the norms?

## Conclusion

The argument we have developed in this chapter centers on the claim that intellectual disability is a composite judgment that cannot be derived from an IQ score alone. Adaptive behavior is the criterion that guards against misdiagnosis; it determines the level of severity in the DSM, and it becomes the primary source for judging severity once an intelligence test reaches its floor. The cut-off of 70 is read in light of the confidence interval and the age of the norms, so a person scoring around 75 may be diagnosed if the adaptive deficits are clear. It follows that hearing and vision screening must precede every instrument, that referral for genetic testing is recommended in every case of unexplained intellectual disability, that the diagnosis of global developmental delay is used provisionally below age five, and that diagnostic overshadowing must be guarded against when other disorders accompany the disability.

Still, the chapter leaves questions that remain matters of professional judgment, beyond what the rules can settle. The Saudi criteria are eligibility criteria that differ from those of the DSM in the intelligence cut-off and in the number of adaptive areas, and within them the Wechsler cut-off differs from the Stanford-Binet cut-off, so a diagnosis may apply to a student who does not meet the eligibility criteria. The educational question therefore remains the one most deserving of attention: What support does the student need, where, and how much?

# 7.11 Summary, key terms, and activities

## Chapter summary

- Intellectual disability is diagnosed when three criteria are met together: intellectual deficits, adaptive deficits, and onset during the developmental period. AAIDD has raised the age of onset to before 22.
- The cut-off of 70 is not absolute; the confidence interval and the age of the norms are taken into account, so a person scoring up to about 75 may be diagnosed if the adaptive deficits are clear.
- Adaptive behavior is the criterion that protects against misdiagnosis; it determines the level of severity in the DSM and shares that role with ability in ICD-11. When an intelligence test reaches its floor, adaptive behavior becomes the primary source for judging severity.
- Below age five, global developmental delay is used as a provisional diagnosis; its ICD-11 counterpart is "disorder of intellectual development, provisional," below age four.
- Hearing and vision screening is a prerequisite before any instrument; referral for genetic testing is recommended in every case of unexplained intellectual disability, and becomes more important when the parents are related.
- The Vineland domains do not map directly onto the DSM domains, whereas the Adaptive Behavior Assessment System yields them directly.
- The Saudi criteria are eligibility criteria that differ from those of the DSM in the intelligence cut-off and the number of adaptive areas, and within them the Wechsler cut-off differs from the Stanford-Binet cut-off.
- Co-occurrence is common, diagnostic overshadowing is a constant risk, and co-occurring symptoms are judged against mental age.

- The most important educational question: What support does the student need, where, and how much?

## Glossary of terms

| | |
|---|---|
| **Intellectual disability** | Significant deficits in intellectual functioning and adaptive behavior that begin during the developmental period. |
| **Adaptive behavior** | The conceptual, social, and practical skills that people learn and perform in their everyday lives. |
| **Global developmental delay** | A provisional diagnosis for a child under five who is delayed in two or more developmental domains. |
| **Disorder of intellectual development, provisional** | An ICD-11 category for a child under four, or for a person who cannot be validly assessed. |
| **Routing subtest** | A short subtest that determines the starting level for the remaining subtests of a scale. |
| **Support needs intensity** | The amount and type of support a person needs to participate in life activities. |
| **Borderline intellectual functioning** | Below-average intellectual ability that does not meet the criteria for intellectual disability, for which the current DSM sets no numerical cut-off. |
| **Chromosomal microarray** | A genetic test that detects small deletions or duplications of chromosomal material that conventional karyotyping cannot see. |

| | |
|---|---|
| **Exome sequencing** | A genetic test that reads the protein-coding portions of nearly all genes in search of a causal variant. |
| **Diagnostic overshadowing** | Attributing new symptoms to the intellectual disability itself, so that a co-occurring condition that could be diagnosed and treated goes unnoticed. |

## Review questions

1. Why did the DSM shift the specification of severity from IQ to adaptive behavior? And how does ICD-11 differ from it in this respect?
2. A student has a full-scale IQ of 66 and an adaptive behavior score of 92. Should he be diagnosed with intellectual disability? And what would you look for next?
3. Using the numbers in Maryam's case, explain why 70 is not an absolute cut-off. Then estimate the range of the corresponding contemporary score had her 74 been obtained on the Stanford-Binet Fifth Edition with its original norms, and explain why this difference is not subtracted mechanically.
4. What is the floor effect? Drawing on Al-Anoud's case, show what an IQ can and cannot tell us at the floor; then choose a suitable intelligence test for a four-year-old suspected of having moderate intellectual disability, and justify your choice.
5. How does a family's culture affect ratings of adaptive behavior? Give two examples.
6. What is the difference between intellectual disability and borderline intellectual functioning? And why is it no longer appropriate to define borderline intellectual functioning by the 71–84 range?

7. A student scored 74 on the Wechsler scale and, two months later, 72 on the Stanford-Binet Fifth Edition, and has deficits in two adaptive areas. Does he meet the intelligence requirement of the Special Education Organizational Guide with each of the two scores? And what does this reveal about the consistency of the criterion?
8. When should the educational team recommend genetic referral? And why do normal newborn screening and premarital screening not make it unnecessary?
9. A student with moderate intellectual disability was rated by his teacher for attention problems on a rating scale normed on the general student population, and his score came out very high. What would you compare before accepting a co-occurring diagnosis of attention-deficit/hyperactivity disorder?
10. **Case for application:** Bassam is eight years old and in second grade. He was given the Arabic version of the Stanford-Binet, whose norm collection year is unknown, and obtained a full-scale score of 72 (67–78), with a verbal score of 70 and a nonverbal score of 76. His mother completed the Vineland form, yielding an Adaptive Behavior Composite of 74 (communication 70, daily living skills 80, socialization 73). His teacher completed the Adaptive Behavior Assessment System form, yielding a general composite of 66 (conceptual 62, social 70, practical 69). He walked at 14 months and said his first words after age two. His mother reports recurrent ear infections in his early years, and his hearing has not been tested since he started school. He received small-group reading support for one semester and progressed slowly. (a) What missing evidence must precede any judgment, and why? (b) Is Criterion A met according to the DSM? And is the intelligence requirement of the Organizational Guide met? (c) How do you interpret the difference between the mother's and the teacher's ratings? (d) What is your decision now: diagnosis, no diagnosis, or deferring judgment? Justify your decision, and specify what you need in order to make it.

**Table 7.15: Mapping chapter objectives to review questions and learning activities.**

| Objective | Assessed by |
|---|---|
| **1. Determine whether all three criteria are met and classify examples of adaptive deficits** | Questions 2 and 5; Activity 2 |
| **2. Determine the severity level and compare it with the international classification** | Question 1; first part of Question 4 |
| **3. Select an instrument and justify the choice** | Question 4; Activity 3 |
| **4. Apply the Saudi eligibility criteria** | Question 7; part (b) of Question 10; Activity 3 |
| **5. Judge the sufficiency of evidence near the boundaries** | Question 3; Question 10 (Case for application); Activity 1 |
| **6. Distinguish intellectual disability from similar and co-occurring conditions, and make medical and genetic referrals** | Questions 6, 8 and 9 |

## Learning activities

Any activity that involves administering an instrument to a child or using a real student's data must be carried out with the consent of the family and the school administration, without any information that could identify the student, and under the supervision of a qualified professional; its results must not be used in any decision concerning the student.

1. **Mapping cases:** On a map like Figure 7.4, place five hypothetical cases of your own design, two of them near the boundaries. For each, record the decision and its rationale, along with the effect of the age of the norms, if any. *Product: a diagram and a short table. Time: about one hour.*
2. **Translating the profile:** Turn the profile in Figure 7.3 into three measurable instructional goals in the conceptual domain that suit Nawaf,

and identify for each goal the adaptive domain it serves. *Product: three written goals. Time: about 45 minutes.*

3. **Instrument profile:** Choose an adaptive behavior scale whose record form and manual are available to you, and prepare a profile for it using the standard template in 7.6, including the Arabic version field. Then map its domains onto the nine domains in the Special Education Organizational Guide and the three domains in the diagnostic manual, and show what it does not cover. *Product: a profile and a mapping table. Time: about two hours.*

## Chapter references

Mohammed, H. A. A. (2016). [Differences between the two Gulf versions of the Stanford-Binet Intelligence Scales, Fifth Edition, in assessing and diagnosing intelligence and mental abilities among persons with disabilities in the Northern Borders Region] [In Arabic]. *Jazan University Journal of Human Sciences, 5*(Suppl.), 236–254. https://search.mandumah.com/Record/1103082

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Health. (2016). *[The national newborn screening program for the prevention of disability: Operational manual]* [In Arabic]. https://www.moh.gov.sa/en/Ministry/About/Health%20Policies/019.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Assistance to States for the Education of Children With Disabilities, 34 C.F.R. pt. 300 (2006) (§300.8(c)(6)).

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

el-Hazmi, M. A., al-Swailem, A. R., Warsy, A. S., al-Swailem, A. M., Sulaimani, R., & al-Meshari, A. A. (1995). Consanguinity among the Saudi Arabian population. *Journal of Medical Genetics, 32*(8), 623–626. https://doi.org/10.1136/jmg.32.8.623

El Mouzan, M. I., Al Salloum, A. A., Al Herbish, A. S., Qurachi, M. M., & Al Omar, A. A. (2008). Consanguinity and major genetic disorders in Saudi children: A community-based cross-sectional study. *Annals of Saudi Medicine, 28*(3), 169–173. https://doi.org/10.5144/0256-4947.2008.169

Emam, M. M., Al-Sulaimani, H., Omara, E., & Al-Nabhany, R. (2020). Assessment of adaptive behaviour in children with intellectual disability in Oman: An examination of ABAS-3 factor structure and validation in the Arab context. *International Journal of Developmental Disabilities, 66*(4), 317–326. https://doi.org/10.1080/20473869.2019.1587939

Emerson, E., & Hatton, C. (2007). Mental health of children and adolescents with intellectual disabilities in Britain. *British Journal of Psychiatry, 191*(6), 493–499.

Harrison, P. L., & Oakland, T. (2015). *Adaptive Behavior Assessment System* (3rd ed.). WPS.

Luckasson, R., Coulter, D. L., Polloway, E. A., Reiss, S., Schalock, R. L., Snell, M. E., Spitalnik, D. M., & Stark, J. A. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.). American Association on Mental Retardation.

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., . . . Shaw, K. A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveillance Summaries, 72*(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Manickam, K., McClain, M. R., Demmer, L. A., Biswas, S., Kearney, H. M., Malinowski, J., Massingham, L. J., Miller, D., Yu, T. W., & Hisama, F. M. (2021). Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: An evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine, 23*(11), 2029–2037. https://doi.org/10.1038/s41436-021-01242-6

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities, 32*(2), 419–436.

Memish, Z. A., & Saeedi, M. Y. (2011). Six-year outcome of the national premarital screening and genetic counseling program for sickle cell disease and β-thalassemia in Saudi Arabia. *Annals of Saudi Medicine, 31*(3), 229–235. https://doi.org/10.4103/0256-4947.81527

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., Faucett, W. A., Feuk, L., Friedman, J. M., Hamosh, A., Jackson, L., Kaminsky, E. B., Kok, K., Krantz, I. D., Kuhn, R. M., . . . Ledbetter, D. H. (2010). Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics, 86*(5), 749–764. https://doi.org/10.1016/j.ajhg.2010.04.006

Reiss, S., Levitan, G. W., & Szyszko, J. (1982). Emotional disturbance and mental retardation: Diagnostic overshadowing. *American Journal of Mental Deficiency, 86*(6), 567–574.

Robertson, J., Hatton, C., Emerson, E., & Baines, S. (2015). Prevalence of epilepsy among people with intellectual disabilities: A systematic review. *Seizure, 29*, 46–62.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Riverside.

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and systems of supports* (12th ed.). AAIDD.

Shevell, M., Ashwal, S., Donley, D., Flint, J., Gingold, M., Hirtz, D., Majnemer, A., Noetzel, M., & Sheth, R. D. (2003). Practice parameter: Evaluation of the child with global developmental delay. *Neurology, 60*(3), 367–380. https://doi.org/10.1212/01.wnl.0000031431.81555.16

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales* (3rd ed.). Pearson.

Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires* (3rd ed.). Paul H. Brookes.

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Jr., Borthwick-Duffy, S. A., Spreat, S., Thissen, D., Widaman, K. F., & Zhang, D. (2012). The construct of adaptive behavior: Its conceptualization, measurement, and use in the field of intellectual disability. *American Journal on Intellectual and Developmental Disabilities, 117*(4), 291–303. https://doi.org/10.1352/1944-7558-117.4.291

Thompson, J. R., Wehmeyer, M. L., Hughes, C., Shogren, K. A., Seo, H., Little, T. D., Schalock, R. L., Realon, R. E., Copeland, S. R., Patton, J. R., Polloway, E. A., Shelden, D., Tanis, S., & Tassé, M. J. (2016). *Supports Intensity Scale–Children's Version: User's manual*. American Association on Intellectual and Developmental Disabilities.

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

# 8 Learning disabilities

Chapter 8

A bright student understands a story when he hears it and discusses its events in depth, then stumbles over a single line of it on the page; in this contradiction lies the essence of learning disabilities. In this chapter we consider how to tell them apart from weakness that stems from inadequate instruction, an unmastered language, or low general ability.

---

**Scene from a third-grade classroom**

In science class, the teacher asked why ice melts faster in the sun. Ahmed's hand went up first, and he explained the answer in words that surprised her. An hour later, in Arabic class, he was asked to read a short, fully vowelized paragraph. He halted at every word. He read "سَحَابَة" (saḥāba, "cloud") as "سَحَبَة," dropping the long vowel; he read "فَرِحَ" (fariḥa, "he was glad") as "فَرَحَ," turning the kasra into a fatḥa even though it was printed right in front of him; and he guessed at words from their first letters. Then his face reddened, and he said, "My eyes are tired."

His mother says he hates homework and cries before spelling tests, and that his older brother "was exactly the same and still hates reading."

The scene carries many of the hallmarks of dyslexia: a gap between oral comprehension and reading; decoding errors of the kind commonly seen among struggling readers of Arabic, such as dropping long vowels and swapping short vowels; and, alongside these, avoidance and a family history. Notably, Ahmed's errors lie in linking letters to their sounds, which we will examine as the core of dyslexia. Yet the scene alone does not amount to evidence for a diagnosis, and we return to Ahmed in the first case.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Distinguish the educational term "learning disabilities" from the diagnostic term "specific learning disorder," and compare how the category is defined in the DSM, the ICD, US law, and Saudi regulations.
2. Match a student's data against the four criteria for specific learning disorder, and specify the affected domain, the subskill, and the level of severity.
3. Classify a struggling reader's profile into one of the four patterns of the simple view of reading, and identify what must be measured to confirm the classification.
4. List the early warning signs in reading, writing, and mathematics for a student at a given age.
5. Select tools for assessing phonological processing, fluency, achievement, number sense, and written expression, and explain the limits of their use with Arab students.
6. Apply to a new case a decision pathway that combines response-to-intervention data, achievement, and exclusion, and justify the decision on each criterion.
7. Distinguish learning disabilities from conditions that resemble them, including developmental coordination disorder, and propose an assessment procedure for each common co-occurring condition.

# Introduction

At their core, learning disabilities involve a contradiction that often gives teachers pause: a student understands what he hears and discusses it in depth, then stumbles over a single line of text or a simple calculation. The diagnostic task is to distinguish this difficulty from weakness explained by inadequate instruction, an unmastered language, low general ability, or a sensory impairment. The question grows more complex in the Arab context, given the distinctive features of Arabic reading in its vowel marks, long vowels, and morphological structure, and given how few standardized Arabic tools exist in this area.

We open the chapter by defining the concept of learning disabilities and its terminology (Section 8.1), then present the diagnostic criteria, the meaning of "substantially below" achievement, and the specifiers (Section 8.2). We then turn to dyslexia in light of the simple view of reading, phonological components, and the features of Arabic reading (Section 8.3), followed by difficulties in writing and mathematics (Section 8.4), early warning signs (Section 8.5), the decision pathway (Section 8.6), and instrument profiles (Section 8.7). We apply all this to four cases (Section 8.8), move on to differential diagnosis and co-occurrence (Section 8.9), and close with the Saudi and Arab context and the tools available within it (Section 8.10).

## 8.1 What are learning disabilities?

**Specific learning disorder** (**Specific learning disorder**) is a neurodevelopmental disorder that shows itself as persistent difficulty in learning and using basic academic skills, such as reading words, understanding what is read, spelling, written expression, calculation, or mathematical reasoning. In education systems it is known as **learning disabilities** (**Learning disabilities**). The word **specific** means that the difficulty lies in one or more particular domains and is not explained by low general ability, a sensory impairment, or a lack of instruction.

In this chapter we follow a single rule of terminology: “learning disabilities” for the educational category served in schools, and “specific learning disorder” when we speak of the DSM criteria. The two terms are close in substance, but the first is governed by an education system and the second by a clinical manual, and a student may qualify under one and not the other when the criteria differ.

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**), uses a single label and then specifies the affected domain. Common terms such as dyslexia, dyscalculia, and dysgraphia are used to describe particular patterns of difficulty, and the manual accepts the first two as alternative terms for describing them. Dyslexia is the most common and most studied of these patterns. Researchers define it as difficulty with accurate or fluent word recognition and with spelling and decoding, typically resulting from a deficit in the phonological component of language, and unexpected in relation to the student’s other abilities and to the effective instruction he has received (Lyon et al., 2003).

**Comparison with ICD-11**

The International Classification of Diseases, Eleventh Revision (**ICD-11**), calls this category "developmental learning disorder" (**Developmental learning disorder**) and distinguishes subtypes within it: with impairment in reading, with impairment in written expression, with impairment in mathematics, and with other specified impairment of learning. It agrees with the DSM in requiring significant and persistent difficulty that emerges as the skills are taught in the early school years, and in excluding intellectual disability, sensory impairment, neurological and motor disorders, lack of availability of education, lack of proficiency in the language of instruction, and adverse psychosocial circumstances. It differs on one point of practical consequence: it describes performance in the affected skill as markedly below what would be expected for **both chronological age and the general level of intellectual functioning**, thereby keeping intellectual ability in the picture, whereas Criterion B of the DSM settles for a comparison with age (World Health Organization, 2024). We will see the effect of this difference in Yazan's case (8.8).

**For the specialist: The definition in US law and the three identification methods**

The US Individuals with Disabilities Education Act (**IDEA**) defines "specific learning disability" as a disorder in one or more of the basic psychological processes involved in understanding or using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. It excludes learning problems that are primarily the result of visual, hearing, or motor disabilities, of intellectual disability, of emotional disturbance, or of environmental, cultural, or economic disadvantage (IDEA, 2004; U.S. Department of Education, 2006, 34 CFR §300.8(c)(10)).

The implementing regulations specify **eight areas** in which the difficulty can be established: oral expression, listening comprehension, written expression, basic reading skills, reading fluency skills, reading comprehension, mathematics calculation, and mathematics problem solving (§300.309). They also set out **three methods of identification**: a state may not require a severe discrepancy between intellectual ability and achievement, must permit a process based on the student's response to scientific, research-based intervention, and may permit other alternative research-based procedures (§300.307). The third method usually encompasses **patterns of strengths and weaknesses** models, which the regulations describe as "a pattern of strengths and weaknesses in performance, achievement, or both, relative to age, State-approved grade-level standards, or intellectual development" (§300.309(a)(2)(ii)). Chapter 6 presents these three models and the critiques of them (6.2 to 6.5). Whatever the method, the team must confirm that the student received appropriate instruction in reading and mathematics delivered by qualified teachers, and that repeated data on his progress were collected (§300.309(b)).

**Common misconceptions**

- **"Dyslexia means the child sees letters backwards."** Reversing letters and their order is common among all children when they begin learning to write. The core problem in dyslexia is linguistic and phonological, in connecting sounds to letters, not visual. That is why the errors of a struggling reader of Arabic are mostly errors of sounds, vowel marks, and syllables, as in Ahmed's scene.
- **"The student is lazy; if he tried, he would read."** Many students with learning disabilities put in far more effort than their peers to reach a poorer result. What looks like laziness is often the avoidance of a painful task.
- **"He will grow out of it."** Learning disabilities do not disappear with waiting. Early intervention is far more successful than late intervention.
- **"Poor handwriting is a learning disability in writing."** Poor handwriting alone is not enough for this diagnosis, and its source may be motor rather than linguistic (see 8.4).

## 8.2 Diagnostic criteria

The **DSM-5-TR** requires four criteria, all of which must be met together (American Psychiatric Association [APA], 2022):

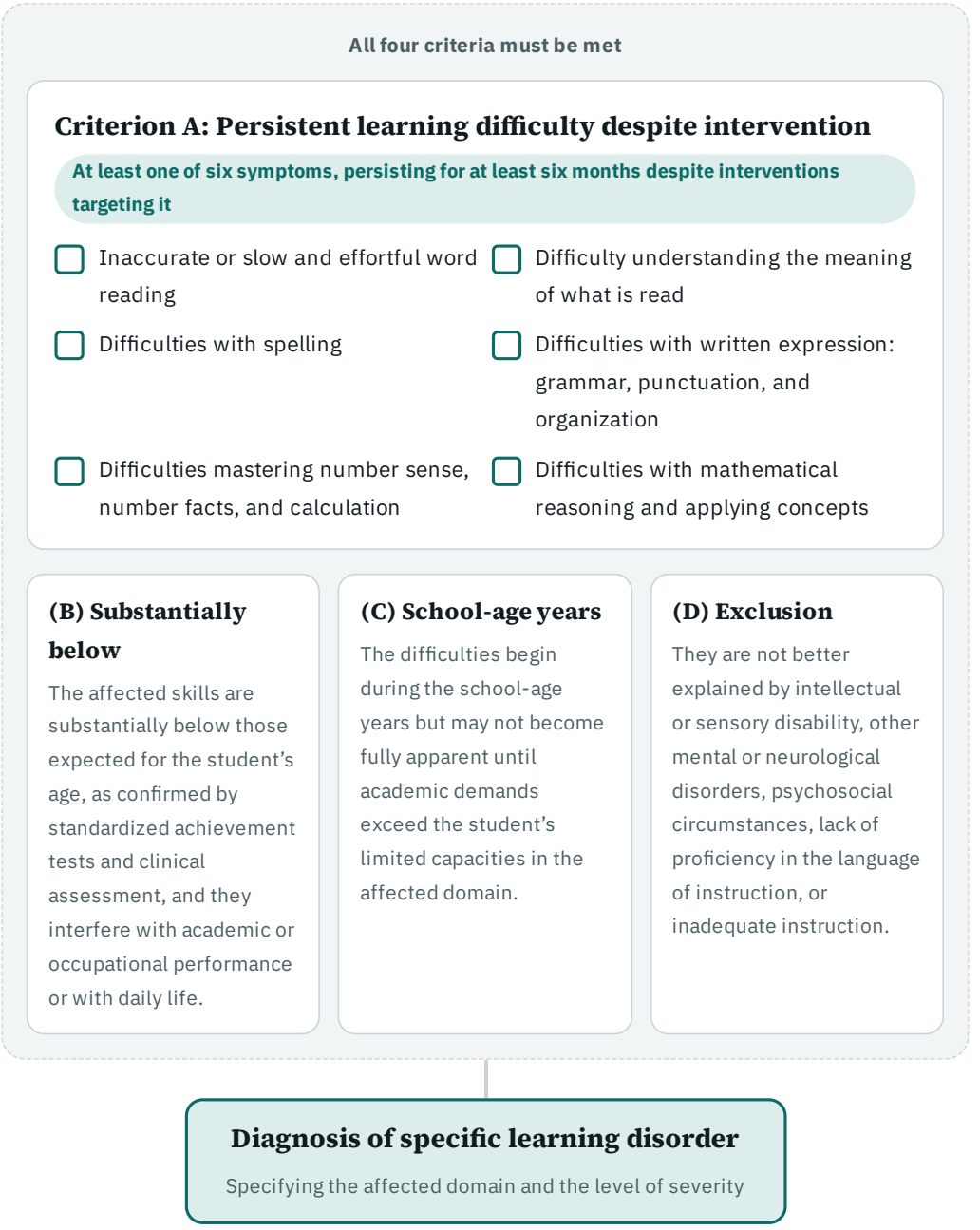



**Figure 8.1:** DSM-5-TR criteria for specific learning disorder. Note that Criterion A requires the difficulty to persist despite intervention; the diagnosis therefore presupposes that the student has first received appropriate instruction.

## What does “substantially below” mean?

For the greatest diagnostic certainty, the DSM suggests an achievement score of about 1.5 standard deviations or more below the mean, that is, about 78 or lower on a test with a mean of 100 and a standard deviation of 15. It does, however, allow a more lenient threshold, down to about one standard deviation, or roughly 85 or lower, when other evidence supports the difficulty, such as data showing nonresponse to intervention, clinical assessment, or school history. Note that the manual **does not require a discrepancy** between intelligence and achievement. The **discrepancy model** once required a large gap between an intelligence score and an achievement score, but it drew extensive criticism, including the charge that it waits for the gap to widen and so delays intervention (see Chapter 6, 6.2).

For individuals aged 17 and older, the manual allows a **documented history** of impairing learning difficulties to substitute for standardized assessment under Criterion B (APA, 2022). This matters in assessing university students and adults whose difficulties were never assessed at school, or for whose ages no standardized tools exist in their language. Such a history includes whatever documents the difficulty independently of the student’s own report: school reports, previous support services, or testing accommodations.

### Specifiers: domain and severity

**Table 8.1: Domains of specific learning disorder and their subskills in the DSM.**

| Area | Subskills specified in the diagnosis |
|---|---|
| **Reading** | Word reading accuracy, reading rate or fluency, and reading comprehension |
| **Written expression** | Spelling accuracy, grammar and punctuation accuracy, and clarity and organization of written expression |
| **Mathematics** | Number sense, memorization of arithmetic facts, accurate and fluent calculation, and accurate mathematical reasoning |

Note that the written expression skills in Table 8.1 do not include handwriting quality. Poor handwriting alone does not suffice for a diagnosis of specific learning disorder in written expression; its source must be sought, as Section 8.4 explains.

Severity is then specified: **mild**, where the student can compensate with accommodations and support; **moderate**, where he needs intensive, specialized teaching for part of his schooling; and **severe**, where he needs intensive, individualized teaching for most of his school years and may not perform the skills efficiently even with support.

## 8.3 Dyslexia and reading in Arabic

### The simple view of reading

One of the most useful frameworks for understanding reading difficulties is a model known as the **simple view of reading** (**Simple view of reading**). The model holds that reading comprehension is the product of two components multiplied together: **decoding**, the conversion of the written word into its spoken form, and **linguistic comprehension**, the understanding of words and

sentences when they are heard (Gough & Tunmer, 1986). Because the relation is multiplicative rather than additive, weakness in either component undermines reading comprehension, however strong the other may be. These two components give rise to four patterns:

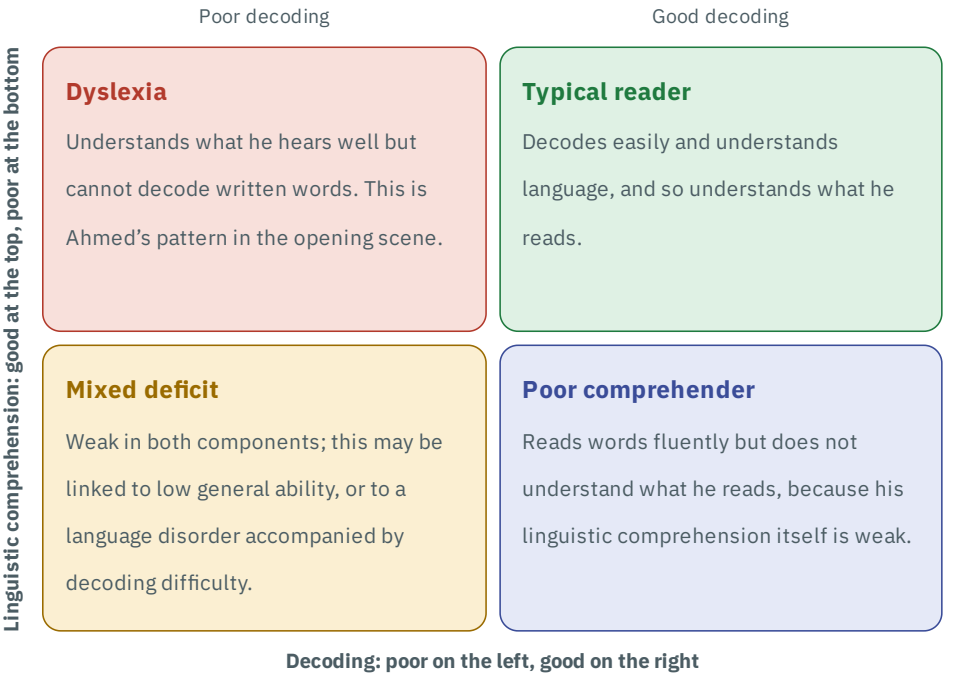



**Figure 8.2:** The four patterns of the simple view of reading. Each pattern points to a different likely diagnosis and a different intervention. A student weak in decoding needs systematic phonics instruction; a student weak in linguistic comprehension needs vocabulary development and work on oral comprehension.

The model bears directly on diagnosis. If a report shows low reading comprehension, both decoding and oral linguistic comprehension must be measured to establish which one is the source of the problem, and hence the type of difficulty.

## Phonological components

In most cases, dyslexia stems from a weakness in **phonological processing** (**Phonological processing**), that is, in the way the brain handles the sounds of language. It is measured in three areas:

- **Phonological awareness:** the ability to perceive and manipulate the sounds of a word, as when a child knows that "باب" (bāb, "door") begins with the sound /b/, deletes a sound from a word, or blends separate sounds into a word.
- **Phonological memory:** holding a sequence of sounds briefly in memory, as in repeating a nonsense word just heard.
- **Rapid naming (Rapid automatized naming, RAN):** the speed of naming a series of familiar symbols, such as colors, letters, or digits. Slowness here is associated with slow reading.

The importance of these components extends to Arabic. A study of Arabic-speaking children found that phonological processing skills contribute to word reading and reading comprehension in Arabic just as they do in other languages (Taibah & Haynes, 2011). Measuring them in Arabic, however, requires attention to the features of the language.

## What sets reading in Arabic apart?

Most diagnostic tools and research on dyslexia were built on English. Arabic differs from English in features that affect both how the difficulty appears and how it is measured (Elbeheri et al., 2006):

**The shift from vowelized to unvowelized text**

كَتَبَ → كتب

Children learn to read with fully vowelized texts in which the relationship between letter and sound is nearly regular; the vowel marks are then gradually withdrawn. Some students' difficulties surface when they are asked to infer the vowels from context (Abu-Rabia, 2001). Among the errors common in struggling beginners is confusing short vowels with long vowels, as in Ahmed's scene. Decoding should therefore be measured with both vowelized and unvowelized texts.

**Diglossia** (Diglossia)

"شنو" → "ماذا" (shinū → mādhā, "what")

Children learn to read in Modern Standard Arabic, which differs in its sounds and vocabulary from the colloquial dialect they speak. They are thus learning to read and, at the same time, learning what is almost a new language (Saiegh-Haddad, 2003). This makes linguistic comprehension in Standard Arabic a separate component of the assessment.

**Changing letter forms and dots**

ع ـعـ ـع · ب ت ث ن

A letter's shape changes with its position in the word, and many letters look alike, distinguished only by their dots. This adds a visual load early in learning.

**Morphological structure**

كتب · كاتب · مكتوب · مكتبة (kataba, kātib, maktūb, maktaba)

Arabic words are built from roots and patterns. Morphological awareness (Morphological awareness), the perception of a word's structure as a root and a pattern, helps the reader recognize unvowelized words, and weakness in it may contribute to the difficulty.

**Figure 8.3:** Four features of Arabic reading that any assessment of dyslexia in Arabic should take into account.

## 8.4 Difficulties in writing and mathematics

Reading receives most of the attention in research and practice, but a student may be referred because of writing or mathematics, and may have difficulties in more than one domain. Each domain raises its own diagnostic questions.

### Written expression and handwriting

Writing progresses from forming and copying letters, to spelling, to building sentences, paragraphs, and organized texts. By impairment in written expression, the DSM means three skills: spelling accuracy, grammar and punctuation accuracy, and clarity and organization of expression (APA, 2022). Poor handwriting and slow handwriting do not, on their own, suffice for this diagnosis, since they may arise from a motor rather than a linguistic source.

Hence the importance of **developmental coordination disorder** (**Developmental coordination disorder, DCD**) in differential diagnosis. It is a neurodevelopmental disorder in which the acquisition and execution of coordinated motor skills fall substantially below what is expected for the child's age and opportunities for learning, so that the child appears clumsy, slow, or imprecise in skills such as catching a ball, using scissors, and handwriting. The difficulty interferes with daily activities or academic productivity, begins in the early developmental period, and is not explained by intellectual disability, visual impairment, or a neurological condition affecting movement (APA, 2022).

A student with poor handwriting may be referred as a case of "writing difficulty," only for the assessment to reveal that his spelling and sentence structure are sound when he dictates his text orally or types it on a keyboard, and that his difficulty lies in fine motor control. Good assessment therefore compares three kinds of performance: copying a given text by hand, spelling from dictation, and free composition written by hand and then typed or

produced orally. An occupational therapist is brought in when a motor source is suspected. The two disorders may also co-occur in the same student, in which case both are assessed.

## Mathematics: number sense and arithmetic fluency

Learning mathematics begins with what is called **number sense** (**Number sense**): an intuitive grasp of quantities and their relationships, as when a child sees without counting that a small set has three objects, that 8 is clearly larger than 5, and that 49 is closer to 50 than to 40, and can place a number in its approximate position on a number line. A longitudinal study found that number competence in kindergarten predicts mathematics achievement years later (Jordan et al., 2009); early weakness in it is therefore a sign worth monitoring.

In its typical form, **dyscalculia** combines three features: weak number sense; difficulty retrieving arithmetic facts automatically, so that the student keeps counting on his fingers for problems his peers have long since memorized; and slow, inaccurate calculation. Mathematical reasoning may be relatively spared, so that the student chooses the right operation and then makes errors in carrying it out (Fletcher et al., 2019). The assessment therefore measures:

- **Number sense:** comparing two numbers, estimating on a number line, and perceiving small quantities without counting.
- **Arithmetic fluency** (**Arithmetic fluency**): accuracy and speed together in basic facts and operations. A student who arrives at the correct result only after lengthy counting has not achieved fluency, even if the answer is right.
- **Mathematical reasoning:** solving word problems, with the problems read aloud to the student so that reading difficulty is not confused with difficulty in mathematics.

**Math anxiety** is common among these students, and it may be a result of repeated failure rather than its cause. The assessment should therefore ask: does performance improve when the pressure is removed, for example when the student solves the same problems individually and untimed? If it does not, anxiety alone does not explain the weakness.

## 8.5 Early warning signs

**Table 8.2: Early warning signs in reading, writing, and mathematics by stage.**

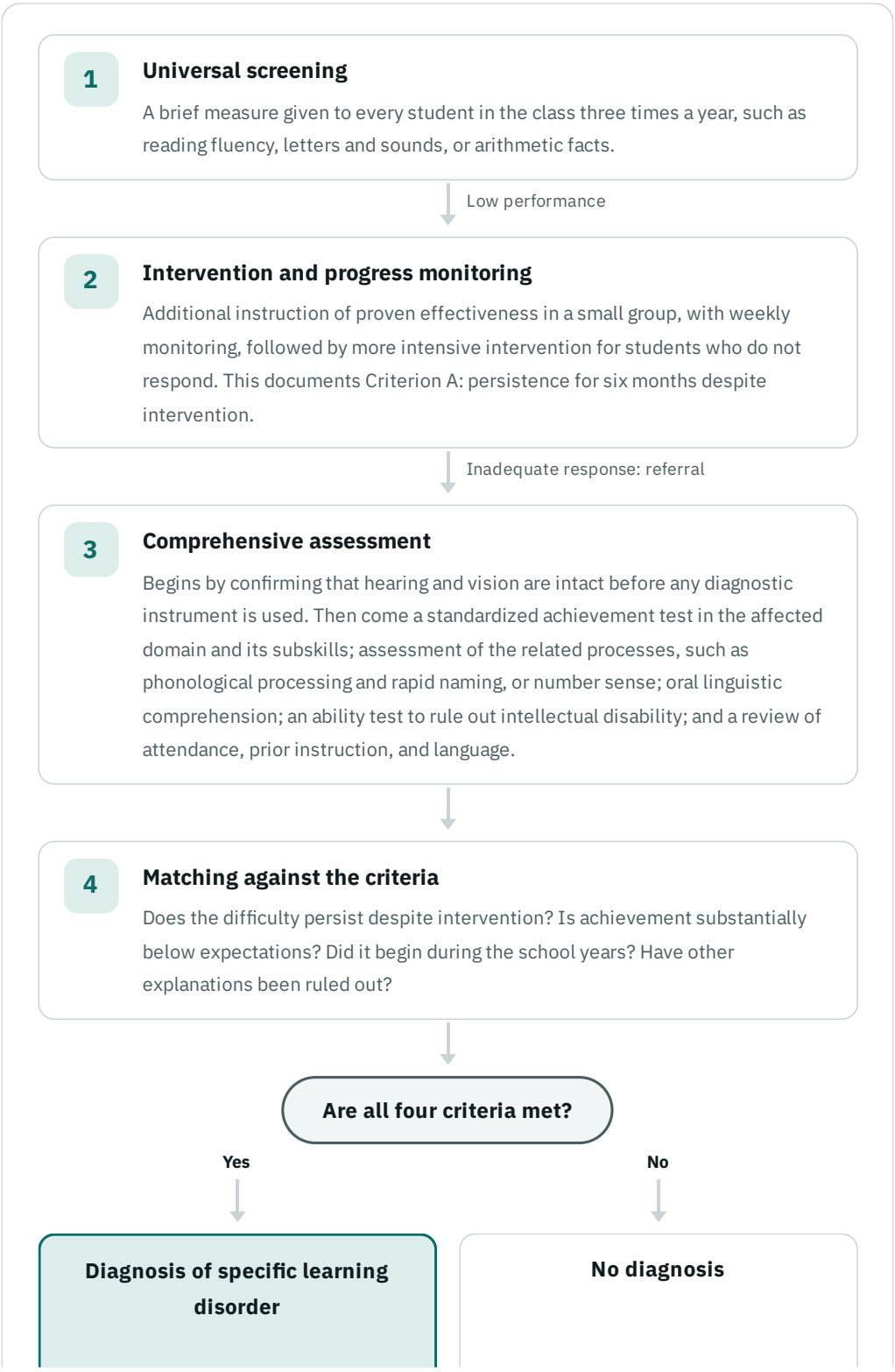
| Stage | In reading and writing | In mathematics |
|---|---|---|
| **Before school** | Difficulty distinguishing words that sound alike at the beginning or end, and difficulty learning letter names, together with a family history of reading difficulties | Difficulty counting in sequence, linking a number to a quantity, and comparing two small quantities |
| **Early grades** | Very slow progress in linking letters to sounds, persistent confusion between short and long vowels, guessing words from their first letters or from pictures, avoidance of reading aloud, and marked slowness in forming and copying letters | Continued finger counting for simple addition, difficulty memorizing basic facts, confusion of mathematical symbols, and difficulty placing numbers on a number line |
| **Upper elementary and middle school** | Slow, effortful reading, markedly weak spelling, avoidance of writing, short and poorly organized written texts, and comprehension that far outstrips reading | Difficulty with multistep problems, repeated procedural errors, slow fact retrieval that hampers learning fractions and algebra, and anxiety about mathematics |

**For the teacher: The question that reveals a great deal**

Read a passage from the student's textbook aloud to him, then ask him about it. Next, have him read a similar passage on his own, and ask him about that one. If he understood what he heard well but did not understand what he read because he stumbled over the words, the problem most likely lies in decoding. If he understood neither, the problem may lie in linguistic comprehension or in general ability. This simple check is not enough for a diagnosis, but it applies the simple view of reading in two minutes. Its counterpart in mathematics is to read a word problem to the student and ask, "What will you do to solve it?" before he calculates; you will then know whether he is stumbling over reasoning or over calculation.

## 8.6 The decision pathway

This pathway follows a hybrid model that combines three kinds of evidence (Fletcher et al., 2019): inadequate response to intervention, meaning that the student has not progressed as he should despite receiving additional instruction of proven effectiveness, which is the heart of the response-to-intervention model; low achievement; and the exclusion of other explanations (see Chapter 6, 6.3 and 6.5). The first kind of evidence is usually gathered through curriculum-based measurement (see Chapter 4, 4.4).


1
Universal screening
A brief measure given to every student in the class three times a year, such as reading fluency, letters and sounds, or arithmetic facts.
Low performance
2
Intervention and progress monitoring
Additional instruction of proven effectiveness in a small group, with weekly monitoring, followed by more intensive intervention for students who do not respond. This documents Criterion A: persistence for six months despite intervention.
Inadequate response: referral
3
Comprehensive assessment
Begins by confirming that hearing and vision are intact before any diagnostic instrument is used. Then come a standardized achievement test in the affected domain and its subskills; assessment of the related processes, such as phonological processing and rapid naming, or number sense; oral linguistic comprehension; an ability test to rule out intellectual disability; and a review of attendance, prior instruction, and language.
4
Matching against the criteria
Does the difficulty persist despite intervention? Is achievement substantially below expectations? Did it begin during the school years? Have other explanations been ruled out?
Are all four criteria met?
Yes
No
Diagnosis of specific learning disorder
No diagnosis

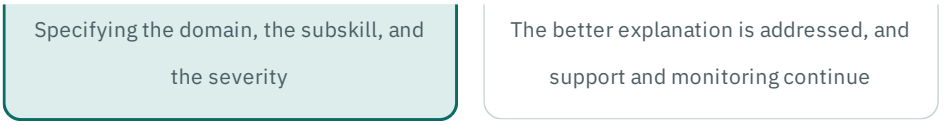



**Figure 8.4:** The decision pathway for learning disabilities. The first and second stages take place in the classroom before referral and generate the evidence that Criterion A requires. Intervention must never become a reason to delay an assessment that the family requests or that the evidence points to.

## 8.7 Instrument profiles

Learning disabilities call for tools that measure the affected skill precisely and that also measure the processes behind it. Here we present six international tools covering reading, mathematics, and writing, with the caveat that all of them are built on English and are therefore suitable for Arab students only in adapted and standardized versions, or through Arabic tools built on the same idea. The aim of presenting them is for readers to know what should be measured and how it should be interpreted, and to know what to ask of any Arabic tool.

The profiles follow a single template. Each profile's table has fields for reliability, classification accuracy, and the year the norms were collected. Where we have not verified a figure in the technical manual itself, we have written "In the technical manual," and we have not carried over figures from secondary sources. Examiners should read these values from the manual before relying on the tool (see Chapter 2). This book does not publish the items of any test; the examples in the profiles were prepared by the author to illustrate the type of task.

### CTOPP-2 Comprehensive Test of Phonological Processing, Second Edition

Phonological processing

### The idea

When a student struggles with decoding, we need to know the reason for his weakness as well as its extent. This test measures the three aspects of phonological processing to which dyslexia is usually traced: phonological awareness, phonological memory, and rapid naming (Wagner et al., 2013).

### What it looks like

An examiner's booklet, rapid naming plates, audio recordings for some tasks, and a record form. It has two age versions: the first for ages 4 to 6, the second for ages 7 to 24. Its task types include deleting a sound from a word and saying what remains, blending separate sounds heard on a recording into a word, repeating digits or nonsense words, and naming rows of digits, letters, colors, or objects as quickly as possible.

### A scene from administration

The examiner places in front of a third grader a plate printed with rows of familiar letters and says, "Name the letters out loud, from the start of each row to the end, as fast as you can," then starts the stopwatch. The student names every letter correctly but takes nearly twice the usual time. The examiner records both the time and the errors.

**Table 8.3: Comprehensive Test of Phonological Processing (CTOPP-2) at a glance.**

| | |
|---|---|
| **Purpose** | Identifying the source of phonological weakness and guiding intervention, as well as supporting diagnosis |
| **Ages** | 4 to 24 years, in two age-based forms |
| **Edition and year** | Second, 2013 |
| **Publisher** | PRO-ED |
| **Qualification level** | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements |

| | |
|---|---|
| **Time** | About 40 minutes |
| **Score types** | Scaled scores for the subtests (mean 10, SD 3), and composite scores for phonological awareness, phonological memory, and rapid naming (mean 100, SD 15) |
| **Cutoff scores** | No diagnostic cut-off; scores are interpreted by descriptive range and in comparison with reading performance |
| **Reliability** | In the technical manual |
| **Classification accuracy** | Not designed for classification on its own |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | We know of no published standardized Arabic version at the time of writing. Arabic tasks built on the same idea have been used in research (Taibah & Haynes, 2011), but they must not be interpreted against the original norms |

### How are the results interpreted?

The profile is read in light of reading performance. Low phonological awareness together with weak decoding of nonsense words points to a core phonological weakness that calls for systematic phonics instruction. Low rapid naming on its own, with acceptable accuracy, points to a problem of fluency more than of accuracy. The combination of the two weaknesses, sometimes called the double deficit, is generally associated with more severe difficulty.

### Its limits

It was built on the sounds and syllables of English, so translating its items does not yield an Arabic tool: a translated word differs in its number of sounds and its syllable structure, and its difficulty changes accordingly. A low score

indicates weakness in the process, and it establishes a learning disability only when it coincides with weakness in reading itself.

## TOWRE-2 Test of Word Reading Efficiency, Second Edition

Decoding fluency

### The idea

A skilled reader reads words accurately, quickly, and effortlessly. This brief test measures accuracy and speed together, and it separates the recognition of familiar words from pure decoding (Torgesen et al., 2012).

### What it looks like

A card with a list of real words of increasing difficulty, which the student reads as quickly as he can in 45 seconds, and a second card with a list of nonsense words (**Pseudowords**), which he reads in another 45 seconds. The test has four equivalent forms, so it can be repeated to monitor progress without the student remembering the list.

### A scene from administration

A fifth grader reads the list of real words at an acceptable pace, then slows down considerably on the list of nonsense words, turning some of them into real words that resemble them. The examiner records the number of words read correctly on each list and notes the types of errors in the margin.

**Table 8.4: Test of Word Reading Efficiency (TOWRE-2) at a glance.**

| | |
|---|---|
| **Purpose** | Measuring word reading fluency and decoding, and progress monitoring through its equivalent forms |
| **Ages** | 6 to 24 years |
| **Edition and year** | Second, 2012 |

| | |
|---|---|
| **Publisher** | PRO-ED |
| **Qualification level** | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements |
| **Time** | About 5 to 10 minutes |
| **Score types** | Standard scores for the two subtests and for a total word reading efficiency index (mean 100, SD 15) |
| **Cutoff scores** | No diagnostic cut-off |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | We know of no published standardized Arabic version at the time of writing. The closest we found is a measure of Arabic word recognition fluency for monitoring the progress of first graders (Abu-Hamour, 2014b), which is a monitoring tool, not a diagnostic test |

### How are the results interpreted?

A student who reads real words acceptably but stumbles a great deal over nonsense words usually relies on memorization and guessing and lacks decoding skill. Low scores on both subtests, by contrast, indicate a general weakness in reading fluency.

### Its limits

It measures the reading of isolated words, not texts, so it cannot replace the measurement of fluency in connected text or of comprehension. The task is easy to construct in Arabic, using vowelized nonsense words that follow the phonological rules of Arabic, but a locally built tool needs norms before its scores can be interpreted normatively.

## WIAT-4 Wechsler Individual Achievement Test, Fourth Edition

Comprehensive achievement

### The idea

Criterion B requires standardized measurement of achievement in the affected domain and its subskills. This battery measures reading, writing, mathematics, and oral language against a single set of norms, which allows the domains to be compared with one another (Breaux, 2020).

### What it looks like

Two stimulus books, a record form, and a response booklet in which the student writes. The examiner selects subtests according to the referral question, so not all of them need to be given. Its value for reading difficulties is that it measures the two components of the simple view of reading and their product: decoding, listening comprehension, and reading comprehension. It also includes subtests of math fact fluency and written expression.

**Table 8.5: Wechsler Individual Achievement Test (WIAT-4) at a glance.**

| | |
|---|---|
| **Purpose** | Standardized measurement of achievement in academic domains and their subskills |
| **Ages** | 4 years to 50 years 11 months |

| | |
|---|---|
| **Edition and year** | Fourth, 2020 |
| **Publisher** | Pearson |
| **Qualification level** | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements |
| **Time** | Varies with the number of subtests selected |
| **Score types** | Standard scores (mean 100, SD 15); percentile ranks, that is, the percentage of the student's peers in the norming sample who scored below him, so that a rank of 84 means he outperformed 84% of them; and composite scores for the domains |
| **Cutoff scores** | No cut-off; the Criterion B thresholds are applied |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The standard score for each subskill is read against the Criterion B thresholds, and the subskills are then read together: low decoding with intact listening comprehension is the dyslexia pattern, whereas low scores on both call for a look at language or general ability.

### Limitations

It is built on English and English-language curricula, so it should be used with Arab students only in an adapted and standardized version. A single score on a short subtest is less reliable than a composite score, so no judgment should rest on it alone.

## KeyMath-3 KeyMath-3 Diagnostic Assessment Mathematics

### The idea

A single score in "mathematics" does not tell us where the student is stumbling. This test divides mathematics into areas and subskills, pinpointing the difficulty precisely enough to guide intervention (Connolly, 2007).

### What it looks like

Two stimulus books from which the examiner presents the problems, and a record form. The subtests are grouped into three areas: **basic concepts**, such as numeration, algebra, geometry, measurement, and data analysis; **operations**, such as mental computation, addition and subtraction, and multiplication and division; and **applications**, such as problem solving. It has two parallel forms.

### A scene from administration

The examiner presents the mental computation problems to a fourth-grade girl one by one and notices that she moves her fingers under the table on every simple addition problem. He notes this in the margin of the record form, because the score alone will not preserve the observation.

**Table 8.6: KeyMath-3 Diagnostic Assessment (KeyMath-3) at a glance.**

| | |
|---|---|
| Purpose | Diagnosing areas of strength and weakness in mathematics and guiding intervention |

| | |
|---|---|
| **Ages** | 4 years 6 months to about 21 years |
| **Edition and year** | Third, 2007 |
| **Publisher** | Pearson |
| **Qualification level** | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements |
| **Time** | Approximately 30 to 90 minutes, depending on age |
| **Score types** | Scaled scores for subtests (mean 10, SD 3), and standard scores for the domains and the total score (mean 100, SD 15) |
| **Cutoff scores** | No cut-off; the Criterion B thresholds are applied |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The examiner compares the three domains: low scores in operations and basic concepts alongside relatively intact applications fit the pattern of dyscalculia, whereas low applications alone may reflect weakness in reading or language. The test is complemented by error analysis, that is, examining the student's errors to identify the faulty procedure they are following (see Chapter 4, 4.5).

### Its limits

It was built on the US mathematics curriculum and its sequence, and some of its items are tied to US currency and measurement, so it cannot simply be translated. Nor does it measure timed arithmetic fluency, so a fluency measure must accompany it.

## TEMA-3 Test of Early Mathematics Ability, Third Edition

Number sense

### The idea

Before mathematics difficulty shows up in written calculation, it shows up in the understanding of number and quantity. This test measures both sides of early mathematics: the **informal** knowledge a child acquires before school, such as counting, comparing quantities, and calculating with objects, and the **formal** knowledge learned at school, such as reading and writing numerals and number facts (Ginsburg & Baroody, 2003).

### What it looks like

A picture book of stimuli, small objects for counting, and a record booklet. The examiner begins at an item suited to the child's age and continues until the discontinue point is reached. Typical tasks include saying which of two sets has more objects, counting a set and stating how many it contains, and reading written numerals. Two equivalent forms are available.

### A scene from administration

The examiner places two cards in front of a first grader, one with four dots and the other with seven, and asks, "Which has more?" The child counts the dots on both cards one by one before answering, although his peers answer at a glance. The examiner records the correct answer and notes how he arrived at it.

**Table 8.7: Profile of the Test of Early Mathematics Ability (TEMA-3).**

| | |
|---|---|
| **Purpose** | Measuring number sense and early mathematics, and identifying children at risk for mathematics difficulties |
| **Ages** | 3 years to 8 years 11 months |
| **Edition and year** | Third, 2003 |
| **Publisher** | PRO-ED |
| **Qualification level** | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements |
| **Time** | About 40 minutes |
| **Score types** | A Math Ability standard score (mean 100, SD 15) and percentile ranks |
| **Cutoff scores** | No diagnostic cut-off |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | Reported in the technical manual; its norms are relatively old |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The standard score locates the child among peers, but item analysis matters more: is the child weak on informal tasks, such as comparing quantities and counting, which signals weak number sense? Or only on formal tasks, such as reading and writing numerals, which may reflect a lack of instruction?

### Its limits

It is not suitable beyond 8 years 11 months, so older children need another test along with informal number-sense tasks. Its norms are dated, and dated norms can inflate a child's standing relative to today's peers. Number-sense tasks, such as comparing quantities and estimating on a number line, depend less on language than other tasks, which makes them easier to build in Arabic, but they too require local norms.

## TOWL-4 Test of Written Language, Fourth Edition

Written expression

### The idea

Writing is the most complex of the academic skills, because it draws on spelling, grammar, punctuation, vocabulary, and the organization of ideas all at once. This test measures it in two ways: **contrived** tasks, each targeting a single skill, and a **spontaneous writing sample** in which the student composes a complete text (Hammill & Larsen, 2009).

### What it looks like

A response booklet in which the student writes, a stimulus picture for story writing, and a scoring manual. Tasks cover vocabulary, spelling, punctuation, constructing logical sentences, and combining short sentences into one. The student then writes a story about the picture after a brief planning period, and the story is scored twice: once for the accuracy of spelling, punctuation, and grammar in context, and once for story construction and organization.

**Table 8.8: Profile of the Test of Written Language (TOWL-4).**

| | |
|---|---|
| Purpose | Measuring the components of written expression and pinpointing areas of weakness |
| Ages | 9 years to 17 years 11 months |

| | |
|---|---|
| **Edition and year** | Fourth, 2009 |
| **Publisher** | PRO-ED |
| **Qualification level** | An examiner trained in administering standardized tests and scoring writing samples, in accordance with the publisher's requirements |
| **Time** | About 60 to 90 minutes; parts can be administered to groups |
| **Score types** | Scaled scores for subtests (mean 10, SD 3), and composite scores for contrived, spontaneous, and overall writing (mean 100, SD 15) |
| **Cutoff scores** | No cut-off; the Criterion B thresholds are applied |
| **Reliability** | Reported in the technical manual; the reliability of story scoring depends on scorer training |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

Comparing the contrived tasks with spontaneous writing is informative: a student who combines sentences and punctuates correctly when asked for one skill at a time, but then neglects these skills in the story, knows the rules but does not apply them under the demands of free writing. A very short, poorly constructed story alongside intact contrived tasks, by contrast, calls for a closer look at planning and organization, or at slow handwriting itself.

### Its limits

It does not directly measure handwriting or handwriting speed, so it cannot by itself distinguish a difficulty in written expression from developmental coordination disorder (8.4). Its spelling, punctuation, and syntax rules are English, so translating it makes no sense. The practical alternative in Arabic remains curriculum-based writing samples scored with an analytic rubric (see Chapter 4).

⚠ **When no standardized Arabic instrument is available**

In this situation, the team gathers evidence from multiple sources: curriculum-based measurement data on reading or arithmetic fluency (see Chapter 4, 4.4) compared against local norms built by the school; decoding, phonological awareness, and number-sense tasks constructed in Arabic; error analysis; and response-to-intervention data. The report states plainly that the judgment on Criterion B rested on this evidence in the absence of a standardized test, and it spells out the limits of that approach. Section 8.10 presents the Arabic instruments we found in the published literature.

## 8.8 When is the judgment made? Four cases

The scores in the following cases are hypothetical and intended for teaching. They assume that the team used instruments standardized for Arabic that measure what the instruments in Section 8.7 measure. This is a teaching assumption; what is actually available is set out in Section 8.10.

## Case 1: Ahmed, 8 years 6 months

Clear dyslexia

**The story:** Ahmed is the student from the opening scene. Universal screening at the start of second grade showed that his reading fluency fell below that of 90% of his classmates. He received small-group intervention for three months, then intensive individual intervention for another three months, yet his fluency grew at a very slow rate. This is what is called a **dual discrepancy** (**Dual discrepancy**): performing below peers in both level of performance and rate of growth. Most of his oral reading errors involve omitting long vowel letters and substituting short vowels.

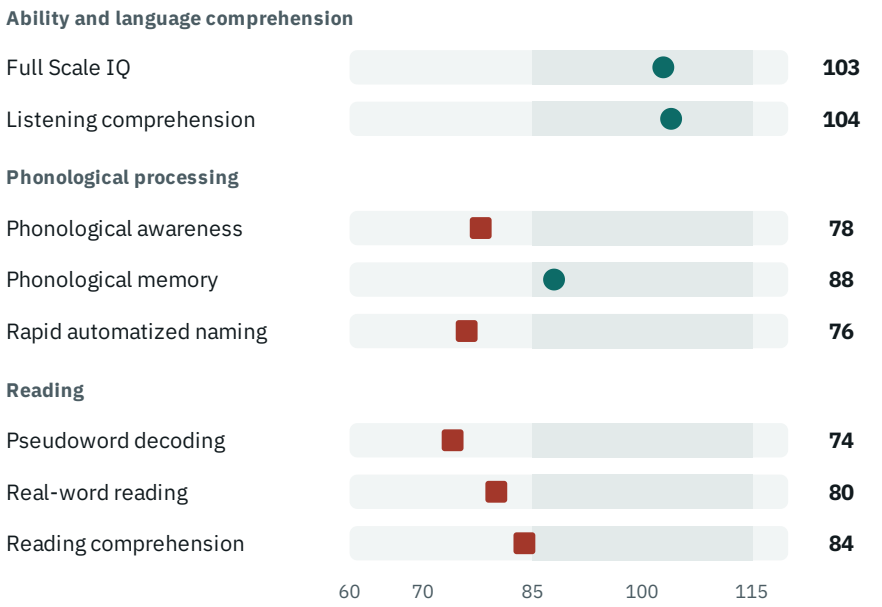



**Figure 8.5:** Ahmed's profile with hypothetical values. The gray band marks the range from 85 to 115, and square dots mark scores below 85. Ability and oral comprehension are in the average range, while decoding, phonological awareness, and rapid naming are low. This is the dyslexia pattern shown in Figure 8.2.

| A ✓ | B ✓ | C ✓ | D ✓ |
|---|---|---|---|
| Persistent for six months despite | Decoding 74 | Since first grade | Average intelligence, intact hearing and |

| | | | |
|---|---|---|---|
| intervention | | | vision, regular attendance |

**Decision:** Ahmed is diagnosed with specific learning disorder with impairment in reading, affecting word reading accuracy and fluency, at a moderate level of severity. The recommendations are systematic, intensive phonics instruction beginning with short vowels, long vowel letters, and syllables, along with accommodations that give him access to content, such as having test questions read aloud to him in non-language subjects, so that he is not denied the chance to show his understanding of science, the very understanding his teacher had noticed.

> **What evidence could change the decision?** If his fluency growth accelerated clearly with systematic phonics instruction until he approached his peers, the severity level would need to be revisited; and if a weakness in listening comprehension emerged when measured with longer texts in Standard Arabic, a co-occurring language disorder would have to be considered.

## Case 2: Lama, 9 years

She responded to instruction

**The story:** Lama moved among three schools in two years because of her father's work, and she was absent for long stretches in second grade owing to family circumstances. Screening showed her reading to be very low, and her mother requested a diagnostic evaluation.

**Table 8.9: Evidence in Lama's case.**

| Source | Result | What it means |
|---|---|---|
| **Screening at the start of the year** | Her fluency fell below that of 95% of her classmates | Needs urgent support |

| Source | Result | What it means |
|---|---|---|
| **Small-group intervention for 12 weeks** | Her fluency grew at twice the rate of her classmates and surpassed the aimline, that is, the expected trajectory set for her progress | Strong response to instruction |
| **Records from previous schools** | Frequent absences and changes in curricula and teachers | Limited opportunity for consistent learning |
| **Phonological awareness tasks** | Average | No core phonological weakness |

| A ✗ | B | C | D ✗ |
|---|---|---|---|
| Responded quickly | Improving | Not applicable | Interrupted schooling explains the weakness |

**Decision:** Lama is not diagnosed with a learning disability. Her weakness is explained by a lack of consistent instruction, and she responded quickly once she received it. She continues to receive support until she catches up with her classmates, with her progress monitored. If her progress later stalls despite support, her case will be reconsidered. Had support been postponed until a formal diagnosis was complete, Lama would have lost months that were enough to remedy her weakness.

**The lesson**

Had Lama been assessed with tests alone at the start of the year, her scores would have been very low, and she might have been misdiagnosed. It was the response-to-intervention data that revealed her real need: consistent instruction.

**What evidence could change the decision?** If her growth slowed or stalled after she settled in one school and attended regularly, or if more fine-grained tasks revealed a weakness in decoding or rapid naming, the lack-of-instruction explanation would weaken, and a comprehensive evaluation would need to be completed.

## Case 3: Dima, 9 years 10 months Dyscalculia

**The story:** Dima is in fourth grade. Her teacher reports that she participates in science and reading lessons and can explain orally how she would solve a word problem, yet she counts 7 + 5 on her fingers, takes twice as long as her classmates on an arithmetic worksheet, and makes many errors. Her mother says she "hates math" and cries before quizzes, and that this began in first grade.

**Table 8.10: Evidence in Dima's case.**

| Source | Result | What it means |
|---|---|---|
| **Universal screening in arithmetic (curriculum-based measurement, 2 minutes)** | Below the 10th percentile among her classmates in third and fourth grades | A persistent, not a passing, weakness |

| Source | Result | What it means |
|---|---|---|
| **Intervention in two phases: 12 weeks in a small group in third grade, then 14 weeks of intensive individual intervention in fourth grade** | Her fluency grew from 8 to 13 correct digits in two minutes, against a goal of 22. Her growth rate is less than half the expected rate | Inadequate response despite more than six months of intervention (dual discrepancy) |
| **Number-sense tasks** | She counts dots one by one even in sets of three, is very slow at comparing two-digit numbers, and places 70 near the middle of a 0–100 line | Weak representation of quantity |
| **Word problems read aloud to her** | She chooses the correct operation on most of them but errs in the calculation itself | Reasoning stronger than calculation |
| **The same facts in a calm, untimed individual session** | She still counts on her fingers, and her accuracy barely improves | Anxiety co-occurs but does not explain the difficulty |
| **Hearing and vision, attendance, and home language** | Hearing and vision intact, regular attendance, and Arabic as her first language | No apparent alternative explanation |

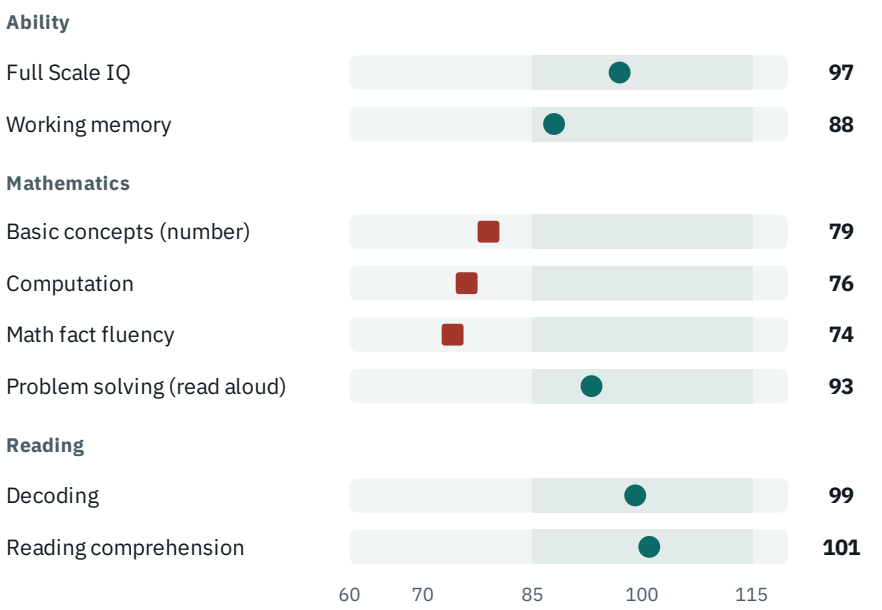



**Figure 8.6:** Dima's profile with hypothetical values. The gray band marks the range from 85 to 115, and square dots mark scores below 85. The weakness is confined to number, calculation, and calculation fluency, with general ability, reading, and mathematical reasoning in the average range. This is the dyscalculia pattern described in 8.4.

| A ✓ | B ✓ | C ✓ | D ✓ |
|---|---|---|---|
| Persistent for more than six months despite intervention | Computation 76, fluency 74 | Since first grade | Average ability, intact senses, consistent instruction; anxiety does not explain the weakness |

**Decision:** Dima is diagnosed with specific learning disorder with impairment in mathematics, affecting number sense, memorization of number facts, and accurate and fluent calculation, with relatively intact mathematical reasoning, at a moderate level of severity. The recommendations are explicit, systematic instruction in number sense, the number line, and fact-derivation strategies, such as working out 7 + 5 from 7 + 3 = 10 and then adding 2, together with brief daily fluency

practice. Accommodations include fewer timed tests and permission to use a fact chart or calculator on reasoning tasks, while instruction in the facts themselves continues.

**The lesson**

It would have been easy to attribute Dima's weakness to anxiety, since she cries before tests. But her performance did not improve when the pressure was removed, so the anxiety here is more a result of years of failure than its cause. Addressing it nonetheless remains part of the plan.

**What evidence could change the decision?** If her accuracy and fluency improved clearly once her anxiety was addressed, or if observation across different settings revealed inconsistent errors linked more to attention than to number, the team would need to revisit the diagnosis and consider an alternative or co-occurring explanation.

## Case 4: Yazan, 11 years

A difficulty masked by giftedness

**The story:** Yazan has been in the gifted program since third grade and solves mathematics problems years beyond his grade level. But his teacher in the program noticed that he refuses to write any report, spends a very long time reading texts, and always asks to explain orally instead. His school grades in Arabic are average, and this caught no one's attention because he is "not weak." His teacher referred him to the learning disabilities teacher at the start of fifth grade.

**Table 8.11: Evidence in Yazan's case.**

| Source | Result | What it means |
|---|---|---|
| **Intervention in fifth grade** | 12 weeks in a small group for reading fluency and spelling, then 12 weeks of intensive individual intervention. His fluency grew at less than half the target rate, and his spelling errors decreased only slightly | Limited response despite six months of intervention |
| **Intelligence test** | Full Scale IQ 128, Verbal Comprehension 132, Processing Speed 92, and **General Ability Index** (**General Ability Index, GAI**) 133 | Very high ability. The full-scale score includes his weak processing speed and so underestimates his ability; the General Ability Index, which excludes the working memory and processing speed indexes, is the more accurate reference |
| **Untimed reading comprehension** | 106 | He understands in the end, but with great effort and time |
| **Reading fluency, and decoding on a timed task** | 82 and 85 | Below average or at its lower bound, and far below his ability |
| **Spelling** | 80 | More than one standard deviation below the mean |

| Source | Result | What it means |
|---|---|---|
| **History** | Delayed in learning to read in first grade; he compensated by memorizing words and through his strong comprehension | The difficulty appeared early but was masked |

| A ✓ | B ✓ | C ✓ | D ✓ |
|---|---|---|---|
| Persistent for six months despite intervention | Spelling 80, fluency 82: the less stringent threshold, with supporting evidence | Since first grade | High ability, intact senses, consistent instruction |

**How did the team read the evidence?** Yazan's school grades and his untimed reading comprehension were in the average range, so he did not show up on the school's usual indicators. But when fluency and spelling were measured with standardized instruments, both fell more than one standard deviation below the mean. This meets the less stringent threshold for Criterion B, because it is accompanied by supporting evidence: a limited response to six months of intervention, a history of early difficulty masked by compensation, and a clear impact on his functioning, as he avoids writing and needs twice the usual time.

**Decision:** Yazan is diagnosed with specific learning disorder with impairment in reading (fluency) and in written expression (spelling accuracy), at a mild level of severity. He is therefore a **twice-exceptional** (**Twice-exceptional**) student, one who combines giftedness with a disability or learning difficulty. He remains in the gifted program and receives accommodations such as extended time and writing on a computer, along with direct instruction in fluency and spelling.

**The point of tension:** What prompted the team to look in the first place was the gap between Yazan's ability and his achievement, that is, the ability-comparison logic underlying the discrepancy model and the patterns of strengths and weaknesses models. The response-to-

intervention model, by contrast, compares the student with peers, so it may not see a student who compensates through ability until compensation fails. This is a well-known tension in identifying twice-exceptional learners: some researchers argue that analyzing cognitive ability profiles adds little to the accuracy of decisions (McGill et al., 2018), whereas specialists in twice-exceptionality argue that removing ability from the picture removes these students from identification. ICD-11 itself compares performance with the level of general intellectual functioning, as we saw in 8.1. The team therefore based its decision on all four criteria being met, and used the gap to explain why the difficulty was detected late. Chapter 6 presents the three models (6.2, 6.3, and 6.4), and Chapter 13 takes up the debate about twice-exceptional learners.

> **What evidence could change the decision?** If fluency and spelling came out in the average range when remeasured with another instrument with more recent norms, or if his avoidance of writing turned out to stem from a fine motor difficulty, the diagnosis in written expression, or in both areas, would need to be revisited.

## 8.9 Differential diagnosis and co-occurrence

### Similar conditions

**Table 8.12: Differential diagnosis of learning disabilities.**

| Condition | Similarity | Distinguishing features |
|---|---|---|
| **Intellectual disability** | Low achievement | Difficulties across all areas, with low general ability and adaptive behavior |
| **Lack of instruction or interrupted schooling** | Very low achievement | Rapid response to consistent instruction, as in Lama's case |

| Condition | Similarity | Distinguishing features |
|---|---|---|
| **Second-language learning** | Weak reading and writing in the language of instruction | Skills in the home language are stronger, and progress follows the trajectory of second-language learners |
| **Hearing or visual impairment** | Difficulty with reading and spelling | The picture improves when the sensory impairment is corrected or compensated for |
| **Attention-deficit/hyperactivity disorder (ADHD)** | Errors in reading and arithmetic, and incomplete tasks | Errors are inconsistent and tied to attention rather than to a specific skill. The two disorders frequently co-occur, so both are assessed |
| **—** | Weak reading comprehension | Oral comprehension itself is weak, i.e., the poor-comprehender pattern in Figure 8.2 |
| **Developmental coordination disorder (DCD)** | Poor handwriting, slow writing, and avoidance of writing | Spelling and syntax are intact when the student dictates orally or types on a keyboard, and the motor difficulty appears in nonacademic tasks such as ball games and using scissors |
| **Anxiety and emotional problems, including math anxiety** | Avoidance and lowered performance | Performance varies considerably between comfortable and stressful situations and improves when the pressure is removed |

## Co-occurring condition

The conditions in Table 8.12 are not always alternatives; many of them co-occur with learning disabilities in the same student. The clearest example is ADHD: published reviews report that a large proportion of individuals with one disorder also have the other, with estimates varying across samples and criteria (DuPaul et al., 2013). Likewise, dyslexia and dyscalculia co-occur more often than chance would predict, and weak reading comprehension overlaps with developmental language disorder (Fletcher et al., 2019). Knowing that two disorders may co-occur is a first step; each combination then calls for an assessment procedure that keeps one from masking or inflating the other:

- **With ADHD:** Attention ratings are gathered from at least two settings, and the examiner observes whether errors on the same skill fluctuate or remain stable. If treatment for attention begins, the academic skill is remeasured afterward; whatever weakness remains is more likely to reflect a learning disability.
- **Between reading and mathematics:** Both domains are assessed whenever a student is referred in either one, and mathematics word problems are read aloud to the student so that a reading weakness is not counted as a weakness in mathematics.
- **With developmental language disorder:** Oral comprehension and vocabulary are measured in both Standard Arabic and the student's dialect, and a speech-language pathologist is involved when oral comprehension is weak (see Chapter 11).
- **With developmental coordination disorder:** Handwritten expression is compared with dictated or keyboarded expression, and an occupational therapist is involved.
- **With anxiety:** Performance under calm, untimed conditions is compared with usual performance, and the student is asked about their experience with the subject, as in Dima's case.

# 8.10 In the Saudi and Arab context

**In the Saudi context: How is a student with learning disabilities identified?**

The Regulations for Special Education Institutes and Programs (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]) set out the conditions for admitting a pupil to learning disabilities programs in Article 7, and the Special Education Organizational Guide retained these conditions in closely similar wording (Ministry of Education, 1436 AH [Arabic]). Table 8.13 compares this framework with the two diagnostic manuals and with US law.

**Table 8.13: Learning disabilities in the diagnostic manuals, US law, and Saudi regulations.**

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Name and definition** | Specific learning disorder (DSM-5-TR) and developmental learning disorder (ICD-11): persistent difficulty learning foundational academic skills | Specific learning disability: a disorder in one or more of the basic psychological processes, manifested in listening, thinking, speaking, reading, writing, spelling, or calculating | Learning disabilities: "disorders in one or more of the basic psychological processes" that are not due to mental, hearing, visual, or other impairment, nor to learning conditions or family care (the Regulations, Article 1) |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Identification logic** | Clearly low achievement with difficulty persisting despite intervention, with no discrepancy requirement in DSM-5-TR; performance below what is expected for age and level of general intellectual functioning in ICD-11 | A severe discrepancy may not be required; response to intervention must be permitted; and alternative research-based procedures may be permitted | "A clear discrepancy between the level of his abilities and the level of his academic achievement" in one of eight areas (oral expression, listening, oral comprehension, writing, reading, reading comprehension, counting, and mathematical reasoning), "or that he has a disorder in one of the psychological processes"; together with evidence "that regular educational services are unsuitable or of little effectiveness" (Article 7). The Organizational Guide does not mention response to intervention by name |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Cut-off point** | About 1.5 standard deviations below the mean, and as little as one standard deviation with supporting evidence (DSM-5-TR); no figure in ICD-11 | No federal figure | The text sets no numerical size for the discrepancy, and we found no published regulatory text that does |
| **Exclusion** | Intellectual and sensory disabilities, other neurological and mental disorders, psychosocial adversity, lack of proficiency in the language of instruction, and inadequate instruction | Visual, hearing, and motor disabilities; intellectual disability; emotional disturbance; and environmental, cultural, or economic disadvantage | The difficulties must not result from mental impairment, a behavioral disorder, or sensory causes (Article 7) |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Who decides** | A qualified clinician, according to each country's licensing rules | A team of qualified professionals together with the parents | The assessment and diagnosis team: the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the student counselor, specialists according to the category, and the parent where possible, after written consent (Organizational Guide) |
| **Instruments** | Standardized achievement tests and a comprehensive clinical assessment | Multi-instrument evaluation; no single procedure may be relied upon | "Instruments and measures approved by the Ministry of Education," "without relying on the results of a single measure" (Organizational Guide). We found no published list of these instruments |

The table shows that the Saudi framework is closer in its wording to the logic of an ability–achievement discrepancy than to the current US guidance, and that it allows a second route: establishing a disorder in a psychological process. Yet it also requires evidence that regular services have been of little effectiveness, which in substance is close to Criterion A

and to the logic of response to intervention. Because the text specifies neither the size of the discrepancy nor how limited effectiveness is to be established, the quality of the decision depends in practice on the evidence the team gathers. Two recent qualitative studies described RTI-like practices in Saudi schools as largely informal and lacking systematic progress monitoring (Almulla & Mahama, 2026), and female learning disabilities teachers cited among the obstacles limited time, a lack of data and training, and the absence of formal adoption of the model (Altheiabi et al., 2025).

As for prevalence, a survey of 5,495 male and female students in grades 3 through 6 in Al-Zulfi Governorate estimated the prevalence of dyslexia at about 5.86%, at 6.54% among boys and 3.83% among girls (Aldakhil et al., 2023). This estimate comes from a single governorate, and at the time of writing we found no published, peer-reviewed national estimate of the prevalence of learning disabilities in general in the Kingdom.

**A question for reflection:** If Yazan's case were brought before a team applying the text of Article 7 literally, which of its two conditions would the team rely on: the ability–achievement discrepancy, or evidence that regular services were of little effectiveness? And would it reach the same decision in Lama's case?

## Available Arabic instruments

Table 8.14 brings together the Arabic instruments and instrument studies relevant to this chapter that we found in the published literature at the time of writing. It makes no claim to completeness, and it is no substitute for verifying the licensing and norms of any instrument before using it in a decision.

**Table 8.14: Published Arabic instruments and studies in the domains of learning disabilities.**

| Area | Study | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| **Oral reading fluency** | Abu-Hamour (2014a) | Jordan; 200 students in grades 1 to 3 (ages 6 to 8) | A pilot study standardizing a curriculum-based measure of reading fluency | Small sample from a single country with a different curriculum |
| **Word recognition fluency** | Abu-Hamour (2014b) | Jordan; first-grade students | Progress monitoring in early reading | A monitoring tool, not a diagnostic test |
| **Word reading fluency** | Mahfouz & Mohamed (2023) | UAE (Al Ain); 70 second-grade students | Monitoring reading progress with curriculum-based measures | Small sample; not norms |
| **Phonological processing** | Taibah & Haynes (2011) | Arabic-speaking children | Research tasks showing the contribution of phonological processing to Arabic reading | A research study, not a standardized diagnostic instrument |
| **Dyslexia assessment in general** | Elbeheri et al. (2006) | A review of assessment issues in Arabic | A framework for what Arabic assessment instruments should take into account | Not an instrument |

| Area | Study | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| **Mathematics and number sense** | At the time of writing, we know of no published standardized Arabic instrument | | | |
| **Written expression** | At the time of writing, we know of no published standardized Arabic instrument | | | |
| **Comprehensive achievement batteries** | At the time of writing, we know of no published standardized Arabic version of the batteries presented in 8.7 | | | |

Nor did we find published Saudi norms for curriculum-based measurement in reading or arithmetic. This is a clear research gap: decisions on Criteria A and B require norms against which to compare the student. Until such norms are published, the practical alternative is for each school to build local norms by measuring all students in a grade three times a year (see Chapter 4, 4.4).

## Conclusion

The argument we have developed in this chapter centers on the idea that a learning disability is recognized first by its persistence despite intervention, and that a gap between intelligence and achievement is not, on its own, enough to establish it. The diagnostic manual requires clearly low achievement and a persistent difficulty not explained by low ability, sensory impairment, or lack of instruction. It follows that response-to-intervention data should be gathered to guard against diagnosing students who simply need instruction; that the pattern of difficulty should be specified, because each pattern has its own intervention; and that errors in Arabic reading involving short vowels, long vowels, and syllables should be read as indicators

of phonological weakness. It also follows that poor handwriting alone should not be taken as sufficient evidence, that giftedness can mask difficulty, and that anxiety may be one of its consequences.

Even so, the sources have not settled on a single position regarding discrepancy: the diagnostic manual does not require it, the international classification keeps the level of general intellectual functioning in the comparison, and the Saudi regulations call for a "clear discrepancy" without a specific figure. Also unresolved are the scarcity of standardized Arabic instruments and the absence of published Saudi norms for curriculum-based measurement, which obliges every report to state the sources of its evidence and their limits.

## 8.11 Summary, glossary, and activities

### Chapter summary

- "Learning disabilities" is an educational term, and "specific learning disorder" is the term of the diagnostic manual; its counterpart in ICD-11 is "developmental learning disorder."
- The category consists of persistent difficulty in a foundational academic skill despite intervention, not explained by low ability, sensory impairment, or lack of instruction.
- DSM-5-TR requires only clearly low achievement with difficulty persisting despite intervention, without requiring a discrepancy between intelligence and achievement. ICD-11 keeps the level of general intellectual functioning in the comparison, and the Saudi regulations call for a "clear discrepancy" without a specific figure.
- The simple view of reading distinguishes four patterns, each with a different intervention. Dyslexia usually stems from a phonological

weakness, which in Arabic shows up as errors with short vowels, long vowels, and syllables.

- Poor handwriting alone is not enough to diagnose a difficulty in written expression; developmental coordination disorder should also be considered. Dyscalculia combines weaknesses in number sense, number facts, and arithmetic fluency.
- Response-to-intervention data guard against diagnosing students who simply need instruction; giftedness can mask difficulty; and anxiety may be a consequence of the difficulty.
- Standardized Arabic instruments in this area are few, and there are no published Saudi norms for curriculum-based measurement, so the report must state the sources of its evidence and their limits.

## Glossary of terms

| | |
|---|---|
| **Specific learning disorder** | The diagnostic manual's term for persistent difficulty in learning foundational academic skills that is not explained by other factors. |
| **Developmental learning disorder** | The counterpart of specific learning disorder in the International Classification of Diseases, Eleventh Revision. |

| | |
|---|---|
| **Learning disabilities** | The educational term for the category of students who qualify for special education services because of a specific learning difficulty, as defined by the education system. |
| **Dyslexia** | A pattern of learning disability marked by poor accuracy and fluency in word recognition and by |

| | |
|---|---|
| | poor spelling, usually stemming from a phonological weakness. |
| **Dyscalculia** | A pattern of learning disability marked by weaknesses in number concepts, number facts, and accurate and fluent calculation. |
| **Number sense** | An intuitive understanding of quantities and their relationships, such as comparing and estimating them and placing numbers on a number line. |
| **Arithmetic fluency** | Performing basic number facts and operations both accurately and quickly. |
| **Developmental coordination disorder** | A neurodevelopmental disorder affecting the acquisition and execution of coordinated motor skills, which may appear as poor handwriting and slow writing. |
| **Simple view of reading** | A model that treats reading comprehension as the product of decoding and language comprehension; in this chapter we refer to it as the simple view of reading. |
| **Decoding** | Converting a written word into a spoken one by linking letters to their sounds. |
| **Phonological processing** | The brain's handling of the sounds of language, including phonological awareness, phonological memory, and rapid naming. |
| **Phonological awareness** | The ability to perceive the sounds in words and to manipulate them through deletion, blending, and |

| | |
|---|---|
| | segmentation. |
| **Rapid automatized naming** | The speed of naming a series of familiar symbols; it is linked to reading fluency. |
| **Diglossia** | The coexistence of two varieties of a language in one community, such as Standard Arabic and the colloquial dialect, each with its own uses. |
| **Pseudowords** | Made-up words that follow the phonological rules of the language, used to measure pure decoding. |
| **Dual discrepancy** | A student's performance falling below that of peers in both level of performance and rate of growth. |
| **General Ability Index** | A composite score on the Wechsler scales that measures reasoning and comprehension without working memory and processing speed. |
| **Twice-exceptional** | Students who combine giftedness with a disability or learning difficulty, so that one may mask the other. |

Response to intervention, curriculum-based measurement, the discrepancy model, and patterns of strengths and weaknesses are terms defined in Chapters 4 and 6.

## Review questions

1. What is the difference between "learning disabilities" and "specific learning disorder"? How might a student qualify under one but not the other? Then name one difference between DSM-5-TR and ICD-11, and one difference between DSM-5-TR and the Saudi regulations. **(Objective 1)**
2. Why does Criterion A require the difficulty to persist for six months despite intervention? How does this relate to Lama's case? **(Objective 2)**

3. A fourth-grade girl has a standard score of 82 in spelling and has received no systematic intervention. Is Criterion B met? What is missing before a judgment can be made? **(Objective 2)**
4. A student reads fluently but does not understand what he reads or what he hears. Which pattern of the simple view of reading does he fit? What would you measure to confirm this? **(Objective 3)**
5. A first-grade teacher asks you which signs warrant monitoring a student in mathematics, and which in writing. Name three for each. **(Objective 4)**
6. Why is reading pseudowords a more precise measure of decoding than reading real words? **(Objective 5)**
7. Which of the instruments in Section 8.7 would you choose for a seven-year-old girl suspected of weak number sense? And for a twelve-year-old boy who avoids writing? What would you write in the report if no standardized Arabic version of the instrument were available? **(Objective 5)**
8. How might a system that compares a student only with peers miss a case like Yazan's? Why did the team use the General Ability Index rather than the Full Scale IQ? **(Objectives 2 and 6)**
9. A student makes inconsistent errors in arithmetic, sometimes right and sometimes wrong on the same problems. Which two diagnoses would you consider? What assessment procedure would distinguish between them? **(Objective 7)**
10. A student's handwriting is very poor, but his spelling and syntax are intact when he dictates his text orally. Which hypothesis would you test first? Whom would you involve in the assessment? **(Objective 7)**

## Case for application

### Muadh, 13 years, seventh grade

**The story:** Muadh's teachers complain that his handwriting is "illegible," that his written answers are very short, and that he does not finish copying what is on the board. His oral participation in class, however, is good. With his parent's consent, the learning disabilities teacher gathered the following data over a month.

**Table 8.15: Data for Muadh's case.**

| Source | Result |
|---|---|
| **Copying a paragraph from the textbook in five minutes** | About half of what the school's median classmate copies; letters vary in size and drift off the line |
| **Standardized spelling test** | Standard score 96 |
| **A composition written by hand, then a similar composition typed on a keyboard** | The first: six poorly organized lines; the second: roughly twice as long, well organized, with few spelling errors |
| **Reading fluency** | Average |
| **Interview with his mother** | Late to ride a bicycle and tie his shoes; avoids playing soccer with peers |
| **Physical education teacher** | "He trips a lot and can't catch the ball" |
| **Vision and attendance** | Wears glasses, with an eye examination this year; attendance is regular |
| **Prior support** | Has received no structured writing intervention |

**Tasks:** (1) Are the criteria for specific learning disorder with impairment in written expression met? Justify your judgment against each criterion. (2) What is the most likely differential diagnosis? What evidence would you need to confirm it, and who would provide it? (3) Propose two accommodations that need not wait for the decision. **(Objectives 2, 6, and 7)**

## Learning activities

Every activity carried out with students is conducted with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a specialist; its results are not used in any decision concerning the student.

1. **A pseudoword list:** Design twenty fully vowelized Arabic pseudowords, graded in length and difficulty, with some that distinguish short vowels from long vowels. Then try them with three students, recording and classifying their errors. *Desk alternative:* Try them with three adult colleagues and analyze which items slowed their reading, and why. **(Objective 5)**
2. **Applying the simple view of reading:** Choose five students and, using two equivalent texts, measure each student's listening comprehension and reading comprehension; then place them within the patterns of Figure 8.2. *Desk alternative:* Write five brief hypothetical profiles, classify them, ask a colleague to classify them independently, and compare your results. **(Objective 3)**
3. **A student profile:** Turn Ahmed's profile in Figure 8.5 or Dima's profile in Figure 8.6 into an intervention plan with three goals, specifying how progress toward each goal will be measured. **(Objective 6)**
4. **Comparing systems:** Apply the Saudi regulations column of Table 8.13 to the cases of Ahmed, Lama, Dima, and Yazan. For each case, write whether

the regulations would have reached the same decision, what evidence that decision rests on, and what the text leaves unresolved. **(Objective 1)**

## Chapter references

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Abu-Hamour, B. (2014a). A pilot study for standardizing curriculum-based measurement oral reading fluency (CBM ORF) in Arabic. *Journal of the International Association of Special Education, 15*(1), 16–26.

Abu-Hamour, B. (2014b). Using Arabic word identification fluency to monitor first-grade reading progress. *Dyslexia, 20*(2), 167–174. https://doi.org/10.1002/dys.1472

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*, 39–59.

Aldakhil, A. F., Ebrahim, M. T., & Gadelrab, H. F. (2023). Diagnostic survey of dyslexia and accompanying behavioral indicators in primary school students in Saudi Arabia. *Research in Developmental Disabilities, 134*, Article 104424. https://doi.org/10.1016/j.ridd.2023.104424

Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications, 13*(1), Article 974. https://doi.org/10.1057/s41599-026-07376-9

Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties, 30*(1), 75–92. https://doi.org/10.1080/19404158.2025.2473425

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Breaux, K. C. (2020). *Wechsler Individual Achievement Test* (4th ed.): *Technical and interpretive manual*. NCS Pearson.

Connolly, A. J. (2007). *KeyMath-3 Diagnostic Assessment*. Pearson.

DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications for DSM-5 and assessment and treatment. *Journal of Learning Disabilities, 46*(1), 43–51.

Elbeheri, G., Everatt, J., Reid, G., & Al Mannai, H. (2006). Dyslexia assessment in Arabic. *Journal of Research in Special Educational Needs, 6*(3), 143–152.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). Guilford.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability* (3rd ed.). PRO-ED.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Hammill, D. D., & Larsen, S. C. (2009). *Test of Written Language* (4th ed.). PRO-ED.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), 20 U.S.C. § 1400 et seq. (2004).

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850–867.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–14.

Mahfouz, M. A. S., & Mohamed, A. H. H. (2023). Using word reading fluency curriculum-based measurements to monitor students' reading progress in grade 2. *Education Sciences, 13*(2), Article 217. https://doi.org/10.3390/educsci13020217

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108–121.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics, 24*(3), 431–451.

Taibah, N. J., & Haynes, C. W. (2011). Contributions of phonological processing skills to reading skills in Arabic speaking children. *Reading and Writing, 24*(9), 1019–1042.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency* (2nd ed.). PRO-ED.

U.S. Department of Education. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final rule (34 CFR Parts 300 and 301). *Federal Register, 71*(156).

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing* (2nd ed.). PRO-ED.

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

# 9 Chapter 9
# Autism spectrum disorder

How do we know that a child has autism spectrum disorder? Who holds the authority to decide, with which instruments, and at what point does the evidence become strong enough to justify a judgment? In this chapter we trace the entire path, from the teacher's first observation in the classroom to the diagnostic decision.

---

**Scene from a kindergarten**

On an ordinary school morning, the kindergarten teacher notices that Yousef does not turn when she calls his name, yet turns at once when he hears a packet of biscuits being opened. He spends free play lining up small cars in a straight row, and if another child moves a single car out of place, he bursts into tears. During morning circle he repeats a line from a television commercial in exactly the tone in which he heard it.

The teacher wonders: Is this autism? Or is he a shy child who needs time to settle in? Or is something wrong with his hearing?

In this chapter we set out to answer the teacher's question systematically, returning to Yousef at each stage of the decision pathway (9.5) to show how a passing observation becomes a referral, then an evaluation, and finally a decision.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Match described classroom behaviors to the five criteria for autism spectrum disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR), and classify the same case using the approach of the International Classification of Diseases (ICD-11).
2. Write an objective behavioral observation suitable for referral, and identify, in a given description, the red flags that call for evaluation without delay.
3. Sequence the steps of the decision pathway, and explain why hearing and vision testing comes before any diagnostic instrument.
4. Classify autism instruments as screening, rating, or diagnostic tools, and select for a described student the appropriate instrument and its form or module.
5. Interpret instrument results as they appear in a diagnostic report, and evaluate the meaning of published figures on instrument accuracy, prevalence, and the status of Arabic versions.
6. Judge whether the evidence is sufficient to make a diagnosis in a new case, distinguish autism from similar and co-occurring conditions, and justify that judgment.

# Introduction

The diagnosis of autism spectrum disorder follows a long path. It usually begins with an observation from a teacher or a family, as in Yousef's kindergarten scene, and ends with a judgment by a trained specialist after evidence has been gathered from multiple sources. Each stop along the way has its own function and its own limits: red flags call for evaluation without delay; screening tools sort but do not diagnose; rating scales convey what is seen by those who know the child; and diagnostic instruments are not sufficient on their own. Precision matters all the more because the disorder may be misdiagnosed when hearing loss or a language disorder is the better explanation, and it may be overlooked when giftedness masks it or when a girl conceals her traits.

We open the chapter by defining the disorder and reading its prevalence figures (Section 9.1), then present the diagnostic criteria in DSM-5-TR (Section 9.2), support levels, specifiers, and the comparison with ICD-11 (Section 9.3), and early red flags and delayed identification in girls (Section 9.4). We then turn to the five-stage decision pathway (Section 9.5), the instrument profiles (Section 9.6), and reading the diagnostic report (Section 9.7), before applying all of this to three cases (Section 9.8). We close with differential diagnosis (Section 9.9), co-occurring conditions and medical evaluation (Section 9.10), and the Saudi and Arab context (Section 9.11).

## 9.1 What is autism spectrum disorder?

Autism spectrum disorder is a neurodevelopmental disorder: it is tied to the way the brain develops and functions, and its first signs appear in the early years of life. It is characterized by two core features: persistent difficulty in social communication and interaction, and restricted, repetitive patterns of behavior and interests (Lord et al., 2020).

The word **spectrum** signals that the disorder does not take a single form. We may meet a child who never speaks and needs help with most daily tasks, and an adolescent who talks fluently and excels in mathematics yet cannot tell when a classmate is joking and when he is serious. Both are on the same spectrum, but their needs differ greatly. For this reason, a report needs more than the statement "has autism": it must also describe the level of support the student needs and any other conditions that accompany the autism.

In 2013, the fifth edition of the Diagnostic and Statistical Manual brought together under this single label several diagnoses that had been separate in the fourth edition, among them autistic disorder, Asperger's disorder, and pervasive developmental disorder not otherwise specified. Teachers may therefore come across a diagnosis of "Asperger's" in older student files. Today it falls within autism spectrum disorder and is usually described as autism without accompanying intellectual impairment and without accompanying language impairment.

## How common is autism, and how should we read prevalence figures?

In the most recent report of the US Autism and Developmental Disabilities Monitoring Network, prevalence among 8-year-olds in 2020 was about 27.6 per 1,000, or roughly one child in 36, and it was about 3.8 times higher in boys (Maenner et al., 2023). In the Arab region, the figures vary widely:

- In a national survey in Qatar of children aged 6 to 11, prevalence was 1.14%, with a confidence interval of 0.89% to 1.46% (Alshaban et al., 2019).
- In Riyadh, the rate was 2.51% (about one child in 40) among 398 children aged 2 to 4 at five hospitals, with three boys for every girl (AlBatti et al., 2022).
- In Makkah and Jeddah, the number of children registered at 37 autism centers came to about 2.81 per 1,000 children (Sabbagh et al., 2021).

- A recent study reports that the official estimate in the Kingdom is about 0.6%, and estimates actual prevalence at between 1.7% and 1.8% (Khamees et al., 2025).

These differences do not mean that autism is ten times more common in one city than in another. Each figure answers a different question. A national survey that actively searches for children differs from a sample of hospital patients, which in turn differs from center registries that count only those who have reached services. Age, the screening instrument, and the case definition also differ. The first question to ask of any prevalence figure, then, is how these children were found. A low registry figure measures not so much prevalence as the gap between those who have the disorder and those who have reached a diagnosis.

**Common misconceptions**

- **"It is caused by a cold mother or a neglectful family."** This is an old hypothesis from the mid-twentieth century that has been shown to be false. Autism is a neurodevelopmental disorder; families do not cause it through the way they raise their children.
- **"Vaccines cause autism."** Large-scale studies have found no association between vaccines and autism, including a meta-analysis covering more than a million children (Taylor et al., 2014).
- **"Every child with autism is a genius in some area."** Exceptional abilities are present in only some. The spectrum spans every level of intellectual ability.
- **"If a child looks me in the eye, he doesn't have autism."** Eye contact is one form of one criterion. Many autistic children do make eye contact, but in ways that do not serve communication.

## 9.2 Diagnostic criteria in DSM-5-TR

Diagnosis relies on one of two manuals: the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**), published by the American Psychiatric Association in 2022 (American Psychiatric Association, 2022), and the International Classification of Diseases, Eleventh Revision (**ICD-11**), published by the World Health Organization (World Health Organization [WHO], 2024), which is used to code diagnoses in health records in most countries. We present the DSM criteria here because they are more detailed and because most assessment instruments and reports are built on them; we then compare them with ICD-11 in 9.3. The manual requires five criteria to be met together. Two describe the symptoms themselves, and three are conditions without which the diagnosis is incomplete. Figure 9.1 summarizes this rule, and we then explain each criterion with examples.

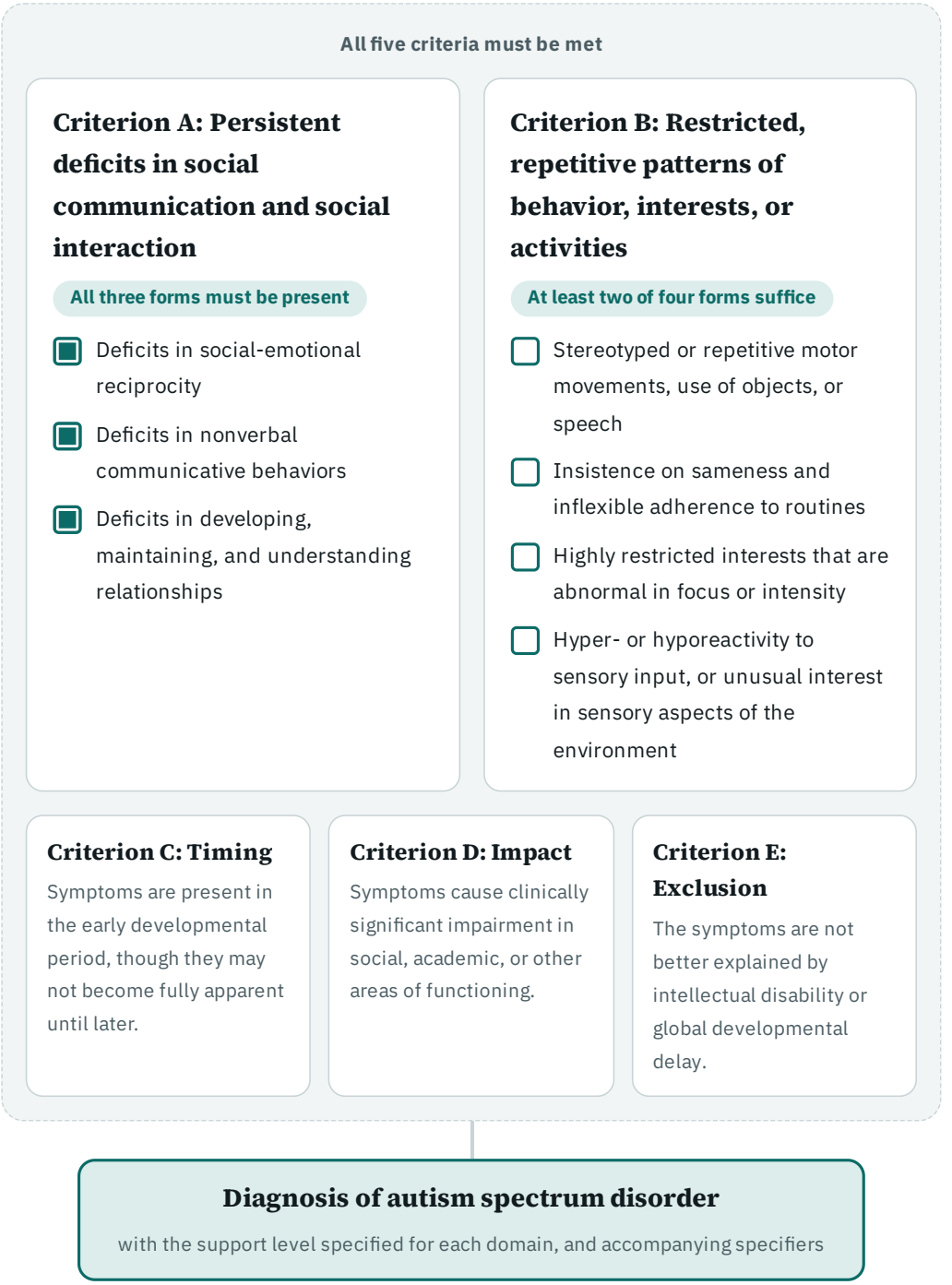



**Figure 9.1:** The five **DSM-5-TR** criteria. The filled boxes under Criterion A indicate that every form is required; the empty boxes under Criterion B indicate that two of the forms are sufficient. The absence of any one of the five criteria halts the judgment.

## Criterion A: Persistent deficits in social communication and social interaction

This criterion requires **all three of the following forms** to be present, either currently or in the child's developmental history as the family reports it. The word “persistent” matters: what is meant is a stable pattern that recurs across different situations and on many days.

### Form 1: Deficits in social-emotional reciprocity

That is, difficulty with the natural give-and-take between two people. Social interaction resembles a game of catch: one person throws, the other catches and throws back. An autistic child may catch the ball and not return it, or throw it without checking whether the other person is ready. Examples at school include the following:

- The child does not initiate conversation with classmates; when he does, it is usually to ask for something he wants, and rarely to share what interests him.
- He does not share his interest with others: he does not show the teacher his drawing to enjoy her praise, or point to an airplane in the sky to draw the attention of those around him. Specialists call this **joint attention** (**Joint attention**).
- In middle school, for example, he talks at length about a favorite topic without noticing that his classmate has lost interest.

### Form 2: Deficits in nonverbal communicative behaviors

That is, in the body language we use without words to convey what we mean:

- Unusual eye contact: very little, an uncomfortable stare, or gaze that is not coordinated with speech.
- Limited use of gestures, such as pointing with a finger, nodding to agree, or waving goodbye.

- Facial expressions that do not fit the situation, or difficulty understanding other people's facial expressions and tone of voice.

#### Form 3: Deficits in developing, maintaining, and understanding relationships

- Difficulty with imaginative play with peers, such as playing “house” or “restaurant.”
- Difficulty adjusting behavior to the situation: behaving in class as on the playground, or addressing the principal as he would a classmate.
- Little interest in peers, or a wish for friendship without knowing how friendships are built and kept.

> **Why all three forms?**
>
> Because each form on its own may appear for other reasons. A shy child may avoid eye contact, a child with delayed language may not initiate conversation, and a child who has just moved to a new school may have no friends yet. It is the combination of all three, persisting across different situations, that points to a deeper pattern.

## Criterion B: Restricted, repetitive patterns of behavior, interests, or activities

This criterion requires **at least two** of the following four forms, currently or in the developmental history:

**Table 9.1: Forms of Criterion B and examples at school.**

| Form | What it means | Examples at school |
|---|---|---|
| **1. Stereotyped movements, speech, or use of objects** | Repeating the same movement or utterance without a clear communicative purpose | Hand flapping when excited. Repeating lines from cartoons in the same intonation, known as **echolalia** (**Echolalia**). Lining up pencils in a row instead of writing with them. |
| **2. Insistence on sameness and adherence to routine** | Intense distress at small changes, and adherence to specific rituals | A crying fit because the teacher changed his seat, or because the line took a different route to the cafeteria. |
| **3. Restricted, intense interests** | Interest that is unusual in its intensity or its focus | Knows the names of every metro station and refuses to talk about anything else, or carries a small fan with him wherever he goes. |
| **4. Unusual sensory response** | Over- or under-reaction to sounds, textures, light, or pain | Covers his ears at the school bell, refuses to wear the uniform because of how it feels, does not seem hurt when he falls, or stares for a long time at a spinning fan. |

Back to Yousef in the opening scene: lining up the cars in a straight row falls under the first form, bursting into tears when one is moved falls under the second, and repeating the commercial's line in its original intonation again belongs to the first. In other words, what the teacher saw in a single morning falls under two forms of Criterion B. Yet this alone is not enough: we still need Criterion A in full, along with the three remaining conditions.

## Criterion C: Timing

Symptoms must be present in the **early developmental period**. The manual acknowledges, however, that they may not become fully apparent until social demands exceed the child's capacities. Some children, particularly those with good language and intellectual abilities, do not reveal their difficulties until middle school, when peer relationships grow more complex and their rules less explicit. This is why the specialist always asks the family about the early years, even when the student arrives for evaluation at age 12.

## Criterion D: Functional impact

Symptoms must cause **clinically significant impairment** in social, academic, occupational, or other areas of functioning. Having some traits that do not interfere with the child's life is not enough for a diagnosis. Many people have deep interests or prefer routine and lead successful lives, and this is not considered a disorder.

## Criterion E: Exclusion

The symptoms must not be better explained by intellectual disability or global developmental delay. To understand this criterion, remember that in a child with intellectual disability all skills develop slowly, social skills included. If the child's social skills are consistent with his overall level of intellectual development, intellectual disability alone explains them. If, however, his social skills are much weaker than his developmental level would predict, he may have both autism and intellectual disability, and the diagnosis records both.

**An example illustrating Criterion E**

Two 7-year-olds, each with intellectual abilities equivalent to those of a 4-year-old. The first plays with younger children, laughs along with them, and brings his toy to his mother to show it to her; his interaction, in other words, resembles that of a 4-year-old. This is explained by intellectual disability. The second does not initiate interaction even with younger children and shares his interests with no one, which is less than 4-year-olds typically do. Here, the presence of autism alongside intellectual disability is considered.

## 9.3 Support levels, specifiers, and comparison with ICD-11

### How much support does the student need?

After diagnosis, the specialist specifies the severity level for each domain separately: social communication, and restricted, repetitive behaviors. The yardstick is how much support the student needs in daily life, regardless of the number of symptoms. The manual calls these "severity levels," but defines each level by the amount of support required, which is why they are commonly referred to as "support levels." We use that term in this chapter to distinguish them from the severity levels of intellectual disability, which are determined by adaptive behavior.

**Table 9.2: Support levels in autism spectrum disorder (adapted from American Psychiatric Association, 2022).**

| Level | Description | What it looks like in practice |
|---|---|---|
| **Level 1: Requiring support** | Noticeable difficulties in the absence of support | A student who speaks in full sentences but stumbles when starting a conversation, has difficulty making friends, and becomes flustered when moving between activities. |
| **Level 2: Requiring substantial support** | Marked difficulties even with support in place | A student who speaks in simple sentences, whose interaction is limited to his own special interests, and whose repetitive behaviors occur often enough for anyone to notice. |
| **Level 3: Requiring very substantial support** | Severe deficits affecting most areas of life | A student who speaks very little or not at all, rarely initiates interaction, and for whom a change in routine causes intense distress that is hard to soothe. |

A student may be at Level 1 in social communication and at Level 2 in restricted behaviors, so the report states both levels. The level may also change with development and intervention: it describes the current situation and is open to revision.

## Accompanying specifiers

The specialist adds to the diagnosis whatever accompanies it, because this changes the student's educational picture considerably. The report states, for example, whether the autism is:

- With or without accompanying intellectual impairment.
- With or without accompanying language impairment.
- Associated with a known medical or genetic condition or environmental factor.

- Associated with another neurodevelopmental, mental, or behavioral disorder, such as attention-deficit/hyperactivity disorder or anxiety.
- With catatonia, meaning motor freezing, holding fixed postures, extreme slowing, or purposeless movements. It is uncommon, but may emerge in adolescence as a sudden decline in movement, speech, and self-care. It calls for urgent medical referral and should not be treated as ordinary "autistic behavior."

Two students diagnosed with autism, one with accompanying intellectual and language impairment and the other with neither, need entirely different educational plans, even though the name of the diagnosis is the same. Section 9.10 returns to the common co-occurring conditions and how they are assessed.

**Comparison with ICD-11**

ICD-11 agrees with DSM-5-TR on the two symptom domains (social communication, and restricted and repetitive behavior, including sensory responses) and on onset in the developmental period, but differs from it in ways that affect the report (WHO, 2024):

- **Two specifiers instead of support levels:** the disorder is classified using two specifiers: the presence or absence of a disorder of intellectual development, and the degree of impairment in **functional language**, that is, the child's ability to use language for his needs and for communication: intact or mildly impaired, markedly impaired, or virtually absent. Sara in 9.8 would thus be classified as "without disorder of intellectual development, with impairment of functional language."
- **Loss of skills:** the classification records whether the child has lost language or social skills he had previously acquired.
- **Descriptions, not checklists:** ICD-11 sets out essential requirements and clinical descriptions, and does not require a fixed number of forms in each domain as DSM-5-TR does.

**For the teacher**

A medical diagnosis under the DSM is one thing; eligibility for special education services is another. A student may hold a medical diagnosis while the decision about educational services still rests on the disorder's impact on his learning and school performance. Your role is central in both, because you see the student in an environment the physician never sees: among his peers, in transitions between lessons, and in unexpected situations.

## 9.4 Early red flags

The earlier autism is identified, the earlier intervention begins and the greater the chances of improvement. Since the late 1990s, medical guidelines have specified signs that, when present, call for evaluation without delay (Filipek et al., 1999), and the current guidelines of the American Academy of Pediatrics have reaffirmed them. These guidelines also recommend screening all children with an autism-specific instrument at the 18- and 24-month visits, without waiting for signs to appear, alongside developmental surveillance at every visit (Hyman et al., 2020). The presence of these signs means that waiting is no longer an appropriate option; the diagnosis itself, however, still depends on a comprehensive evaluation.

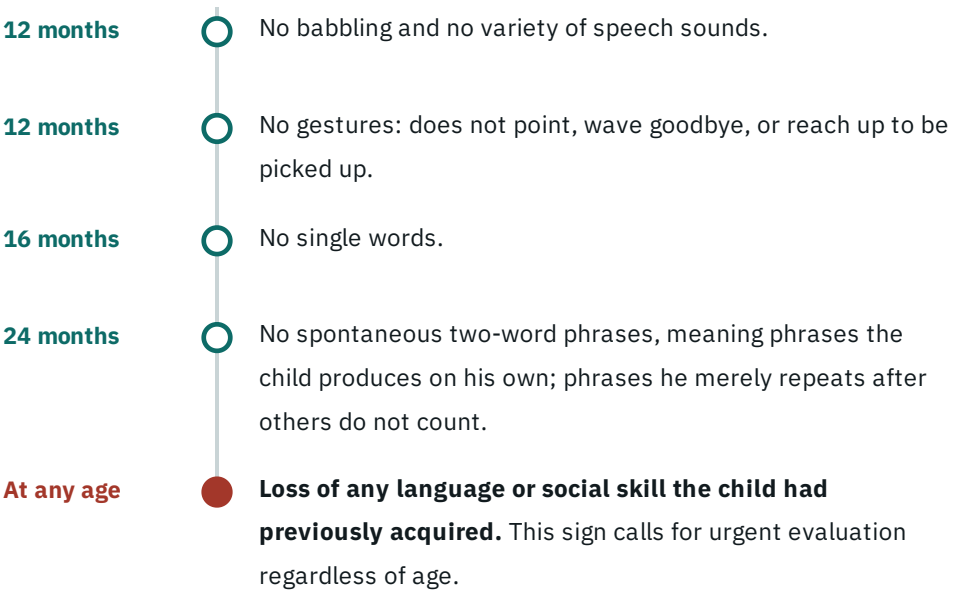



**Figure 9.2:** Early red flags that call for evaluation without delay, from Filipek et al. (1999); they remain in current guidelines (Hyman et al., 2020).

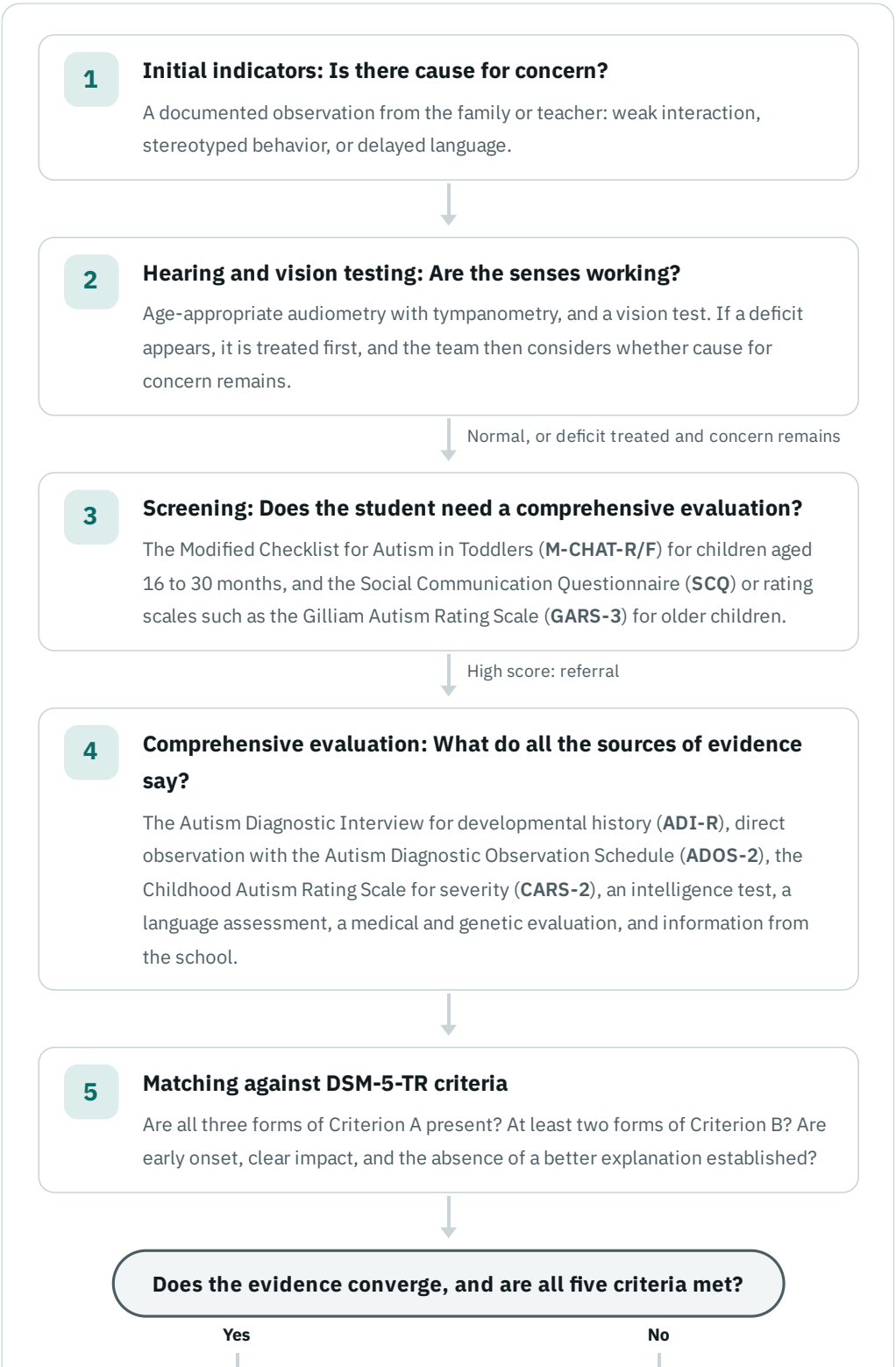
### Why is autism identified later in girls?

Girls are diagnosed with autism far less often than boys. A large systematic review estimates the true ratio at about three boys to every girl, lower than the ratios found in diagnostic records (Loomes et al., 2017). Researchers believe that part of the gap arises because many girls learn to **camouflage** their traits (**Camouflaging**): they watch their female classmates, imitate the way they talk and move, and memorize ready-made phrases for social situations (Hull et al., 2017). Their intense interests may also center on topics that attract little attention, such as animals or novels, and so do not seem "odd" in the way a boy's interest in train timetables might.

As a result, a girl may appear well adjusted at school and then collapse from exhaustion at home. So when a mother says that her daughter "gets very tired after school and needs to be alone for hours," this is information worth heeding.

## 9.5 The decision pathway: From observation to judgment

The student passes through five stages before a decision is made, and each stage has its own question and its own instruments. The central idea of this pathway is that its stages are completed in order, so that the judgment is made only after each stage has answered its question. Hearing and vision testing comes at the second stage as a prerequisite for any instrument, because any autism measure administered to a child who does not hear well yields a doubtful result, whatever that result may be.


1
Initial indicators: Is there cause for concern?
A documented observation from the family or teacher: weak interaction, stereotyped behavior, or delayed language.
2
Hearing and vision testing: Are the senses working?
Age-appropriate audiometry with tympanometry, and a vision test. If a deficit appears, it is treated first, and the team then considers whether cause for concern remains.
Normal, or deficit treated and concern remains
3
Screening: Does the student need a comprehensive evaluation?
The Modified Checklist for Autism in Toddlers (M-CHAT-R/F) for children aged 16 to 30 months, and the Social Communication Questionnaire (SCQ) or rating scales such as the Gilliam Autism Rating Scale (GARS-3) for older children.
High score: referral
4
Comprehensive evaluation: What do all the sources of evidence say?
The Autism Diagnostic Interview for developmental history (ADI-R), direct observation with the Autism Diagnostic Observation Schedule (ADOS-2), the Childhood Autism Rating Scale for severity (CARS-2), an intelligence test, a language assessment, a medical and genetic evaluation, and information from the school.
5
Matching against DSM-5-TR criteria
Are all three forms of Criterion A present? At least two forms of Criterion B? Are early onset, clear impact, and the absence of a better explanation established?
Does the evidence converge, and are all five criteria met?
Yes
No

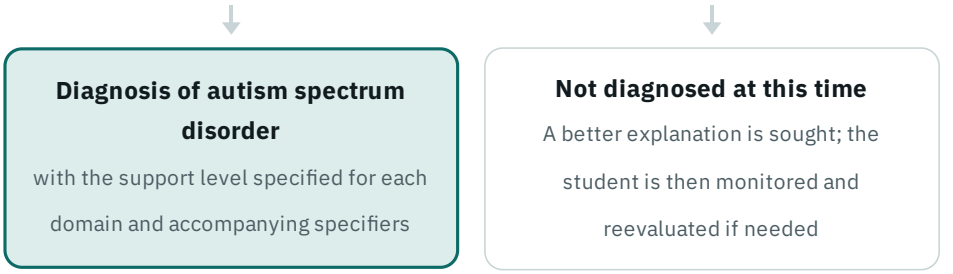



**Figure 9.3:** The decision pathway in diagnosing autism spectrum disorder. Sensory testing precedes every instrument, a high screening score means referral only, and the judgment is made only after matching against the criteria.

## Stage 1: Initial indicators

The story usually begins with an observation from the family or a teacher. Here lies the teacher's most important role, because the teacher sees the student every day among peers, in situations the physician never sees in the clinic. An observation helps the team, however, only if it is **documented and objective**. Compare these two observations about Yousef:

**An observation that does not help the team**

"Yousef is odd and doesn't like his classmates, and I think he has autism."

**An observation the team can build on**

"Over two weeks, Yousef did not once initiate conversation with a classmate. I called his name from two meters away ten times, and he did not turn on seven of them, yet he turned immediately at the sound of a cookie box being opened. He lined up the toy cars in a straight row on 8 of 10 free-play days."

The first observation is a personal judgment; it even contains a diagnosis, which lies outside the teacher's remit. The second describes a specific behavior, in a specific context, with the number of times it occurred. It also draws attention to something important: Yousef hears the biscuit packet, so his failure to respond to his name is probably not due to hearing loss, although a hearing test remains necessary. A record such as the following helps the teacher document what she sees:

**Table 9.3: A sample observation record for teachers to use before referral, with two entries from Yousef's record. It may be copied for classroom use.**

| Date and time | Situation | What happened before the behavior | Behavior as I saw it | What happened after |
|---|---|---|---|---|
| **Sunday 8:40** | Morning circle | The teacher began the morning song | Covered his ears with his hands, lowered his head, and did not join in | Calmed down when the song ended |
| **Sunday 10:15** | Free play | A classmate took a car from his row | Screamed and cried for three minutes, then put the car back in its place | Went back to lining up the cars alone |

The record captures what came before and after the behavior, because context is part of the evidence.

## Stage 2: Hearing and vision testing

Before any autism instrument is used, the question is: Does the child hear and see well? Not turning to his name, delayed speech, and limited interaction in noisy settings can all arise from hearing loss alone. A child who is administered an observation schedule while unable to hear the specialist clearly may appear less interactive than he really is, and the result loses its

validity. Passing a newborn hearing screening is not enough, since some types of hearing loss appear later or fluctuate with middle ear infections. The team therefore requests a recent, age-appropriate hearing assessment with tympanometry, together with a vision test.

The kindergarten referred Yousef for a hearing test first. His hearing proved normal in both ears, his tympanometry was normal, and his vision test was clear. Only then did screening become meaningful.

## Stage 3: Screening

Screening is a sorting tool; diagnosis is a later stage that requires broader evidence. The function of screening is to determine who needs a more in-depth evaluation, so that not every quiet child is referred and no child who genuinely needs evaluation is overlooked.

> **An analogy that clarifies the difference**
>
> Screening is like the metal detector at an airport. If it beeps, the traveler needs a hand search to settle the matter. A high screening score works the same way: it means "look more closely," not "he has autism."

The accuracy of a screening tool is measured by two concepts: **sensitivity** (**Sensitivity**), the proportion of those who actually have the disorder whom the tool detects, and **specificity** (**Specificity**), the proportion of those without the disorder whom the tool correctly rules out. Chapter 2 (Section 2.6, Table 2.5) presented a worked example showing that a tool with 85% sensitivity and 95% specificity, applied to 1,000 children among whom the disorder has a prevalence of 2%, yields positive results of which roughly three quarters are false. The lesson is that the value of a positive result depends as much on the proportion of those screened who have the disorder as on the tool itself.

Real-world data tell the same story. In the validation study of the **M-CHAT-R/F**, administered to more than 16,000 children at primary care visits, fewer than half of those who still screened positive after the Follow-Up interview were diagnosed with autism. Most of the rest, however, had developmental delays or other concerns warranting intervention, so their results were not "false alarms" in the full sense (Robins et al., 2014):

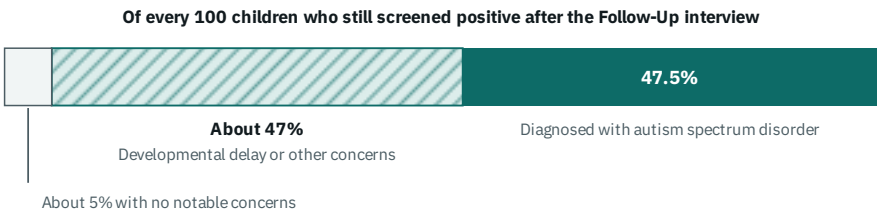



**Figure 9.4:** Outcomes of evaluation for children who still screened positive on the **M-CHAT-R/F** after the Follow-Up interview: 47.5% were diagnosed with autism, and 94.6% overall had developmental delays or concerns warranting follow-up (Robins et al., 2014). A positive result is therefore an invitation to closer examination.

Yousef is 5 and thus past the age range of the **M-CHAT-R/F**. His mother therefore completed the Lifetime form of the Social Communication Questionnaire (see its profile in 9.6). He scored 21, above the cut-off score of 15, and was referred to a diagnostic center together with his teacher's record.

## Stage 4: Comprehensive evaluation

This stage is carried out by a multidisciplinary team that may include a pediatrician or psychiatrist, a psychologist, a speech-language pathologist, and a special education specialist. The team gathers evidence from three independent sources:

1. **Developmental history** from the family, through a structured interview such as the **Autism Diagnostic Interview–Revised, ADI-R**: When did the symptoms begin? What was the child like at age four?

2. **Standardized direct observation** of the child, using an instrument such as the **Autism Diagnostic Observation Schedule, ADOS-2**: How does he interact now with an unfamiliar person in situations designed to elicit social behavior?
3. **Supplementary assessments** that rule out other explanations and reveal co-occurring conditions: an intelligence test to establish intellectual ability, a language assessment, and a medical evaluation that asks about seizures, sleep, feeding, and gastrointestinal problems. The American Academy of Pediatrics recommends offering genetic testing to every child diagnosed with autism, beginning with chromosomal microarray and testing for fragile X syndrome (Hyman et al., 2020).

Information from the school is added to these: the teacher's ratings on the rating scale, her observation records, and samples of the student's work.

In Yousef's case, the specialist interviewed his mother using the diagnostic interview, and his scores exceeded the cut-off scores in all three domains; his mother reported that he had not pointed or waved before the age of two. He was administered Module 1 of the observation schedule, because he does not use spontaneous phrases and most of his speech repeats what he has heard; the classification was “autism.” His nonverbal ability was 88, in the average range.

## Stage 5: Matching against the criteria

The team lays out all the evidence and asks about each criterion in turn: Is it met? What is the evidence for it? From which source did that evidence come? If the sources converge on all five criteria being met, the diagnosis is made, with the support level specified. If the sources conflict, or if a better explanation such as hearing loss emerges, the judgment is suspended and that explanation is investigated.

Yousef's evidence converged: the teacher's record, his mother's report, and direct observation. He was diagnosed with autism spectrum disorder without accompanying intellectual impairment, with accompanying language

impairment, at support Level 2 in social communication and Level 1 in restricted behaviors, and he was referred for genetic evaluation. In this way the first morning observation, once documented and followed by hearing testing, screening, and evaluation, became a decision grounded in evidence.

☆ **A golden rule**

A decision not to diagnose may be the most accurate outcome of an evaluation. A student misdiagnosed with autism loses the services he actually needs, and his family bears the burden of a diagnosis that does not fit him.

## 9.6 Instrument profiles

In this section we present six of the most widely used instruments in the diagnosis and identification of autism. They fall into three families with different functions:

**Screening tools**

Sort out the children who need a more in-depth evaluation. Quick, and do not require advanced qualifications.

M-CHAT-R/F · SCQ

**Rating scales**

Someone who knows the child rates how strongly the traits appear, yielding a severity score.

CARS-2 · GARS-3

**Diagnostic instruments**

Administered by a specialist with formal training; regarded as the primary reference in diagnosis.

ADI-R · ADOS-2

Each instrument has a profile following the book's standard template: its underlying idea, what it actually looks like, and a scene from its administration, followed by an information card with twelve fields (purpose, ages, edition and year, publisher, qualification level, administration time, score types, cut-off scores, reliability, classification accuracy, norming year,

and Arabic version), then how its results are interpreted, and finally its limitations. Where we do not know a figure from a source we trust, we say so and refer the reader to the instrument's manual rather than estimating it.

> ⚠ **A caution on test security and Arabic versions**
>
> This book describes instruments and explains how to administer and interpret them, but it does not publish their items or reproduce their record forms. The items are the property of the publishers, and publishing them undermines an instrument's validity when examinees see them in advance. The drawings representing forms in this chapter illustrate only the layout of the page; their gray lines do not represent real items.
>
> Many instruments also circulate in unlicensed translations without local norms. Each profile therefore states the status of the Arabic version, and readers should verify a version's license and norms before relying on it.

## M-CHAT-R/F Modified Checklist for Autism in Toddlers, Revised, with Follow-Up

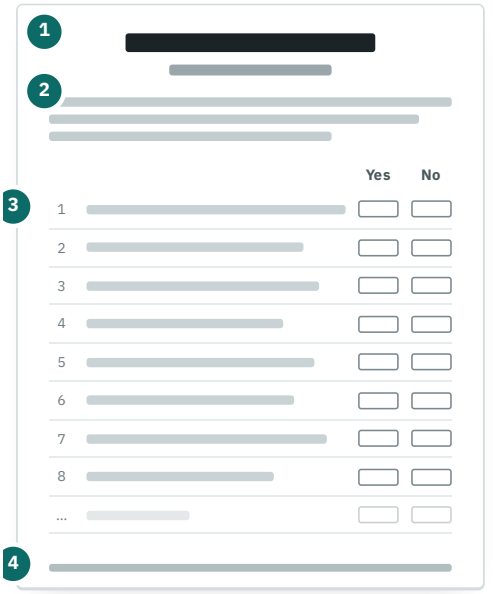
Screening tool

### The idea

This checklist was designed to be administered to **all** children at routine health care visits, including those who raise no concern at all. The aim is to identify children whom the family may have missed, at the age when intervention is most beneficial (Robins et al., 2014).

### What it looks like

A single page of 20 simple questions that parents answer yes or no. They ask about everyday behaviors every parent knows, such as pointing with a finger, looking at what others point to, and pretend play. The checklist comes with a **Follow-Up interview** guide that the professional uses for the questions whose answers indicated risk, asking the family for real-life examples before scoring the answer.




1. The instrument name, the child's details, and age in months.
2. Brief instructions for parents: answer according to your child's usual behavior.
3. Twenty questions, each with a Yes box and a No box beside it.
4. A copyright line at the foot of the page, which the authors require to be retained whenever the form is printed.

**Figure 9.5:** An original illustration of the checklist's page layout. The gray lines do not represent the actual questions; the original form can be downloaded from the official website.

### A scene from administration

At a primary care center, the nurse hands the checklist to the mother of Raghad, a 20-month-old, to complete while they wait for the doctor. She finishes it in five minutes. The doctor tallies the answers that signal risk and counts four, so he sits down with the mother and goes through each one in detail: “When you said she doesn't point, does she point if she wants juice from the shelf? Does she point to a cat in the street to show it to you?” After the follow-up, three at-risk answers remain, and the doctor refers Raghad for a diagnostic evaluation.

**Table 9.4: M-CHAT-R/F information card.**

| | |
|---|---|
| **Purpose** | Early screening for possible autism, ahead of diagnosis |
| **Ages** | 16 to 30 months |
| **Edition and year** | The revised version with Follow-Up interview; its validation study was published in 2014 |
| **Publisher** | Its authors; it can be downloaded free of charge from their official website for clinical, educational and research use, under conditions they set |
| **Qualification level** | Completed by parents; the follow-up interview is conducted by a physician, nurse or specialist trained in it; no psychological assessment license required |
| **Time** | About 5 minutes for the checklist and 5 to 10 minutes for the follow-up |
| **Score types** | The number of at-risk answers, from 0 to 20 |
| **Cutoff scores** | 3 and 8, then 2 after the follow-up (Table 9.5) |
| **Reliability** | The instrument is not interpreted against norms, and its published studies focus on classification accuracy rather than reliability coefficients |
| **Classification accuracy** | Of the children who still screened positive after the follow-up, 47.5% were diagnosed with autism, and 94.6% had developmental delays or concerns (Robins et al., 2014; Figure 9.4) |
| **Year of norm collection** | It has no age norms. The cut-off scores derive from the validation study in US primary care |

| | |
|---|---|
| **Arabic version** | The first version of the checklist was studied in nine Arab countries with 228 children, 122 of whom screened positive; sensitivity was 0.86, specificity 0.80, and positive predictive value 0.88 (Seif Eldin et al., 2008). This high value reflects a sample in which more than half screened positive, and it can be expected to drop in universal screening. The official website lists the approved translations of the revised version, but at the time of writing we know of no published Arabic standardization study for it |

### How are the results interpreted?

The at-risk answers are summed and read in three ranges:

**Table 9.5: M-CHAT-R/F score ranges and the action for each.**

| Score | Risk level | Action |
|---|---|---|
| **0 to 2** | Low | No action for now; continue monitoring at future visits. If the child is younger than 24 months, rescreen after the second birthday. |
| **3 to 7** | Average | Conduct the Follow-Up interview. If a score of 2 or more remains afterward, refer the child for evaluation. |
| **8 to 20** | High | Refer directly for diagnostic evaluation and early intervention; the Follow-Up interview may be skipped. |

### Limitations

It may miss some children, particularly those whose symptoms are mild or emerge later. A high result signals the need for a fuller evaluation, as Figure 9.4 shows. Reproducing the checklist in a book or on a website requires the authors' permission.

## SCQ Social Communication Questionnaire

Screening tool

### The idea

The **M-CHAT-R/F** stops at 30 months. What, then, of a five-year-old, or a ten-year-old student whose teacher is worried? This questionnaire was designed to carry the questions of the diagnostic interview with the family, in condensed form, onto a form a parent can complete in minutes, sorting out who needs a comprehensive evaluation (Rutter, Bailey, et al., 2003).

### What it looks like

Forty yes/no questions in two forms. The **Lifetime** form asks about the child's entire history, and some items ask specifically about ages four to five; this is the form used for screening and referral. The **Current** form asks about the past three months and is useful for monitoring. The first question establishes whether the child speaks in phrases, so the language items are dropped for a child who does not.

### A scene from administration

At an early intervention center, the specialist gives Yousef's mother the Lifetime form and asks her to answer according to what she genuinely remembers of his behavior. The mother pauses at a question about pointing at things to show them to others, then says, “He never did that.” The specialist tallies the relevant answers: the score is 21. She notes in his file that it is above the cut-off and refers him for a comprehensive evaluation, without offering the mother “autism” as a verdict.

**Table 9.6: SCQ information card.**

| | |
|---|---|
| Purpose | Screening for the social communication traits associated with autism in those past the age range of the toddler checklist |
| Ages | 4 years and older, provided mental age exceeds 2 years |

| | |
|---|---|
| **Edition and year** | 2003, with Lifetime and Current forms |
| **Publisher** | **WPS** |
| **Qualification level** | Completed by a parent; scored and interpreted by a trained professional in accordance with the publisher's requirements |
| **Time** | Under 10 minutes to complete |
| **Score types** | A raw total score from 0 to 39, since the first question is not scored |
| **Cutoff scores** | 15 or above on the Lifetime form warrants referral for a comprehensive evaluation |
| **Reliability** | High in the development studies; see the manual for figures |
| **Classification accuracy** | Better in clinical samples of older children and lower in young children and in community screening, so a score below 15 does not rule out the disorder when concerns persist |
| **Year of norm collection** | It has no age norms. The cut-off score derives from British diagnostic samples in the late 1990s |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The total score is read against the cut-off alone: above it, refer; below it, no referral for now unless other strong concerns exist. The score is not used to estimate severity or to determine the level of support.

### Limitations

It depends on the parent's memory and understanding of the questions, and an anxious or an optimistic parent may answer in keeping with that anxiety or optimism. It misses many girls who camouflage their traits, as well as students

whose difficulties surface only in middle school. A closely related instrument is the **Social Responsiveness Scale, Second Edition** (**SRS-2**), completed by a parent, a teacher or the individual, from age two and a half to adulthood. It has 65 items and yields a T-score (a standard score with a mean of 50 and a standard deviation of 10): 59 or below is within normal limits, 60 to 65 mild, 66 to 75 moderate, and 76 or above severe (Constantino & Gruber, 2012). Its advantage is that it measures traits on a continuum, which makes it well suited to comparing home and school ratings.

## CARS-2 Childhood Autism Rating Scale, Second Edition

Rating scale

### The idea

Once the specialist has observed the child and gathered information about him, an organized way is needed to turn those impressions into numbers that can be compared and tracked. That is what this scale offers: fifteen domains, in each of which the specialist rates how far the child's behavior departs from what is expected of same-age peers (Schopler et al., 2010).

### What it looks like

An examiner's manual and three booklets, chosen according to the child:

- **Standard Version ST:** for children under 6, or older children with marked communication difficulties or below-average cognitive ability.
- **High-Functioning Version HF:** for fluent speakers aged 6 or older with an IQ above 80.
- **Questionnaire for Parents or Caregivers QPC:** completed by the parents and used by the specialist to inform the ratings; it is not scored.

The Standard Version's domains include relating to people, imitation, emotional response, body use, object use, adaptation to change, visual and listening response, verbal and nonverbal communication, and activity level.

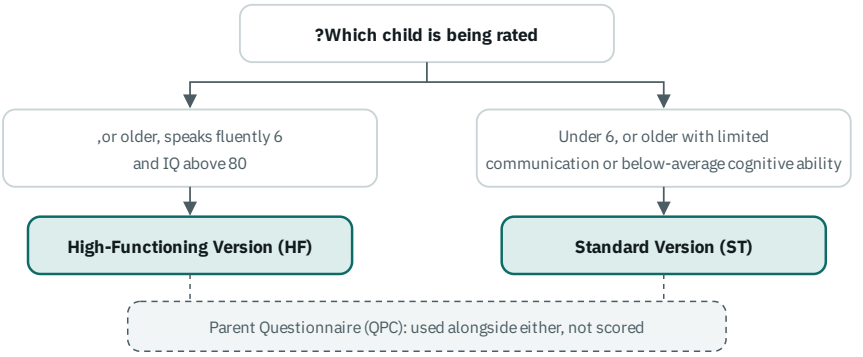



**Figure 9.6:** An original schematic for choosing the right booklet. A mistake here changes the meaning of the score, as in Layan's case (9.8).

## A scene from administration

The specialist observes Hisham, who is five, for about an hour as he plays and responds to simple activities, and reads through his mother's questionnaire. He then takes the rating booklet and gives each domain a score from 1 to 4. In "adaptation to change," for instance, he noted that Hisham cried hard when his toy train was taken away so that he could be given another toy, and kept crying for several minutes despite efforts to soothe him; he rated it 3, moderately abnormal. In "imitation," Hisham copied most simple movements after they were repeated, so he rated it 2.5.

**Table 9.7: CARS-2 information card.**

| | |
|---|---|
| **Purpose** | Rating the severity of autistic traits and supporting diagnosis |
| **Ages** | 2 years and older |
| **Edition and year** | Second, 2010 |

| | |
|---|---|
| **Publisher** | **WPS** |
| **Qualification level** | A specialist trained in autism assessment, after observing the child and gathering information about the child |
| **Time** | 5 to 10 minutes for the rating itself, after observation |
| **Score types** | Each domain is rated from 1 (age-appropriate) to 4 (severely abnormal), with half-points allowed, so the raw total ranges from 15 to 60. The manual provides a T-score and a percentile rank relative to a sample of individuals with autism. The percentile rank is the percentage of the comparison sample who scored lower than the child; a rank of 84 means the child's score exceeds those of 84% of that sample |
| **Cutoff scores** | On the Standard Version for children under 13: 30 and 37 (Figure 9.7). The High-Functioning Version and those aged 13 or older have other cut-off scores given in the manual |
| **Reliability** | The manual reports high internal consistency and interrater agreement; see it for figures |
| **Classification accuracy** | The manual reports good accuracy at the cut-off scores. Results should be read with caution in severe intellectual disability, because some domains are affected by overall developmental level |
| **Year of norm collection** | The Standard Version retains the items and scoring of the first edition (1988), and the autism sample from which the T-scores are calculated was gathered shortly before the second edition appeared |

| | |
|---|---|
| **Arabic version** | An Arabic edition was published by the American University of Beirut Press, prepared by Leila Akoury-Dirani and Mayssam Alameddine (Akoury-Dirani & Alameddine, 2016 [Arabic]); we found no published description of its sample. An Arabic version was studied with 301 children aged 2 to 12, yielding an internal consistency of 0.79, and factor analysis supported a three-factor, 14-item model (Alotaibi & Alotaibi, 2021). An adapted Iraqi version of the Standard Version, studied with 286 children in Baghdad, has also been published (Al-Shibly et al., 2025). We know of no published Arabic norms for the T-scores |

### How are the results interpreted?

The specialist sums the scores of the fifteen domains and then reads the total on the following scale. If Hisham's total is 32.5, his score falls in the “mild to moderate” range.

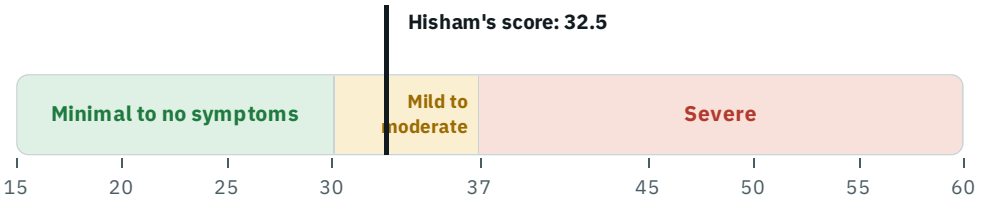



**Figure 9.7:** Total score ranges on the Standard Version for children under 13: below 30, minimal to no symptoms; 30 to 36.5, mild to moderate; 37 or above, severe.

In the report, the specialist writes something like: “The total score on the Standard Version indicates mild to moderate autistic traits,” and then links it to the rest of the evidence. He does not write “he has mild autism,” because the scale rates severity; it does not issue the diagnosis.

### Its limits

The High-Functioning Version, and those aged 13 or older, have other cut-off scores given in the manual. Using the wrong version can do a student an injustice: a fluent adolescent may score low on the Standard Version because its items do not capture his subtler difficulties. Publisher's page

## GARS-3 Gilliam Autism Rating Scale, Third Edition

Rating scale

### The idea

At home and at school, a child behaves in ways that may never surface in a short clinic session. That is why this scale was designed to be completed by someone who knows the student well and sees him every day: a parent or the **teacher**. It is among the instruments teachers are most likely to be asked to complete (Gilliam, 2014).

### What it looks like

An examiner's manual and a form that combines the items with the summary page. The form contains 56 statements describing behaviors; for each, the respondent rates how well it applies to the student on a four-point scale, from "not at all like the individual" to "very much like the individual." (The previous edition asked how often the behavior was observed, so mind the difference if you come across an old form.) The items are distributed across six subscales: Restrictive/Repetitive Behaviors, Social Interaction, Social Communication, Emotional Responses, Cognitive Style, and Maladaptive Speech.

### A scene from administration

The school psychologist sends the form to the second-grade teacher and asks him to rate Omar's behavior as he has seen it at school over recent weeks. The teacher sits down for ten minutes, reading each statement and recalling actual

situations before choosing. Omar's mother completes the same form about his behavior at home. The psychologist then converts the raw scores into **scaled scores** (**Scaled scores**), standard scores with a mean of 10 and a standard deviation of 3, and compares the two ratings.

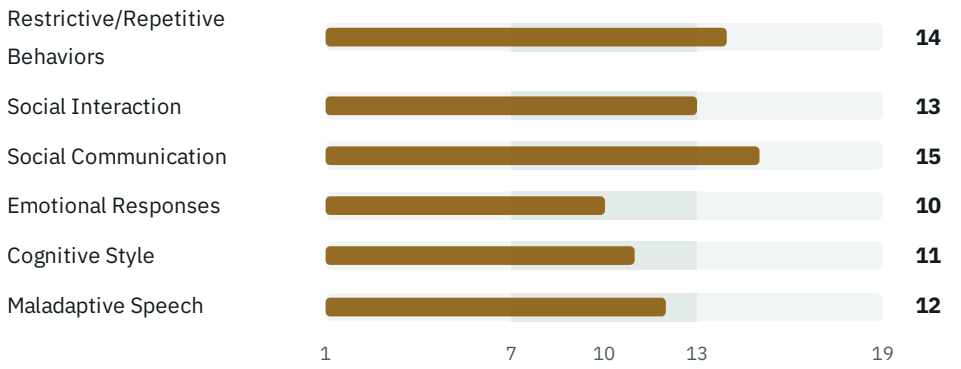



**Figure 9.8:** An illustrative profile of a teacher rating, using hypothetical values. The mean scaled score is 10, and the gray band represents the range from 7 to 13, one standard deviation either side of the mean. On this scale, a higher score means more autism-related behavior.

**Table 9.8: GARS-3 information card.**

| | |
|---|---|
| **Purpose** | Estimating the probability and severity of autism from the observations of someone who sees the student daily |
| **Ages** | 3 to 22 years |
| **Edition and year** | Third, 2014 |
| **Publisher** | **PRO-ED** |
| **Qualification level** | Completed by parents, a teacher or a specialist; interpreted by a qualified specialist |
| **Time** | 5 to 10 minutes |

| | |
|---|---|
| **Score types** | Subscale scaled scores with a mean of 10, and an Autism Index with a mean of 100 and a standard deviation of 15 |
| **Cutoff scores** | The probability of autism and the severity level are read from the manual's tables according to the Index score |
| **Reliability** | The manual reports high internal consistency for the Index and the subscales |
| **Classification accuracy** | The manual reports high accuracy, but independent studies of the first edition found that it placed a large share of children with a rigorously confirmed autism diagnosis in the lower probability categories (South et al., 2002) |
| **Year of norm collection** | A US reference sample of individuals with autism; see the manual for the year of data collection and sample size |
| **Arabic version** | An Arabic version was studied in Saudi Arabia with 178 participants aged 3 to 22 from educational and clinical settings, with a Cronbach's alpha of 0.971 (Alasmari et al., 2024a, 2024b). We know of no published Saudi norms for the Index, so scores are read against the US norms |

### How are the results interpreted?

The specialist chooses between two indexes: one built from four subscales for individuals whose speech is absent or very limited, and one built from all six subscales for those with functional speech. The reason is that the remaining two subscales revolve around speech, so rating them makes no sense for a child who does not talk. The specialist then reads from the manual's tables the probability of autism and the severity level that correspond to the Index score.

### Its limits

The scale reflects what the respondent perceives of the behavior. A teacher with thirty students may notice different things from a mother, and an anxious family may exaggerate or play things down. A disagreement between the two

ratings is therefore important information in its own right, and the scale alone is not enough for diagnosis. Independent studies have also clearly criticized the diagnostic accuracy of this scale series, particularly its tendency to underestimate the probability of autism in children who have been diagnosed (South et al., 2002). The high internal consistency of the Arabic version does not answer this question: consistency shows that the items measure a single thing, whereas accuracy in separating those with autism from those with look-alike conditions requires diagnostic validity studies. The scale is thus used to convey what the teacher and the family see, while judgments of probability are left to the rest of the evidence. Publisher's page

**For the teacher: How do you complete a rating scale honestly?**

1. Read each statement and bring to mind actual situations you have seen for yourself.
2. Rate what you have personally observed during the period the manual specifies.
3. If you have had no opportunity to observe a particular behavior, tell the specialist rather than guessing.
4. Do not try to “help” the student qualify for services by inflating the rating, or to “protect” him from a diagnosis by lowering it. An honest rating is what serves him.

## ADI-R Autism Diagnostic Interview–Revised

Diagnostic instrument

### The idea

Autism is a developmental disorder, meaning its roots go back to the earliest years, and no one knows those years like the parents. Yet asking them a general question, such as “Was your child sociable?”, does not yield precise information; most parents will say yes. That is why this interview is structured: it asks about specific behaviors and requests real examples for every answer (Rutter, Le Couteur, et al., 2003).

### What it looks like

An interview booklet from which the specialist reads 93 questions in a set order, recording and coding the answers, then transferring the codes to an algorithm form that computes scores in each domain. Nothing is handed to the family to fill in; the whole interview is a conversation.

### A scene from administration

Jumana's mother sits with the specialist for nearly two hours. The specialist asks, “When Jumana was between four and five, if she found something she liked, what would she do? Would she bring it to show you?” The mother answers, “She'd play with it by herself in the corner.” The specialist asks, “Can you remember a particular time?” And so it goes with every question, because the interview is after the actual behavior, not the family's opinion of it.

**Table 9.9: ADI-R information card.**

| | |
|---|---|
| **Purpose** | A diagnostic instrument that gathers the developmental history through a structured interview with the family |
| **Ages** | Individuals with a mental age above 2 years |
| **Edition and year** | Revised, 2003 |
| **Publisher** | **WPS** |
| **Qualification level** | A specialist with formal training in administering and coding the interview |

| | |
|---|---|
| **Time** | Roughly an hour and a half to two and a half hours |
| **Score types** | Algorithm scores for each domain: reciprocal social interaction, communication, and restricted and repetitive behaviors, plus age of symptom onset. It does not yield standard scores |
| **Cutoff scores** | Social interaction 10; communication 8 for verbal and 7 for nonverbal individuals; restricted behaviors 3; with signs evident at or before 36 months |
| **Reliability** | The manual reports high agreement among trained interviewers |
| **Classification accuracy** | Good in clinical samples, and lower in very young children and those with very low mental age |
| **Year of norm collection** | It has no age norms. The algorithm is based on diagnostic samples |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

Each domain score is compared with that domain's cut-off on the algorithm form (Table 9.9). The result is consistent with autism when scores exceed the cut-offs in all three domains and signs of the disorder were evident at or before 36 months.

### Limitations

It is long and time-consuming, and it depends on the accuracy of the parents' memory and their candor. It can be hard to administer when the child has been in the care of people other than the parents from an early age, or when the family cannot recall the details of the early years.

## ADOS-2 Autism Diagnostic Observation Schedule, Second Edition

Diagnostic instrument

### The idea

If the interview just described looks at the past through the parents' eyes, this schedule looks at the present directly through the specialist's. It goes beyond watching the child play to creating carefully designed situations that invite interaction and communication, and then recording how the child responds. That is why it is described as the reference standard in diagnostic observation (Lord et al., 2012).

### What it looks like

A large wheeled plastic case containing the manual, a protocol booklet for each module, and more than 100 stimulus materials: blocks, noise-making toys, soap bubbles, wordless picture books, and dolls and props for a pretend birthday party. The stimulus materials are sold only as part of the complete kit.

Each module has a protocol booklet in which the specialist records observations during the session and codes them afterward; at the end is the algorithm page, where the codes become a score and a classification. Figure 9.9 summarizes this path:

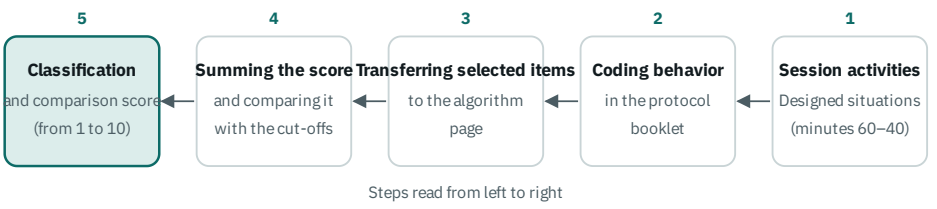



**Figure 9.9:** An original schematic of the path from session to classification in the observation schedule. Selected items from those coded, depending on the module, are carried over to the algorithm, and the comparison score is available only in Modules 1 to 3.

### How is the session room arranged?

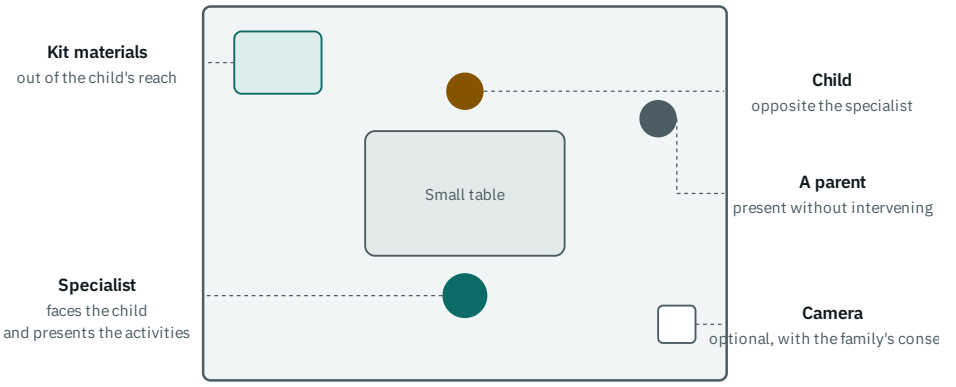



**Figure 9.10:** A suggested layout of the session room, seen from above. The materials are kept out of the child's reach so that he has to ask for them, since how he asks is part of what is observed.

### The five modules

From the schedule, the specialist selects **a single module** suited to the child's language level and age, and the choice begins with language before age:

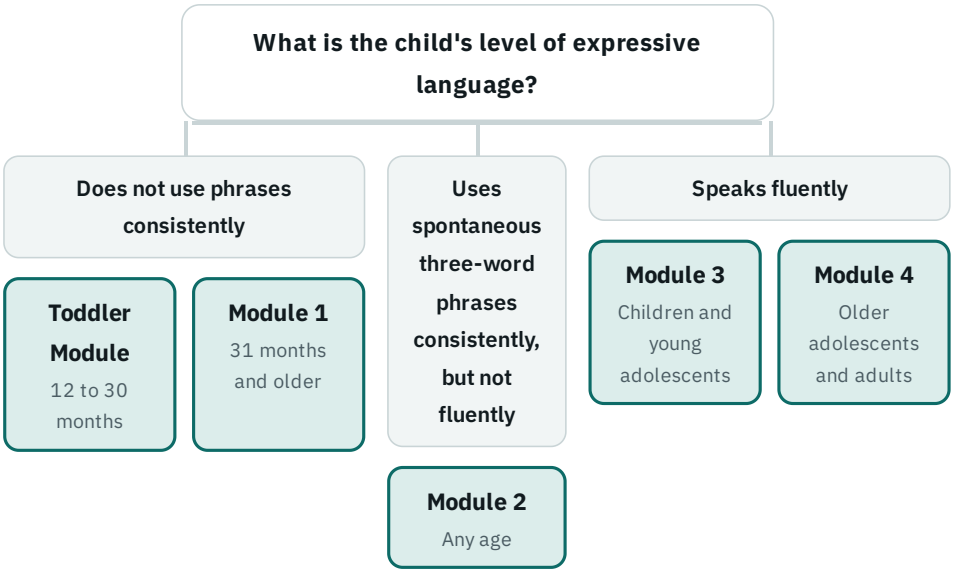



**Figure 9.11:** Choosing the right module. A ten-year-old who does not use phrases consistently is given Module 1 despite his age. Phrases here means phrases of three or more words that the child generates and uses consistently; memorized or echoed phrases do not count. The specialist may switch to another module if the child's language level turns out to differ from what was expected.

### Scene from administration: Module 3

Ziyad, a fluent nine-year-old, walks into a room with a table and two chairs. The specialist starts with a simple construction task, then asks him to describe a picture crowded with detail, and the two of them leaf through a wordless picture book together while Ziyad tells the story. Midway through the session, the conversation turns to friends: “Do you have friends? What makes someone your friend?” Then to feelings: “What makes you happy? What bothers you?”

To Ziyad, it all feels like ordinary conversation and play. But the specialist is watching closely: Does he look at him while talking? Does he share what catches his attention, or talk only about himself? Does he understand how the characters in the story feel? Does he ask the specialist about himself, even once? Straight after the session, the specialist codes the behaviors in the protocol booklet and then transfers the codes to the algorithm page.

**Table 9.10: ADOS-2 information card.**

| | |
|---|---|
| **Purpose** | A diagnostic instrument based on standardized direct observation |
| **Ages** | 12 months to adulthood, in five modules |
| **Edition and year** | Second, 2012 |
| **Publisher** | **WPS** |
| **Qualification level** | A specialist with formal training in the instrument; the publisher restricts purchase to those with advanced qualifications |
| **Time** | 40 to 60 minutes |
| **Score types** | Item codes, a total algorithm score, a classification, a comparison score from 1 to 10 in Modules 1 to 3, and ranges of concern in the Toddler Module |
| **Cutoff scores** | They vary by module, and by age and language level within a module, and are read from the algorithm page |
| **Reliability** | The manual reports good agreement among trained coders |
| **Classification accuracy** | Good in clinical samples; specificity drops with severe intellectual disability and in very young children |
| **Year of norm collection** | It has no age norms. The algorithms and comparison score are based on US diagnostic samples |
| **Arabic version** | At the time of writing, we know of no publisher-approved Arabic translation and no published standardized Arabic version |

### How are the results interpreted?

In Modules 1 to 4, the algorithm score is calculated and compared with cut-offs, yielding one of three classifications: “autism”; “autism spectrum,” which indicates milder symptoms; or “non-spectrum.” Modules 1 to 3 also provide a

“comparison score” from 1 to 10, which shows the level of symptoms relative to children diagnosed with autism of the same age and language level and is useful for tracking a child over time. The Toddler Module, by contrast, gives “ranges of concern” rather than a classification, because a formal classification may not be appropriate at such an early age.

### Its limits

It describes the child's behavior in a single session with a stranger. A tired or frightened child may appear worse than he really is, and some highly able children may perform well in a short, structured situation with an understanding adult. It is therefore always supplemented by the developmental history and information from school. Publisher's page

**For the teacher: Where do you fit in with these instruments?**

You will often be asked to complete the **GARS-3** or the **SRS-2**; the family may be asked to fill in the **SCQ**; you may be asked for information the specialist draws on for the **CARS-2**; and you may take part in screening young children in early intervention programs. The **ADI-R** and **ADOS-2**, however, are administered by a trained specialist; your role is to read their results in the report and understand what they mean for teaching the student.

## 9.7 How do you read a diagnostic report?

What usually reaches the teacher is a report several pages long, dense with numbers and terminology. The skill required is knowing where to look and what to extract, more than memorizing cut-off scores. Here is a hypothetical excerpt from Ziyad's report, with annotations showing how to read it:

Psychoeducational Evaluation Report

Name: (withheld) · Age: 9 years, 3 months

Module 3 of the Autism Diagnostic Observation Schedule[1] was administered because Ziyad speaks fluently in complex sentences. Ziyad was cooperative and gave correct, detailed answers to direct questions, but he rarely commented on the examiner's remarks or asked about the examiner, and for most of the conversation he returned to the topic of airplanes[2]. In the picture-story narration, he described the events accurately but mentioned the characters' feelings only when asked about them.

The total score fell within the "autism spectrum" classification, with a comparison score of 6 out of 10 in this session[3]. These results are consistent with the diagnostic interview with the family and with the teacher's rating on the Gilliam scale.

On the basis of the evidence as a whole, the criteria for autism spectrum disorder under DSM-5-TR are met, without accompanying intellectual impairment and without accompanying language impairment, at support Level 1 in both domains[4]. Direct instruction in reciprocal conversation skills is

1. **Which instrument, and why this module?** The choice of Module 3 tells you that Ziyad's expressive language is good.
2. **The behavioral description** is the most valuable part of the report for the teacher, because it pinpoints what needs teaching: initiating, sharing in conversation, and understanding feelings.
3. **The classification and comparison score** describe the severity of symptoms in the session; on their own, they do not constitute the diagnosis.
4. **The decision statement** brings all the evidence together and states the specifiers and the level of support. It is what the service decision rests on.

recommended, along with advance preparation before unfamiliar group activities.

**Figure 9.12:** A hypothetical excerpt from a diagnostic report, worded by the author, with annotations showing what the teacher takes from each part.

When you finish reading, ask yourself three questions: What strengths does the report mention that I can build on? What difficulties does it describe that show up in my classroom? And which recommendations can I put into practice tomorrow?

## 9.8 When is the judgment made? Three cases

Having presented the criteria and the instruments, we are left with the hardest question: how does all of this come together in a single decision? In this section we follow three cases from the moment of referral to the decision. In the first, the judgment is reached; in the second, it is not; and in the third, autism nearly escapes notice because the student's giftedness masks it.

All three cases are hypothetical, written for teaching purposes; their figures are illustrative and do not represent real children.

### Case one: Sara, 4 years 6 months

When the evidence converges

**The story:** Sara started kindergarten at the beginning of the year. Two months in, her teacher noticed that she spent free playtime alone, lining up colored blocks in the same order every day. When another girl tried to join her, Sara pushed her away and screamed. The teacher also noticed that she cried hard whenever the order of activities changed, and that she covered her ears whenever songs were played.

**What the teacher did:** Rather than relying on a general impression, she kept dated notes for two weeks in a log similar to Table 9.3. She then spoke with the mother, who said that Sara had been late to talk and began using short phrases only after age three; today she uses spontaneous three-word phrases consistently, though her speech is not fluent. The mother added that Sara pulls her mother's hand toward what she wants instead of pointing at it. The teacher passed her notes to the kindergarten administration, and Sara was referred first for hearing and vision screening and then to a diagnostic center.

### The evidence the team gathered:

**Table 9.11: Evidence in Sara's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision screening (before any instrument)** | Both normal | Her lack of response is not due to the senses, so the autism instruments become meaningful |
| **GARS-3, teacher rating** | High Autism Index, high probability | The kindergarten sees clear traits |
| **GARS-3, mother's rating** | Also a high Index | Home sees the same picture, so the traits are not tied to one setting |
| **CARS-2-ST** | 34 | Mild to moderate traits |
| **ADI-R with the mother** | Exceeded the cut-offs in all three domains, with signs before 36 months | The developmental history confirms that the symptoms are longstanding |

| Source | Result | What it means |
|---|---|---|
| **ADOS-2, Module 2** | "Autism spectrum" classification | Module 2 was chosen because she uses spontaneous three-word phrases consistently but not fluently (Figure 9.11). Direct observation confirms the traits at present |
| **Nonverbal intelligence test** | 92 | Her cognitive ability is average, so intellectual disability does not explain her symptoms |

| A ✓ | B ✓ | C ✓ | D ✓ | E ✓ |
|---|---|---|---|---|
| All three manifestations | Three of four manifestations | Before age three | Impairs participation | Average intelligence |

**How the team matched the evidence to the criteria:** Criterion A was met in all three of its manifestations: she does not share her interests with anyone, she neither points nor waves, and she does not play with peers. Three manifestations of Criterion B were met: lining up blocks in a fixed pattern, intense distress at changes in routine, and covering her ears during songs. The mother reported signs before age three, the symptoms keep Sara from taking part in kindergarten activities, and her intelligence is in the average range.

**Decision:** Sara is diagnosed with autism spectrum disorder, without accompanying intellectual impairment, with accompanying language impairment, at support Level 2 for social communication and Level 2 for restricted behaviors.

You may be wondering: how can the support level be 2 when her score on the **CARS-2** is in the mild-to-moderate range? The answer is that the team judges the level of support from Sara's day-to-day functioning at

kindergarten and at home; the scale score is only one of its inputs, and she genuinely needs substantial support to take part in any group activity.

**What does this mean for the teacher?** Sara will need a picture schedule that introduces activities in advance, preparation before any change, noise-reducing headphones during song time, and direct teaching of shared play skills, first with one child and then with a small group.

> **What evidence could change the decision?** If her participation in group activities improved markedly after months of the picture schedule and direct instruction, the support level would need to be reviewed; and if a comprehensive language evaluation showed that her language comprehension is much weaker than it appeared, the language impairment specifier and the interpretation of some of her symptoms would have to be reconsidered.

## Case two: Fahd, 7 years

When there is a better explanation

**The story:** Fahd's second-grade teacher complains that Fahd "lives in his own world." He sometimes does not turn when called, rarely talks with classmates, and sits alone at recess. The teacher had attended a training course on autism and began to see its signs in Fahd; he completed the Gilliam scale for him, and the probability of autism came out high. But the teacher had also noticed something that puzzled him: in some weeks Fahd responded well, and in others it was "as if he can't hear anyone." He took this to mean that Fahd "chooses when to engage."

**What the team did first:** Before any autism instrument, the team began with the second step of the pathway: hearing and vision screening. Fahd failed the school hearing screening in both ears and was referred to an

audiologist.

### The evidence the team gathered:

**Table 9.12: Evidence in Fahd's case, in the order it was gathered.**

| Source | Result | What it means |
|---|---|---|
| **GARS-3, teacher rating (reason for referral)** | High probability of autism | The teacher sees worrying traits at school |
| **School hearing screening** | Failed in both ears | Everything stops until the cause is known |
| **Audiological evaluation** | Fluid in the middle ear on both sides, indicated by tympanometry, and a mild conductive hearing loss. The family reports recurrent ear infections since kindergarten | **This is the key to the case**: hearing loss that fluctuates from week to week |
| **GARS-3, father's rating** | Low probability | Home does not see the same picture; this **discrepancy** deserves a pause |
| **Developmental interview with the father** | In his early years he pointed, waved and shared his interests, and played pretend games with his siblings | His history shows no restricted social or behavioral pattern |
| **Structured classroom observation after eight weeks of medical treatment and moving his seat to the front** | Turned to his name on 9 of 10 occasions and joined two classmates in a game at recess | Most of what worried the teacher disappeared once the hearing barrier was removed |

| Source | Result | What it means |
|---|---|---|
| **ADOS-2 and ADI-R** | Not administered | The team decided they were not needed once a better explanation emerged and the concerns resolved |

| A ✗ | B ✗ | C | D ✓ | E |
|---|---|---|---|---|
| Sound interaction when he can hear | No restricted patterns | Not applicable | Academic impact before treatment | Not applicable |

**How the team read the evidence:** Fluid in the middle ear makes hearing weaken and recover with the state of the ear, so a child hears well one week and poorly the next. This explains what puzzled the teacher: what he took for a "choice" to engage was hearing that came and went, especially when the teacher called him from behind amid the noise of thirty students. Fahd avoided his classmates at recess because he could not follow their talk over the noise, and he felt embarrassed. At home, where his father speaks to him face to face, his interaction is sound.

**Matching the criteria:** Criterion A was not met, because Fahd showed appropriate social reciprocity and nonverbal communication once the hearing barrier was removed, and throughout his developmental history. Criterion B was not met either. When the two core criteria are absent, there is no need to consider the remaining conditions. Specialists call what happened here **differential diagnosis** (**Differential diagnosis**): the visible symptoms found a better explanation in the hearing loss.

**Decision:** Fahd is not diagnosed with autism. He will be followed by an ear, nose and throat physician and an audiologist, with periodic hearing checks because the fluid may return with further infections; he will sit at

the front of the class near the teacher; and he will receive language support if the years of fluctuating hearing have left a mark on his language. The team will revisit his social functioning if difficulties persist once his hearing has stabilized.

**The lesson**

Had the team relied on the rating scale alone, Fahd would have been misdiagnosed and denied the services he actually needs. Had the observation schedule been administered before the hearing test, its result would have been open to doubt, whatever it showed. Sequence is part of the evidence: the senses first, then the autism instruments. Disagreement among sources and week-to-week fluctuation are two signals that call for stopping and investigating.

**What evidence could change the decision?** If his difficulties with social initiation and shared interest persisted for months after his hearing had stabilized, or if restricted, repetitive patterns of behavior that had not been evident emerged during follow-up, the assessment would need to be completed with the diagnostic interview and the observation schedule.

## Case three: Layan, 12 years

When giftedness masks autism

**The story:** Layan is one of the strongest sixth graders in mathematics and science, and she was nominated for the gifted program. Her program teacher, however, noticed something striking: Layan solves problems expertly on her own but falls apart in group projects. She insists on doing things her way and cannot understand why her classmates get upset when she corrects their mistakes in front of everyone. At every opportunity she talks about astronomy in endless

detail, without noticing that those around her have grown bored. Her mother says she comes home from school exhausted and spends hours alone in her room.

**Why was she identified so late?** Because her academic excellence led everyone to explain her behavior away as "the quirkiness of high achievers." Her excellent language covered her difficulties in social communication. She had also memorized polite phrases and used them in the right places, so she appeared well adjusted, as we explained in our discussion of girls in Section 9.4. Her difficulties became clearly visible when relationships among classmates grew more complex at this stage, which is exactly what Criterion C means when it states that symptoms may not become apparent until social demands exceed the child's capacities.

**What set her assessment apart:**

- The **High-Functioning Version CARS-2-HF** was used rather than the Standard Version, because she speaks fluently and her intelligence is high. Had the Standard Version been used, her score would have come out unfairly low.
- **Module 3** of the **ADOS-2** was administered; it includes a conversation about friendship and emotions, and it was there that her difficulties clearly emerged.
- In the interview, her mother recalled that Layan was reading by age four, yet she would move away from the other children in the family to line up and count objects, and she became upset if anyone moved them.

| A ✓ | B ✓ | C ✓ | D ✓ | E ✓ |
|---|---|---|---|---|
| Evident in conversation and interview | Routines and intense interest | Indicators in early childhood | Social impact and exhaustion | High intelligence |

**Decision:** Layan was diagnosed with autism spectrum disorder without intellectual disability and without language impairment, at support Level 1. She is therefore a **twice-exceptional** (**Twice-exceptional**) student: gifted and autistic at the same time (Chapter 13 returns to twice-exceptional learners in detail).

**What does this mean?** Layan remains in the gifted program and receives both services together: enrichment suited to her abilities, and support in social skills and in understanding others' perspectives, along with clear roles in group work. Here the diagnosis preserves her giftedness, explains what had been puzzling in her behavior, and opens the door to the support she needed.

> **What evidence could change the decision?** If follow-up showed that her social difficulties ease considerably when she finds peers who share her interests and abilities, or if an anxiety assessment revealed that her breakdowns in group projects are better explained by recent social anxiety than by a pattern extending back to early childhood, the team would need to revisit the diagnosis.

## 9.9 Differential diagnosis

The three cases have taught us that a single symptom can arise from different causes. Not responding to one's name may stem from autism, from hearing loss, or from deep absorption in a toy. This is why the team always makes sure that the symptoms are not better explained by another condition. The following are the main conditions that can resemble autism:

**Table 9.13: Conditions that can resemble autism and what distinguishes them.**

| Condition | Similarity to autism | Distinguishing features |
|---|---|---|
| **Social (pragmatic) communication disorder** | Difficulty using language socially: initiating conversation, taking turns, and understanding hints | No restricted, repetitive patterns of behavior, meaning that Criterion B is not met. This is the diagnosis in the manual closest to autism |
| **Intellectual disability** | Weak communication and social skills | The child's social skills are consistent with his overall level of intellectual development. The two disorders may co-occur |
| — | Delayed speech and little spontaneous talk | The child wants to interact, compensates with pointing, gestures and facial expressions, and shows no restricted patterns of behavior |
| **Hearing impairment** | Not responding to name, and delayed language | Eye contact and gestures are intact, and interaction improves when the child can hear well. Hearing loss may fluctuate with middle ear infections, so responsiveness looks "selective," as in Fahd's case |
| — | Social difficulties and impulsivity | The child understands social cues, but impulsivity and poor attention keep him from following them. The two disorders frequently co-occur |
| **Social anxiety and selective mutism** | Withdrawal and little speech in front of others | The child has social skills but fear holds them back, and he interacts normally at home and with familiar people |

| Condition | Similarity to autism | Distinguishing features |
|---|---|---|
| **Severe environmental deprivation** | Poor interaction and attachment | A history of severe neglect, with rapid improvement once care and environment improve |
| **Giftedness** | Deep, narrow interests, and difficulty with peers | The child's interests are flexible and he enjoys sharing them, and his difficulty eases when he finds peers of similar ability. The two may co-occur, as in Layan's case |

**Questions teachers ask themselves before referral**

1. Does the behavior appear in every situation, or only in one setting or with one person?
2. Have the child's hearing and vision been checked recently, beyond the newborn screening? Does his responsiveness fluctuate from one week to the next?
3. Does he interact normally with familiar people, at home or with a relative?
4. Has there been a major recent change in his life, such as a move, a loss, or a new baby?
5. Does he understand what is asked of him but not do it, or does he not understand it at all?

## 9.10 Co-occurrence and medical evaluation

Differential diagnosis asks: which of the two conditions? Co-occurrence, by contrast, asks: is there another condition alongside autism? In autism, co-occurrence is the rule rather than the exception. In the 2020 U.S. surveillance

data, 37.9% of autistic children with available ability data had intellectual disability (Maenner et al., 2023). A meta-analysis of published studies estimated the prevalence of attention-deficit/hyperactivity disorder among autistic people at about 28%, anxiety disorders at about 20%, sleep disorders at about 13%, and depression at about 11%, with wide variation across studies depending on sample and age (Lai et al., 2019). Since its fifth edition, the diagnostic manual no longer prohibits diagnosing autism and ADHD together.

**Table 9.14: Common co-occurring conditions in autism spectrum disorder and what their assessment requires.**

| Co-occurring condition | What we know about its frequency | Assessment guidance |
|---|---|---|
| **Intellectual disability** | Present in more than one third of autistic children in the U.S. surveillance data (Maenner et al., 2023) | Autism is diagnosed alongside it only if social communication is below what would be expected for general developmental level (Criterion E). The specificity of the observation schedule declines as mental age falls substantially, so false positives become more frequent; its result is read cautiously in severe disability, and developmental history is given greater weight |
| — | About 28% in the meta-analysis (Lai et al., 2019) | Some items on attention scales overlap with the social features of autism, so scores rise for reasons other than attention. The same attention and impulsivity symptoms are therefore asked about in situations that do not depend on social interaction, and from more than one source |

| Co-occurring condition | What we know about its frequency | Assessment guidance |
|---|---|---|
| **Anxiety** | About 20% (Lai et al., 2019) | It often shows up as a more rigid insistence on routine or as meltdowns at transitions, and is thus confused with restricted behavior. The questions to ask: Has the behavior increased with a new stressor? Can the student describe his fear? |
| **Sleep disorders** | About 13% as diagnoses, with complaints considerably more common (Lai et al., 2019) | Lack of sleep weakens attention and increases outbursts at school. The family is therefore asked about sleep before any decline in behavior is interpreted |
| **Epilepsy and gastrointestinal problems** | More common than among children in general; epilepsy is more frequent with intellectual disability | Unnoticed seizures, constipation and pain can surface as aggressive behavior or sudden withdrawal in a child who does not express himself in words. The child is referred to a physician before the new behavior is attributed to autism itself |

Medical and genetic evaluation is part of the comprehensive assessment, not an appendix to it. Current guidelines recommend offering genetic testing to every child diagnosed with autism, beginning with chromosomal microarray and fragile X testing, and asking about seizures, sleep, feeding and the gastrointestinal system (Hyman et al., 2020). Genetic referral becomes all the more important in communities where consanguineous marriage is common, Saudi society among them; its rate reached 56% in a national sample of mothers (El Mouzan et al., 2008). The working rule is that after every autism diagnosis, the team asks two questions: Is there anything in the file that autism alone does not explain? And has that area been measured with a tool suited to the student's developmental level?

## 9.11 In the Saudi and Arab context

**In the Saudi context**

Table 9.15 sets what this chapter has presented against the U.S. Individuals with Disabilities Education Act (**IDEA**) and against two Saudi documents: the Regulations for Special Education Institutes and Programs (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]) and the Special Education Organizational Guide, the most recent published edition we were able to find (Ministry of Education, 1436 AH [Arabic]). Educational criteria are criteria of **eligibility for services**, and their function differs from that of medical diagnostic criteria.

**Table 9.15: Autism spectrum disorder in the diagnostic manual and the international classification, in U.S. law, and in Saudi regulation.**

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Terminology** | Autism spectrum disorder | **Autism** | "Autistic disorder" in the Regulations, and "autism spectrum disorder" in the Guide |

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Definition** | Persistent deficits in social communication, and restricted, repetitive patterns | A developmental disability significantly affecting verbal and nonverbal communication and social interaction | Regulations: “a disorder that occurs in the child before the age of 36 months.” Guide: “persistent deficits in social communication and social interaction across multiple contexts,” worded to match Criterion A in the fifth edition of the diagnostic manual |
| **Age of onset** | The early developmental period, though symptoms may not become apparent until later | Generally evident before age three, which does not preclude identifying a child whose characteristics emerged later if the criteria are met | Regulations: before 36 months, an age restriction carried over from older definitions |
| **Impact** | Clinically significant functional impairment (Criterion D) | Adversely affects educational performance | Implicit in the definition of a student with a disability: a deficit “to the extent that requires the provision of special education services” |

| Element | DSM-5-TR and ICD-11 | IDEA (United States) | Saudi regulations |
|---|---|---|---|
| **Exclusion** | Symptoms are not better explained by intellectual disability | The category does not apply if the educational impact is primarily due to an emotional disturbance | We found no autism-specific exclusion clause in the documents we reviewed |
| **Who decides, and with what tools** | A qualified professional under each country's regulations | A team of professionals together with the parents, using multiple tools; no single procedure is sufficient | The assessment team, including the school principal, the special education teacher, the general education teacher, the student counselor, specialists according to the category, and the parent or guardian; “not relying solely on the results of a single scale,” with tools “approved by the Ministry of Education” |
| **Consent and timelines** | Not addressed | Parental consent, with the evaluation completed within 60 days of consent | Written consent after formal notice; no response within 14 days “is considered permission and consent.” The diagnosis is repeated 6 to 18 months after the initial diagnosis |

Sources: American Psychiatric Association (2022); WHO (2024); Assistance to States for the Education of Children With Disabilities (2006, §300.8(c)(1), §300.301, §300.304); Ministry of Education [Al-Ma'arif] (1422 AH [Arabic], Article 1); Ministry of Education (1436 AH [Arabic]).

Three observations follow from this comparison. **The first** is that the older Regulations' definition restricts onset to before 36 months, whereas the current diagnostic manual acknowledges that symptoms may not become apparent until social demands exceed the child's capacities. A student like Layan (9.8) could be excluded if the older text were read literally, even though her early indicators surfaced in the interview. **The second** is that treating silence as consent after fourteen days differs from the U.S. requirement of explicit consent, and it places an ethical burden on the school to make sure the family has actually understood the notice. **The third** is that the Guide requires tools "approved by the Ministry of Education," yet at the time of writing we found no published list of such tools. The Law on the Rights of Persons with Disabilities affirms the right of persons with disabilities to the "provision and adaptation of … modern measurement and diagnostic tools" and to "early intervention programs" (Article 8), and to "early detection of disability" (Article 9) (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic]).

Table 9.16 summarizes what we know about the Arabic versions of the instruments covered in this chapter:

**Table 9.16: Arabic instruments available for autism assessment, based on what we found published at the time of writing.**

| Instrument | What we know about its Arabic version | What is missing |
|---|---|---|
| **M-CHAT and M-CHAT-R/F** | Validation of the first edition in nine Arab countries: sensitivity 0.86, specificity 0.80 (Seif Eldin et al., 2008) | An Arabic norming study of the revised version with follow-up, in universal screening rather than a referred sample |
| **CARS-2** | An Arabic edition (Akoury-Dirani & Alameddine, 2016 [Arabic]), a factor-structure study with 301 children (Alotaibi & Alotaibi, 2021), and an adapted Iraqi version (Al-Shibly et al., 2025) | Arabic T-score norms, and classification-accuracy studies with comparison samples |
| **GARS-3** | An Arabic version studied in Saudi Arabia with 178 participants, with high internal consistency (Alasmari et al., 2024a, 2024b) | Saudi norms for the index, and sensitivity and specificity studies |
| **SCQ and SRS-2** | We know of no published standardized Arabic version | Standardization and accuracy studies |
| **ADI-R and ADOS-2** | We know of no published authorized or standardized Arabic version | A publisher-authorized translation, official training available in Arabic, and validation studies |

**A question for reflection:** If the reference diagnostic instruments have no authorized Arabic versions, on what evidence can a team at your school build a decision it can be confident in, and how should the report state the limitations of the tools it used?

## Conclusion

The argument we have developed in this chapter centers on the idea that a judgment of autism spectrum disorder is reached when evidence from developmental history, direct observation, and information from home and school converges to show that all five criteria are met, which makes it broader than anything a single scale can say. It follows that hearing and vision testing comes before any instrument, that a screening result is treated as an invitation to comprehensive assessment, that disagreement between sources is taken as a signal to look for another explanation, and that medical and genetic evaluation forms part of the comprehensive assessment, because co-occurrence is so common. The teacher plays a pivotal role in all of this: documented, objective observation; honest completion of rating scales; and an informed reading of the diagnostic report.

Questions about the numbers and the instruments nonetheless remain open. Prevalence figures vary with the way children are sought out, so they should not be read without asking how those children were reached, and most screening and diagnostic instruments lack published Arabic norms. Caution remains essential with those who tend to be identified late, such as gifted students and girls who mask their traits: giftedness can conceal autism, and autism does not cancel out giftedness.

# 9.12 Summary, key terms and activities

## Chapter summary

- Autism spectrum disorder is diagnosed when all five **DSM-5-TR** criteria are met: the three forms of social communication deficit, at least two of the four forms of restricted behavior, early onset, clear impact, and the absence of a better explanation by intellectual disability. **ICD-11** classifies the disorder by the presence of disorder of intellectual development and the degree of impairment in functional language, rather than by support levels.
- Prevalence figures vary with the way children are sought out, so the first question to ask of any figure is: how were these children found?
- Red flags call for evaluation without delay; the loss of any acquired skill calls for it at any age; and all children are screened at 18 and 24 months.
- Hearing and vision testing precedes any autism instrument, because an instrument's result for a child who does not hear well is open to doubt whatever it turns out to be.
- Screening tools sort but do not diagnose; rating scales convey what is seen by those who know the child; and diagnostic instruments are administered by a trained specialist. Most of these instruments lack published Arabic norms.
- No single scale is sufficient for reaching a judgment. The judgment is made when evidence from developmental history, direct observation, and information from home and school converges.
- Disagreement between sources is a signal to look for another explanation, such as hearing loss or a language disorder; co-occurrence is very common; and medical and genetic evaluation is part of the comprehensive assessment.

- Giftedness can mask autism, many girls camouflage their traits, and autism does not cancel out giftedness.
- The teacher's role is pivotal: documented, objective observation; honest completion of rating scales; and an informed reading of the diagnostic report.

## Glossary of terms

| | |
|---|---|
| **Joint attention** | When a child directs another person's attention to something, or follows that person's attention to it, so that the two share an interest in it, as when a child points to an airplane and looks at his mother. |
| **Echolalia** | Repeating words or phrases the child has heard from others, either immediately or after a delay, often in the same tone of voice. |

| | |
|---|---|
| **Screening** | A quick procedure applied to a large group to determine who needs a more in-depth assessment; it does not yield a diagnosis. |
| **Rating scale** | An instrument on which someone who knows the child rates the extent to which specific behaviors appear, using a graded scale. |
| **Cut-off score** | The score that separates two ranges in interpreting a scale, such as a score of 30 on the Standard Version of the **CARS-2**. |
| **Scaled score** | A standard score with a mean of 10 and a standard deviation of 3, used on subtests to compare a |

| | |
|---|---|
| | student's performance with the norming sample. |
| **Sensitivity** | The proportion of those who actually have the disorder whom the instrument correctly identifies. |
| **Specificity** | The proportion of those who do not have the disorder whom the instrument correctly rules out. |
| **Positive predictive value** | The probability that the disorder is actually present in someone with a positive result; it depends on the prevalence of the disorder in the group being tested. |
| **Differential diagnosis** | Establishing that the symptoms are not better explained by another condition. |
| **Co-occurrence** | The presence of another disorder alongside autism at the same time, such as intellectual disability, ADHD or anxiety; each is diagnosed and planned for. |
| **Functional language** | The child's ability to use language for everyday needs and communication; it is one of the two specifiers for classification in **ICD-11**. |
| **Catatonia** | Freezing in movement, holding fixed postures, extreme slowness, or purposeless movements; it may accompany autism and calls for urgent medical referral. |
| **Camouflaging** | A person's attempt to hide their traits by imitating others and memorizing ready-made phrases for |

| | |
|---|---|
| | social situations; it is more common among females. |
| **Twice-exceptional** | A gifted student who at the same time has a disability or disorder, such as autism or a learning disability. |

## Review questions

1. Why does the manual require all three forms of Criterion A to be met, while requiring only two of the four forms in Criterion B?
2. A 22-month-old child scored 5 on the **M-CHAT-R/F**. What is the next step? Can we say that he has autism? Use Figure 9.4 to justify your answer.
3. A fourteen-year-old student speaks fluently and has an IQ of 115. Which version of the **CARS-2** should be used with him? Which module of the **ADOS-2**? Why?
4. The teacher's rating on the **GARS-3** was high and the family's was low. Give three possible explanations for this disagreement, then explain why the high internal consistency of the Arabic version is not enough to justify a judgment based on the index alone.
5. Returning to Fahd's case: what evidence could the teacher have observed in class that might have alerted him to possible hearing loss? Why did the team place the hearing test before the observation schedule?
6. In Sara's case, her **CARS-2** score fell in the mild-to-moderate range, and her support level was set at Level 2. Is this a contradiction? Explain.
7. How would you distinguish between a gifted student with deep interests and a twice-exceptional student who is autistic?
8. Classify Sara's and Layan's cases using the **ICD-11** approach, then explain what the description of support levels in the diagnostic manual adds and

what the description of functional language in the international classification adds.

9. Compare the prevalence figures for Riyadh (2.51%) and for Makkah and Jeddah (2.81 per 1,000). Give three methodological reasons for the difference, and explain which figure is better suited to estimating an education district's need for services, and why.

10. An autistic student's teacher rated his attention symptoms on a rating scale, and his score came out very high. What would you examine before accepting a co-occurring diagnosis of attention-deficit/hyperactivity disorder?

11. **Case for application:** Hatoon is fourteen and in the eighth grade (second year of middle school). The student counselor referred her because of repeated absences, isolation at recess, and crying spells before oral presentations. Her achievement is high in science and art. Her hearing and vision were tested last year and found to be normal. Her mother completed the Social Responsiveness Scale, yielding a T-score of 72; her science teacher completed it, yielding 58. Her mother also completed the Lifetime form of the Social Communication Questionnaire, with a score of 13. The mother recalls that as a young child she was "quiet, never asking for anything," lined up her stickers in a fixed order, became upset when the route to school changed, and has had an intense interest in horses since age six. Module 3 of the observation schedule was administered and yielded a classification of "non-spectrum," but the specialist noted that many of her responses seemed to be "memorized phrases." Hatoon says: "I get tired of playing the part of a normal girl all day." (a) Which criteria in the manual does the evidence support, and from which source? (b) How do you read the observation schedule result and the questionnaire score, given that both fall below the cutoff? (c) What is the strongest alternative explanation, and what evidence would distinguish it from autism? (d) What is your decision now: diagnosis, no diagnosis, or deferring judgment? Justify your decision, and specify what you would need in order to reach it.

**Table 9.17: Mapping chapter objectives to review questions and learning activities.**

| Objective | Assessed by |
|---|---|
| **1. Mapping behaviors to the criteria and the international classification** | Questions 1 and 8, and part (a) of Question 11 |
| **2. Writing objective observations and identifying red flags** | Activity 1 |
| **3. Sequencing the decision pathway and placing sensory screening first** | Question 5 and Activity 6 |
| **4. Classifying instruments and choosing the instrument and its version or module** | Questions 2 and 3, and Activities 2 and 5 |
| **5. Interpreting results and evaluating published figures and Arabic versions** | Questions 4, 6 and 9, part (b) of Question 11, and Activities 4 and 5 |
| **6. Judging the sufficiency of evidence and distinguishing autism from similar and co-occurring conditions** | Questions 7 and 10, Question 11 (case for application), and Activity 3 |

## Learning activities

Any activity that involves observing a child or using a real student's data is carried out with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a qualified specialist; its results are not used in any decision concerning the student and are communicated to the family only through official channels.

1. **Structured observation:** Observe a child in a natural setting over three sessions and record your observations in a log similar to Table 9.3, then rewrite a nonobjective observation in objective form, as in the first stage of 9.5. *Product: a three-session log and a rewritten observation. Time: three 20-minute sessions, plus about an hour for writing.*

2. **Examining the instrument:** Read the scene with Ziyad in the **ADOS-2** profile and the description of Module 3, then design a table listing the activities it mentions, the behavior each activity elicits, and the criterion it serves. If you have access to publisher-licensed training materials as part of official training, compare your table with them. *Product: a three-column table. Time: about 45 minutes.*
3. **Conflicting sources (group activity):** The group is given conflicting results from the teacher and the family and writes down what additional evidence is needed before any judgment, where it should be gathered, and in what order. *Product: an ordered list of evidence with a rationale for the order. Time: about an hour.*
4. **Reading a report:** Analyze the excerpt from Ziyad's report in Figure 9.12, or a real report with all student-identifying data removed, identify where each criterion was met, and then turn the recommendations into three measurable classroom goals. *Product: a criteria table and three goals. Time: about an hour.*
5. **A local instrument profile:** Choose an autism scale used in diagnostic centers in your region whose booklet and manual are available to you, and prepare a profile of it using the standard template in 9.6, with its twelve fields, including a statement of its licensing status and local norms. *Product: a completed profile. Time: about two hours.*
6. **The pathway in your school:** Consult the Special Education Organizational Guide and your education directorate's procedures, map the decision pathway as it actually operates, and then compare it with Figure 9.3: Who refers? When are hearing and vision tested? Who diagnoses? Which tools are approved? How long does it take from referral to decision? *Product: a comparison diagram and a critical paragraph. Time: about two hours.*

# Chapter references

Akoury-Dirani, L., & Alameddine, M. (2016). [Arabic edition of the Childhood Autism Rating Scale, Second Edition] [In Arabic]. American University of Beirut Press. https://www.aub.edu.lb/aubpress/pages/cars.aspx

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alasmari, M., Alduais, A., Qasem, F., Almaghlouth, S., & AlAmri, L. (2024a). Examining language, speech and behaviour characteristics: A cross-sectional study in Saudi Arabia using the Arabic version of Gilliam Autism Rating Scale–Third Edition. *Children, 11*(4), Article 472. https://doi.org/10.3390/children11040472

Alasmari, M., Alduais, A., Qasem, F., Almaghlouth, S., & AlAmri, L. (2024b). Predicting language, cognition, and social skills in individuals with autism spectrum disorder using the Arabic validation of Gilliam Autism Rating Scale–Third Edition. *Applied Neuropsychology: Child*. Advance online publication. https://doi.org/10.1080/21622965.2024.2417812

AlBatti, T. H., Alsaghan, L. B., Alsharif, M. F., Alharbi, J. S., BinOmair, A. I., Alghurair, H. A., Aleissa, G. A., & Bashiri, F. A. (2022). Prevalence of autism spectrum disorder among Saudi children between 2 and 4 years old in Riyadh. *Asian Journal of Psychiatry, 71*, Article 103054. https://doi.org/10.1016/j.ajp.2022.103054

Alotaibi, B. M., & Alotaibi, A. A. (2021). Exploratory and confirmatory factor analyses of the Arabic version of the Childhood Autism Rating Scale. *Research in Autism Spectrum Disorders, 86*, Article 101827. https://doi.org/10.1016/j.rasd.2021.101827

Alshaban, F., Aldosari, M., Al-Shammari, H., El-Hag, S., Ghazal, I., Tolefat, M., Ali, M., Kamal, M., Abdel Aati, N., Abeidah, M., Saad, A. H., Dekair, L., Al Khasawneh, M., Ramsay, K., & Fombonne, E. (2019). Prevalence and correlates of autism spectrum disorder in Qatar: A national study. *Journal of Child Psychology and Psychiatry, 60*(12), 1254–1268. https://doi.org/10.1111/jcpp.13066

Al-Shibly, M. N. R., AlHusseiny, A. H., Lami, F., Jasim, S. M., Jaber, O. A., Nayeri, N. D., Sabet, M. S., & Al-Gburi, G. (2025). Psychometric properties of an Arabic modified Childhood Autism Rating Scale 2 Standard Form. *International Journal of Methods in Psychiatric Research, 34*(4), Article e70047. https://doi.org/10.1002/mpr.70047

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Assistance to States for the Education of Children With Disabilities, 34 C.F.R. pt. 300 (2006) (§300.8(c)(1), §300.301, §300.304).

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale* (2nd ed.). WPS.

El Mouzan, M. I., Al Salloum, A. A., Al Herbish, A. S., Qurachi, M. M., & Al Omar, A. A. (2008). Consanguinity and major genetic disorders in Saudi children: A community-based cross-sectional study. *Annals of Saudi Medicine, 28*(3), 169–173. https://doi.org/10.5144/0256-4947.2008.169

Filipek, P. A., Accardo, P. J., Baranek, G. T., Cook, E. H., Jr., Dawson, G., Gordon, B., Gravel, J. S., Johnson, C. P., Kallen, R. J., Levy, S. E., Minshew, N. J., Prizant, B. M., Rapin, I., Rogers, S. J., Stone, W. L., Teplin, S., Tuchman, R. F., & Volkmar, F. R. (1999). The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 29*(6), 439–484. https://doi.org/10.1023/A:1021943802493

Gilliam, J. E. (2014). *Gilliam Autism Rating Scale* (3rd ed.). PRO-ED.

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M.-C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 47*(8), 2519–2534. https://doi.org/10.1007/s10803-017-3166-5

Hyman, S. L., Levy, S. E., Myers, S. M., & Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics, 145*(1), Article e20193447.

Khamees, H., Gulati, A., Vatsa, A., Aldosari, M., & Hoque, E. (2025). Estimating autism prevalence among children in the Kingdom of Saudi Arabia: A comprehensive multisource benchmarking approach. *Saudi Medical Journal, 46*(11), 1276–1279. https://doi.org/10.15537/smj.2025.46.11.20250781

Lai, M.-C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry, 6*(10), 819–829.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 56*(6), 466–474.

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers, 6*, Article 5. https://doi.org/10.1038/s41572-019-0138-4

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule* (2nd ed.). WPS.

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., . . . Shaw, K. A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveillance Summaries, 72*(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37–45. https://doi.org/10.1542/peds.2013-1813

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. WPS.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview–Revised*. WPS.

Sabbagh, H. J., Al-Jabri, B. A., Alsulami, M. A., Hashem, L. A., Aljubour, A. A., & Alamoudi, R. A. (2021). Prevalence and characteristics of autistic children attending autism centres in 2 major cities in Saudi Arabia: A cross-sectional study. *Saudi Medical Journal, 42*(4), 419–427. https://doi.org/10.15537/smj.2021.42.4.20200630

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood Autism Rating Scale* (2nd ed.). WPS.

Seif Eldin, A., Habib, D., Noufal, A., Farrag, S., Bazaid, K., Al-Sharbati, M., Badr, H., Moussa, S., Essali, A., & Gaddour, N. (2008). Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry, 20*(3), 281–289. https://doi.org/10.1080/09540260801990324

South, M., Williams, B. J., McMahon, W. M., Owley, T., Filipek, P. A., Shernoff, E., Corsello, C., Lainhart, J. E., Landa, R., & Ozonoff, S. (2002). Utility of the Gilliam Autism Rating Scale in research and clinical populations. *Journal of Autism and Developmental Disorders, 32*(6), 593–599. https://doi.org/10.1023/A:1021211232023

Taylor, L. E., Swerdfeger, A. L., & Eslick, G. D. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine, 32*(29), 3623–3629.

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

Chapter 10

# 10 Attention-deficit/hyperactivity disorder and behavioral and emotional disorders

Restlessness, distraction, anger and sadness come and go in every child's life; the essential question is when they reach the threshold of a disorder. This chapter covers categories in which hasty judgments are common in two opposite directions: an active child who is misdiagnosed, and a quiet adolescent girl who suffers in silence while no one notices.

---

**Two scenes in a single day**

In second grade, Faisal cannot stay in his seat for more than a few minutes. He blurts out answers before the teacher has finished the question, takes his classmates' pens without asking, and forgets his notebook almost every day. His teacher says, “He never settles down.”

> In sixth grade, Haya causes no trouble at all. She sits quietly, yet she misses half of what is said, leaves her assignments unfinished, and her grades have been slipping since the start of the year. Her teacher says, "Quiet and polite, but she drifts off a lot."

Which of the two students is in greater need of assessment? Perhaps both. Faisal, however, is the more likely to be referred, because his behavior disrupts the class, while Haya may be left waiting for a long time. And what each of them is going through may stem from something quite different from what appears on the surface. We return to them in the first and second cases, and then turn to two more: a first grader and a girl in middle school.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Match a student's data against the five criteria for attention-deficit/hyperactivity disorder, specify the presentation and the severity level, and name one difference between the DSM and the ICD.
2. Explain, with evidence, how the relative age effect and the quiet presentation can lead to misdiagnosis or to being overlooked, and propose a school procedure that limits each.
3. Classify a given student's problems as externalizing or internalizing, and distinguish among the abbreviated criteria for oppositional defiant disorder, conduct disorder, disruptive mood dysregulation disorder, anxiety, depression and posttraumatic stress disorder.
4. Apply the definition of the emotional disturbance category in US law, and its counterpart in Saudi regulations, to a given case, and critique the exclusion of "social maladjustment."

5. Select a suitable screening tool, rating scale and sources of information for a given student, interpret the resulting T-scores and the discrepancies among raters, and explain the limits of computerized tests.
6. Identify, in a given case, the causes that must be ruled out before diagnosis and where each fits in the decision pathway.
7. Judge whether the evidence in a new case is sufficient to warrant a judgment, and justify that judgment criterion by criterion.

## Introduction

Restlessness, distraction, anger and sadness are a familiar part of every child's life, and the line between what is expected at a given age and what rises to the level of a disorder is not always clear. This is why hasty judgments are so common in these categories, and in two opposite directions: an active child, or one younger than his classmates, is misdiagnosed because his behavior draws attention, while a quiet girl who is struggling goes unnoticed because she causes no disruption, much as the two scenes at the opening of the chapter illustrate. The matter is complicated further by the fact that the clinical categories of the medical manuals do not map onto the educational category, and that clinical diagnosis and eligibility decisions rest with two bodies that hold different authority.

We begin the chapter with attention-deficit/hyperactivity disorder (Section 10.1), then present its diagnostic criteria, presentations and specifiers (Section 10.2), and ask who is misdiagnosed and who is overlooked (Section 10.3). We then take up the clinical pictures of behavioral and emotional disorders (Section 10.4), the educational category of emotional disturbance (Section 10.5), the signs that warrant attention (Section 10.6), the decision pathway and universal behavior screening (Section 10.7), and instrument profiles (Section 10.8). We then apply all of this to four cases (Section 10.9), set out differential diagnosis and co-occurrence (Section 10.10), and close with the Saudi and Arab context (Section 10.11).

## 10.1 Attention-deficit/hyperactivity disorder

Attention-deficit/hyperactivity disorder (**Attention-Deficit/Hyperactivity Disorder, ADHD**) is a neurodevelopmental disorder characterized by a persistent pattern of inattention, hyperactivity-impulsivity, or both, to a degree that is inconsistent with the child's developmental level and that interferes with functioning in more than one setting.

A large worldwide review put its prevalence among children and adolescents at about 5% (Polanczyk et al., 2007). A later meta-analysis, pooling studies from many countries, estimated the prevalence of mental disorders overall in these age groups at about 13.4%: anxiety disorders at about 6.5%, disruptive behavior disorders at about 5.7%, attention-deficit/hyperactivity disorder at about 3.4%, and depressive disorders at about 2.6% (Polanczyk et al., 2015). In Saudi Arabia, a meta-analysis of fourteen Saudi studies arrived at a pooled prevalence for the disorder of 12.4%, with a wide confidence interval running from 5.4% to 26% (Aljadani et al., 2023). The width of this interval means that the pooled studies differ considerably from one another. The figure should therefore be read as a rough general estimate, and the gap between it and the global estimates does not justify concluding that the disorder is more prevalent in Saudi Arabia, at least not before differences in samples, instruments and criteria across the studies have been taken into account.

To understand this disorder, it helps to remember that attention is a set of related abilities: focusing on one thing while ignoring others, sustaining that focus long enough, shifting between tasks, and holding back the first response long enough to think. These abilities are tied to what are called **executive functions** (**Executive functions**), the capacities that regulate behavior and direct it toward a goal: planning, organization, working memory and self-control.

**Common misconceptions**

- **"If he can focus on video games for hours, he doesn't have an attention problem."** The problem is not the ability to focus as such, but directing it where it is needed. Fast-paced, stimulating activities capture attention automatically; it is the dull task that demands effort that proves difficult. The DSM states that signs may be minimal or absent when the child is receiving frequent rewards for appropriate behavior, is under close supervision, is in a novel setting, or is engaged in an especially interesting activity. The disappearance of the behavior during a favorite activity therefore does not rule out the disorder.
- **"It's caused by bad parenting."** This is a neurodevelopmental disorder with a strong genetic component. How the child is handled at home affects the severity of the accompanying problems, but it does not cause the disorder.
- **"Girls don't get it."** Boys are diagnosed more often, but the real difference is smaller than it appears, because girls more often show the quiet, inattentive presentation that does not attract notice.

## 10.2 Diagnostic criteria and specifiers

The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, text revision (**DSM-5-TR**), sets out two lists of nine symptoms each and requires five criteria to be met together (American Psychiatric Association, 2022):

**All five criteria must be met**

### (A1) Inattention

**At least six symptoms, for at least six months**

1. Fails to attend to details and makes careless mistakes
2. Has difficulty sustaining attention in tasks or play
3. Seems not to listen when spoken to directly
4. Does not follow through on instructions and fails to finish assignments
5. Has difficulty organizing tasks and activities
6. Avoids tasks that require sustained mental effort
7. Loses things necessary for tasks
8. Is easily distracted by external stimuli
9. Is often forgetful in daily activities

### (A2) Hyperactivity-impulsivity

**At least six symptoms, for at least six months**

1. Fidgets with hands or feet or squirms in seat
2. Leaves seat in situations where remaining seated is expected
3. Runs about or climbs in inappropriate situations
4. Is unable to play quietly
5. Is constantly on the go, as if driven by a motor
6. Talks excessively
7. Blurts out answers before the question is finished
8. Has difficulty waiting his or her turn
9. Interrupts or intrudes on others' activities

### (B) Before age 12

Several inattentive or hyperactive-impulsive symptoms were present before age twelve.

### (C) Two or more settings

Several symptoms are present in at least two settings, such as home and school, with friends or relatives, or in other activities.

### (D) Impact on functioning

Clear evidence that the symptoms interfere with, or reduce the quality of, social or academic functioning.

### (E) Exclusion

The symptoms do not occur exclusively during the course of schizophrenia or another psychotic disorder, and are not better explained by another mental disorder, such as a mood, anxiety, dissociative or personality disorder, or by substance intoxication or withdrawal.

**Figure 10.1:** **DSM-5-TR** criteria for attention-deficit/hyperactivity disorder, in abbreviated form. Five symptoms rather than six suffice for those aged 17 and older. The symptoms must be inconsistent with developmental level, and they are not counted if they are solely a manifestation of oppositional behavior, defiance, hostility or failure to understand the task. Note that a learning disability is treated as a differential or co-occurring diagnosis, outside the exclusion criterion (see 10.10), and that autism spectrum disorder has no longer been an exclusion since the fifth edition of the manual, so the two diagnoses may be given together.

Note criterion (C) in particular: **several symptoms** must be present in **at least two settings**. A child who is very active in only one class, or with only one teacher, calls for questions about that class or that teacher before questions about the child. This is why diagnosis draws on information from both the family and the school. The criterion does not, however, require the full symptom count in every setting: six symptoms at school and three at home may satisfy it. The question for the second setting is whether several genuine symptoms appear there and whether they have an impact; reaching six is not required.

## The three presentations

**Predominantly inattentive presentation**

**Inattention 6+, hyperactivity fewer than 6**

The child seems dreamy and slow, forgets, and loses things. Such a child bothers no one and is therefore identified late; this presentation is more common among girls.

**Combined presentation**

**Both lists 6+**

Attention difficulties occur together with excessive movement and impulsivity. The most common presentation in clinics.

**Predominantly hyperactive-impulsive presentation**

**Hyperactivity 6+, inattention fewer than 6**

Seen more often in early childhood; it may shift to the combined presentation with age as the demands on attention increase.

**Figure 10.2:** The three presentations. The presentation is specified on the basis of the past six months; it describes the current picture and may change with development.

## Severity and other specifiers

The manual supplements the presentation with a specification of current severity, and it provides labels for cases that do not fit the full picture:

**Table 10.1: Severity levels and other specifiers for attention-deficit/hyperactivity disorder in the DSM.**

| Specifier | When it is used |
|---|---|
| **Mild** | Few, if any, symptoms in excess of those required for the diagnosis, resulting in no more than minor impairment in social or academic functioning. |
| **Average** | Symptoms or functional impairment between mild and severe. |
| **Severe** | Many symptoms in excess of those required, or several symptoms that are particularly severe, or marked impairment in social or academic functioning. |
| **In partial remission** | Full criteria were previously met; fewer than the full criteria have been met for the past six months, while symptoms still result in impaired functioning. |
| **Other specified ADHD** | Symptoms cause clinically significant distress or impairment without meeting the full criteria, and the clinician states the reason, for example, that the inattentive symptoms fall short of the required number. |
| **Unspecified ADHD** | The same picture, but the clinician does not state the reason, or there is insufficient information for a more specific diagnosis, as in some emergency visits. |

These labels matter in practice when a report is written. “Combined presentation, moderate” tells the teacher and the family far more than the word “hyperactivity,” and it allows comparison at reevaluation. “Other

specified" guards against two things: a student being denied support because the symptom count fell one short, and a student being given a full diagnosis the evidence does not support.

**Comparison with ICD-11**

The International Classification of Diseases, eleventh revision (**ICD-11**), describes the disorder by the same core features: a persistent pattern of inattention or hyperactivity-impulsivity, lasting at least six months, emerging during the developmental period, exceeding what is expected for age and level of intellectual functioning, appearing in more than one situation, and interfering with functioning. It also has presentations that parallel those of the DSM. The clearest difference is that it describes the manifestations and leaves the judgment of their sufficiency to the clinician, **without a fixed number of symptoms**. Among the disruptive behavior categories, the ICD-11 has no separate category corresponding to "disruptive mood dysregulation disorder" (10.4); instead, it adds to oppositional defiant disorder a specifier, "with chronic irritability-anger." It names conduct disorder "conduct-dissocial disorder," and applies the specifier "with limited prosocial emotions" both to this disorder and to oppositional defiant disorder (World Health Organization, 2024). A teacher may thus read two reports on the same student that use two different labels; the difference lies in the manual used, not in the student.

# 10.3 Who is misdiagnosed and who is overlooked?

## The relative age effect in the classroom

A single class may hold children whose oldest and youngest members are separated by almost a full year. At age six, a whole year is a large difference in the ability to sit still, pay attention and exercise self-control. When a teacher compares a child with his classmates, the immaturity of the youngest may look like a disorder. This is known as the **relative age effect** (**Relative age effect**).

Layton and colleagues studied more than 400,000 American children in states where children enter kindergarten if they turn five before September 1. They found that children born in August, the youngest in their class, were diagnosed with the disorder at a clearly higher rate than those born in September, the oldest in their class (Layton et al., 2018):




**Figure 10.3:** Rate of attention-deficit/hyperactivity disorder diagnosis per 10,000 children by birth month, in states where September 1 is the school-entry cutoff date, from Layton and colleagues (Layton et al., 2018). The difference did not appear in states that use other cutoff dates.

The phenomenon recurs outside the United States. In Canada, Morrow and colleagues followed nearly one million children aged six to twelve in the province of British Columbia, where the cutoff falls at the end of December. Boys born in December, the youngest in their class, were about 30% more likely to be diagnosed than those born in January, and girls born in December were about 70% more likely (Morrow et al., 2012). These are two observational

studies: they establish that the likelihood of diagnosis is linked to relative age, but they do not establish that every additional case is a misdiagnosis. They do, however, make it likely that immaturity is sometimes read as disorder.

In Saudi Arabia, a child's age for admission to first grade is calculated on the first day of the school year: the minimum is six years, and a child up to ninety days younger may be admitted. The regulations for 1448 AH required that these younger children have completed a full year of kindergarten (Ministry of Education, 2025 [Arabic]; Sabq, 2026 [Arabic]). A single first-grade class may thus include one child who turns six only months after school begins and another who is nearly seven, a gap of almost a year, which may widen if a child enrolled late or repeated a grade. At the time of writing we found no published Saudi study of the relative age effect in diagnoses of the disorder, but there is no reason to assume that children here are immune to it.

The practical lesson for the teacher: before referring a child for restlessness and inattention, look at his date of birth, and compare him with the classmates closest to him in age rather than with the whole class.

## Girls, the quiet presentation and adolescents

Many girls, by contrast, are overlooked, and with them boys with the predominantly inattentive presentation. Teachers tend to refer the students who disrupt the class, and a girl whose mind wanders quietly disturbs no one. Her difficulty may not become clearly visible until middle or high school, when the volume of coursework and the need for self-organization increase; her grades fall, and everyone assumes she has grown careless. A difficulty that surfaces late does not negate criterion (B), provided that several symptoms were present before age twelve, even if no one noticed them. The evaluator therefore extends the inquiry to elementary school records and the observations of her teachers there, alongside the present.

The same holds, even more clearly, for emotional problems. Anxiety and depression increase with the onset of adolescence and show up as absences, physical complaints, falling grades and withdrawal from friends, signs that are easily read as laziness or as "a phase that will pass." We will see this in the case of Lamya (10.9).

## 10.4 Behavioral and emotional disorders: clinical pictures

This category covers many conditions that share one feature: the student's behavior or emotions interfere with learning and relationships to a clear and persistent degree. It helps to divide the problems into two kinds: externalizing problems (**Externalizing**) and internalizing problems (**Internalizing**):

### Externalizing problems

Directed at others and the environment, and visible to everyone

- Aggression and fighting
- Anger, defiance and refusal to follow instructions, as in oppositional defiant disorder
- Rule-breaking and harming others, as in conduct disorder
- Severe, recurrent temper outbursts, as in disruptive mood dysregulation disorder
- Impulsivity and hyperactivity

### Internalizing problems

Directed inward, and seen only by those who pay attention

- Excessive worry and fears
- Withdrawal and silence, including selective mutism
- Persistent sadness and loss of interest, which in children and adolescents may appear as irritability
- Recurrent physical complaints with no medical cause, such as stomachaches on school-day mornings
- Effects of trauma, such as avoiding reminders of the event and hypervigilance

**Figure 10.4:** Externalizing and internalizing problems. The former are often referred because they disturb others; the latter are often overlooked because they disturb no one but the student. Both kinds may occur in the same student.

Teachers are not expected to diagnose these disorders, but knowing their abbreviated criteria helps in two ways: it enables the teacher to describe what he or she sees with a precision useful to the specialist, and it makes clear that a single behavior may stem from quite different disorders. A student who refuses to do homework may be defiant, anxious about failing, depressed and without interest in anything, or simply unsure what is being asked. Table 10.2 summarizes the main disorders as the DSM presents them.

**Table 10.2: Abbreviated criteria for the main behavioral and emotional disorders of childhood and adolescence in the DSM (adapted and abridged).**

| Disorder | Core picture | Duration and count | Specific to children and adolescents |
|---|---|---|---|
| **Oppositional defiant disorder** | A pattern of angry/irritable mood, argumentative/defiant behavior, or vindictiveness, in three groups comprising eight symptoms | At least four symptoms, for six months, with at least one individual who is not a sibling | Under age five: on most days; age five and older: at least once a week. Severity depends on the number of settings: mild in one setting, moderate in two, and severe in three or more |

| Disorder | Core picture | Duration and count | Specific to children and adolescents |
|---|---|---|---|
| **Conduct disorder** | A repetitive and persistent pattern of violating the rights of others or major age-appropriate societal norms, in four groups: aggression toward people and animals, destruction of property, deceitfulness or theft, and serious violations of rules | At least three of fifteen behaviors in the past twelve months, with at least one in the past six months | A childhood-onset type (at least one symptom before age ten) and an adolescent-onset type. The specifier **with limited prosocial emotions** (**With limited prosocial emotions**): at least two of four traits, over twelve months, across more than one relationship and setting: lack of remorse or guilt, callousness and lack of empathy, unconcern about performance, and shallow affect |

| Disorder | Core picture | Duration and count | Specific to children and adolescents |
|---|---|---|---|
| **Disruptive mood dysregulation disorder (Disruptive mood dysregulation disorder, DMDD)** | Severe, recurrent temper outbursts, verbal or behavioral, out of proportion to the situation and inconsistent with developmental level, with irritable or angry mood most of the day between outbursts, observable by others | Three outbursts a week on average, for twelve months, with no symptom-free period of three months, in two of three settings and severe in at least one | Not diagnosed for the first time before age six or after age eighteen, with onset before age ten. It is not given together with oppositional defiant disorder; if the criteria for both are met, only this diagnosis is given |
| **Anxiety disorders** | Excessive fear or anxiety out of keeping with developmental level, with avoidance and an impact on functioning. The most common at school are separation anxiety, generalized anxiety, social anxiety, specific phobia and selective mutism | Separation anxiety: at least four weeks in children and adolescents; generalized and social anxiety: typically six months; selective mutism: at least one month, not limited to the first month of school | In generalized anxiety, one physical or associated symptom rather than three suffices in children. In social anxiety, the anxiety must occur in peer settings, not only with adults, and the child may express it by crying, tantrums, freezing, clinging or not speaking |

| Disorder | Core picture | Duration and count | Specific to children and adolescents |
|---|---|---|---|
| **Major depressive disorder** | At least five of nine symptoms, including depressed mood or loss of interest or pleasure, representing a change from previous functioning | At least two weeks | In children and adolescents, mood may be **irritable** rather than sad, and weight loss may appear as failure to make expected weight gains. In persistent depressive disorder, one year suffices for them rather than the two years required for adults |
| **Posttraumatic stress disorder (Posttraumatic stress disorder, PTSD)** | After exposure to a traumatic event: intrusive re-experiencing of the event, avoidance of reminders, negative changes in thoughts and mood, and heightened arousal and reactivity | More than one month | Re-experiencing may appear as repetitive play in which themes of the event recur, and as frightening dreams without recognizable content. Children aged six and younger have separate criteria that rely less on language |

Readers will notice that irritability recurs in more than one row of the table: in oppositional defiant disorder, in disruptive mood dysregulation disorder and in adolescent depression, and it may also accompany anxiety and trauma. Describing a student as "short-tempered" therefore calls for detail: when the student gets angry, how often, what the student is like between outbursts, when this began, and whether it represents a change from before. These very questions are what distinguish the disorders in the table from one another.

> ⚠ **When the matter cannot wait for assessment**
>
> If a student expresses, in speech, writing or drawing, a wish to harm himself or herself or to end his or her life, shows signs of injury, or discloses having been harmed, these are not data to be added to the assessment file. They are a situation that calls for immediate action under school procedures: informing the school counselor and the school principal the same day, not leaving the student alone, and involving the family and the competent authority as the procedures require. Completing the assessment comes afterward.

## 10.5 The educational category: emotional disturbance

The disorders in Table 10.2 are clinical diagnoses. The education system, however, asks a different question: does this student need special education services because of his or her behavior or emotions? For this reason, education systems often group these conditions in a broad educational category with its own definition. Under the US Individuals with Disabilities Education Act (**IDEA**), reauthorized in 2004 as the Individuals with Disabilities Education Improvement Act, a student qualifies for services under the category of **emotional disturbance** (**Emotional disturbance**) if he or she exhibits one or more of the following five characteristics over a long period of time and to a

marked degree, adversely affecting educational performance (Individuals with Disabilities Education Improvement Act, 2004; U.S. Department of Education, 2006, 34 CFR §300.8(c)(4)):

1. An inability to learn that cannot be explained by intellectual, sensory or health factors.
2. An inability to build or maintain satisfactory relationships with peers and teachers.
3. Inappropriate types of behavior or feelings under normal circumstances.
4. A general pervasive mood of unhappiness or depression.
5. A tendency to develop physical symptoms or fears associated with personal or school problems.

The definition then adds two sentences that weigh heavily on the decision: **that the category includes schizophrenia**, and **that it does not apply to students who are "socially maladjusted"** (**Socially maladjusted**) **unless it is determined that they have an emotional disturbance** under the characteristics above, according to the US Code of Federal Regulations (34 CFR §300.8(c)(4)(ii)).

Note the three shared conditions: a long period of time, a marked degree, and an adverse effect on learning. A student who has gone through two difficult weeks after the death of a grandfather is not placed in this category, although he or she may need support. Note also that the educational category does not coincide with clinical diagnosis: a student may be diagnosed with anxiety without its effect on learning reaching the level that qualifies for the category, and a student may qualify for the category on the strength of clear, documented characteristics before any particular clinical diagnosis has been completed.

**For the specialist: the debate over excluding "social maladjustment"**

US regulations do not define the term "social maladjustment," so states and school districts have interpreted it in different ways. Some equate it with conduct disorder or delinquency and see in it "willful" behavior, in which the student knows the rules and chooses to break them, as opposed to an "involuntary" disorder that merits services. This distinction has been widely criticized in the behavioral assessment literature (Merrell, 2008), for reasons that include the following:

- **No measurable criterion:** no standardized instrument separates "maladjustment" from "emotional disturbance," so the decision comes to rest on the team's judgment.
- **Actual overlap:** many students with externalizing problems also have anxiety, depression, the effects of trauma or attention-deficit/hyperactivity disorder, so separating the two kinds within a single student is often artificial.
- **The impact of exclusion:** the students most at risk of dropping out and of school discipline may be denied services, even though their need for structured support is plainer.
- **Risk of bias:** in the absence of a criterion, the label "maladjusted" may be shaped more by the student's background and reputation than by the evidence (see Chapter 14).

Defenders of the exclusion argue, on the other hand, that the category was created for students whose emotional disturbance impedes their learning, and that folding every behavioral infraction into it would empty it of meaning. The practical position this book adopts is that the exclusion

should be applied only after a comprehensive assessment that looks explicitly for the five characteristics, and that the evidence on which it rests should be stated in the report.

Saudi regulations, for their part, treat "behavioral and emotional disorders" as a distinct category of special education, alongside "attention deficit and hyperactivity disorders" (Ministry of Education, 1436 AH [Arabic]). Section 10.11 sets out what we found in these texts and what we did not.

## 10.6 Signs that warrant attention

The following signs are warning signals. Their value lies in the teacher documenting them and discussing them with the school counselor or the support team, especially when they persist, appear in more than one situation, or represent a clear change from the student's previous state.

**Table 10.3: Signs that warrant attention, by school level.**

| Stage | Attention and activity | Externalizing behavior | Emotion (internalizing) |
|---|---|---|---|
| **Kindergarten and early grades** | Restless movement exceeding that of the classmates closest in age, great difficulty waiting for a turn, and repeated accidents caused by impulsivity | Temper outbursts far more intense and prolonged than expected for the child's age, and repeated aggression toward peers | Intense crying at separation that lasts for weeks, complete silence at school with normal speech at home, and recurrent morning physical complaints |

| Stage | Attention and activity | Externalizing behavior | Emotion (internalizing) |
|---|---|---|---|
| **Upper elementary grades** | Assignments left unfinished despite understanding, repeated careless errors, and materials constantly lost | Constant arguing with adults, refusal to follow instructions with more than one teacher, and lying or taking others' belongings | Excessive fear of mistakes or of tests, avoidance of participation, withdrawal from play, and sadness or irritability lasting for weeks |
| **Middle and high school** | A sudden decline as the number of subjects grows, inability to plan projects and deadlines, and quiet inattention that had gone unnoticed before | Unexcused absences, serious rule violations, aggression, associating with peers who break rules, and frequent intense irritability | Increasing absences, falling grades, withdrawal from friends, changes in sleep or eating, and expressions of hopelessness or worthlessness (see also the warning box in 10.4) |

# 10.7 The decision pathway

**1 Documented observation**

Record the behavior with an operational definition, that is, a description that can be observed and counted, across different situations, together with what precedes and follows it. Consider the student's relative age within the class and any recent change in his or her life.

**2 Sensory check, screening and classroom support**

Confirm that hearing and vision are intact as a precondition for any diagnostic instrument. Then conduct universal behavior screening that captures externalizing and internalizing problems together, or use a broad screening measure such as the Strengths and Difficulties Questionnaire. Alongside these, make adjustments to the environment, instruction and behavior management, and monitor their effect.

Persists despite support: referral

**3 Ruling out other causes**

Ask about sleep and health, assess achievement to rule out a learning disability, and ask about family circumstances, recent changes and traumatic events. Formal referral begins with the parent's written consent.

**4 Comprehensive multi-informant assessment**

Rating scales from the family and from two teachers where possible, and from the student if old enough; a developmental history interview; direct observation; and a functional behavioral assessment (**Functional Behavioral Assessment, FBA**) when the function the behavior serves needs to be understood.

**Are the criteria met in two settings, with impact, and with no better**

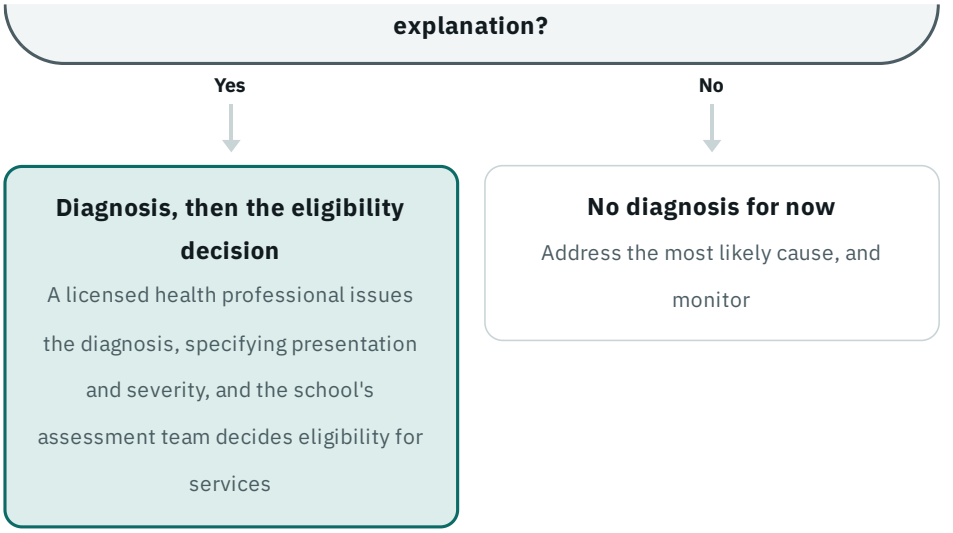



**Figure 10.5:** The decision pathway for attention-deficit/hyperactivity disorder and behavioral and emotional disorders. Hearing and vision screening comes in the second step, before any diagnostic instrument, and the third step deliberately precedes the comprehensive assessment, because inattention and difficult behavior have many possible causes.

## Universal behavior screening

Schools usually wait for a student to catch the teacher's eye, which means that a student who disturbs no one may never be seen at all. For this reason, schools that operate a multi-tiered system of supports are turning to **universal behavior screening** (**Universal behavior screening**): the teacher rates every student in the class, referred or not, with a brief tool two or three times a year, as is done for reading. Table 10.4 presents three well-known examples.

**Table 10.4: Examples of universal behavior screening systems in schools.**

| System | Method | What it captures |
|---|---|---|
| **Student Risk Screening Scale (Student Risk Screening Scale, SRSS) and its extended version for internalizing problems (SRSS-IE)** | The teacher rates every student on a single list of a few items using a short scale, in one sitting for the whole class | The original version targets externalizing problems; the extended version adds items for internalizing problems |
| **Behavioral and Emotional Screening System (Behavioral and Emotional Screening System, BESS)** | Brief teacher, parent and student forms, part of the Behavior Assessment System for Children family (Reynolds & Kamphaus, 2015) | An overall risk level (normal, elevated or extremely elevated) covering both externalizing and internalizing problems |
| **Systematic Screening for Behavior Disorders (Systematic Screening for Behavior Disorders, SSBD)** | Successive gates: the teacher ranks students on externalizing and internalizing lists, then rates the highest-ranked students with more detailed checklists; in the original version, direct observation in the classroom and on the playground follows | Both kinds together; it was designed from the outset so that internalizing problems would not be missed |

These systems are useful only under two conditions. First, they must be preceded by whatever notification of families school procedures require, and their results must be kept confidential. Second, the school must have something to offer the students the screening identifies, such as support groups for social skills or anxiety management, and follow-up with the school counselor. Screening that identifies students and then leaves them without support defeats its own purpose. The screening result determines who needs a

closer look; the decision rests on the assessment that follows. At the time of writing, we know of no published standardized Arabic version of any of these three systems.

**For the teacher: the limits of your role**

You are a key source of information about the school setting, and no assessment is complete without your ratings. But do not tell a family that their son "has hyperactivity" or that their daughter "is depressed," and do not suggest medication. Diagnosis and treatment fall within the authority of the licensed health professional, and the decision on eligibility for services within that of the school's assessment team. Describe what you see, provide support in the classroom, and refer according to procedures.

## 10.8 Instrument profiles

**Broadband screening**

Captures externalizing and internalizing problems together and highlights strengths.

SDQ

**Rating scales**

Measure disorder symptoms or a broad range of problems across multiple informants.

Vanderbilt · Conners 4 · ASEBA · RCADS

**Computerized measures**

Computer-based attention tasks, used only as a supplement.

CPT

The profiles follow a single template: the idea behind the tool, its actual format, a scene from its administration, a table of its key information, how its results are interpreted, and finally its limitations. Where we have not verified a figure in the technical manual itself, we write "in the technical manual" or "in published studies," and we have not carried over figures from secondary sources; examiners should read these values from their original sources before

relying on the tool (see Chapter 2). This book does not reproduce the items of any instrument, and the drawings of forms are meant only to illustrate the layout of the page.

**For the specialist: Why do informants disagree?**

A well-known meta-analysis pooled the results of dozens of studies on informant agreement about children's behavioral and emotional problems. The mean correlation was about 0.60 between informants who see the child in the same setting, such as two parents; about 0.28 between informants who see the child in different settings, such as a parent and a teacher; and about 0.22 between children's self-ratings and the ratings of others (Achenbach et al., 1987). A later, extensive review concluded that this discrepancy is information in its own right, and that averaging it away squanders that information: it tells us where the behavior appears and with whom, and it reveals problems that only the child can see (De Los Reyes et al., 2015). The practical rule is this: for externalizing problems, teachers and parents are strong sources; for internalizing problems, especially in adolescents, the student's self-report is essential, because anxiety and sadness may go unnoticed by anyone else.

## SDQ Strengths and Difficulties Questionnaire

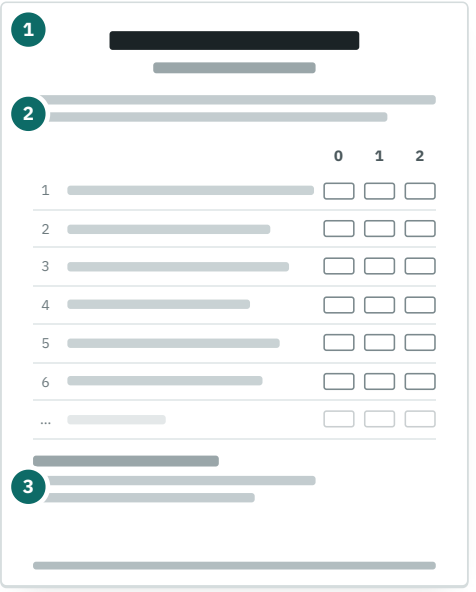
Broadband screening

### The idea

A brief tool completed by someone who knows the student. Within minutes it screens for the most prominent behavioral and emotional problems while also measuring a positive domain, which makes it suitable for screening in both schools and clinics (Goodman, 1997).

### What it looks like

A single page of 25 statements, divided equally among five subscales: emotional problems, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. For each statement, the respondent chooses one of three options: “Not true,” “Somewhat true,” or “Certainly true.” There are parent and teacher forms, a self-report form from age 11, and an optional supplement asking about the impact of the difficulties on the student’s life.




1. Details of the student and the respondent, with brief instructions: answer based on the student’s behavior over the past six months.
2. Twenty-five statements in which the five scales alternate, each followed by three boxes scored 0, 1 and 2, with reverse scoring for the positively worded statements.
3. Impact supplement: Do the difficulties affect home life, schoolwork and relationships? How long have they lasted?

**Figure 10.6:** Illustrative drawing of the layout of the Strengths and Difficulties Questionnaire form. The gray lines do not represent the actual statements; the original form can be downloaded in its various languages from the tool’s official website (https://www.sdqinfo.org).

### A scene from administration

At the start of the semester, a sixth-grade teacher completes the questionnaire for a student referred by a colleague who teaches her another subject. It takes five minutes, and she notices that she would never have paid attention to the statements about fears and physical complaints had she not been asked about them. The parent form is sent to the mother with a letter explaining its

purpose, and the student completes the self-report form with the school counselor. The counselor then gathers the three forms and compares the subscales across them.

**Table 10.5: Strengths and Difficulties Questionnaire (SDQ) at a glance.**

| | |
|---|---|
| **Purpose** | Broadband screening for behavioral and emotional problems, with a measure of prosocial behavior |
| **Ages** | Parent and teacher forms for ages 2 to 17, in age-specific versions, and a self-report form for ages 11 to 17 |
| **Edition and year** | Original version 1997, with later versions and supplements |
| **Publisher** | The tool's author, via its official website; the paper version is free for noncommercial use |
| **Qualification level** | Completed by parents, teachers and students; interpreted by a trained professional |
| **Time** | About 5 minutes |
| **Score types** | A score for each subscale (0 to 10), a total difficulties score (0 to 40) summing the four problem scales, and an impact score |
| **Cutoff scores** | Bands derived from the original norms, originally set so that about 80% of the sample fell in the normal band, about 10% in the borderline band, and about 10% in the high band |
| **Reliability** | In published studies |
| **Classification accuracy** | In published studies; varies by population and informant |
| **Year of norm collection** | The original norms are British; some countries have their own norms |

| | |
|---|---|
| **Arabic version** | Yes. An Arabic translation is available on the official website, and its validity was examined with Yemeni children aged 5 to 12 in a clinical sample (87 children) and a community sample (100 children); the questionnaire discriminated well between the two samples (Alyahri & Goodman, 2006). We are not aware of published Saudi norms for it at the time of writing |

### How are the results interpreted?

The subscales are read before the total score, because the same total can result from externalizing problems in one student and internalizing problems in another. The forms are then compared: elevated emotional problems on the student's own form alone, with low scores on the teacher and parent forms, may mean that what the student is going through is hidden from those around them, rather than that the student is exaggerating. A high prosocial behavior score is a reminder that the student has strengths alongside the difficulties, and it helps in building a support plan on those strengths.

### Limitations

It is a screening tool, so a high band signals the need for a more detailed assessment. Five statements per scale are not enough to distinguish specific disorders, such as anxiety from depression. The bands are derived from British norms, so they should be used with caution with Arab students unless local norms are available.

## Vanderbilt Vanderbilt ADHD Diagnostic Rating Scales

Rating scale

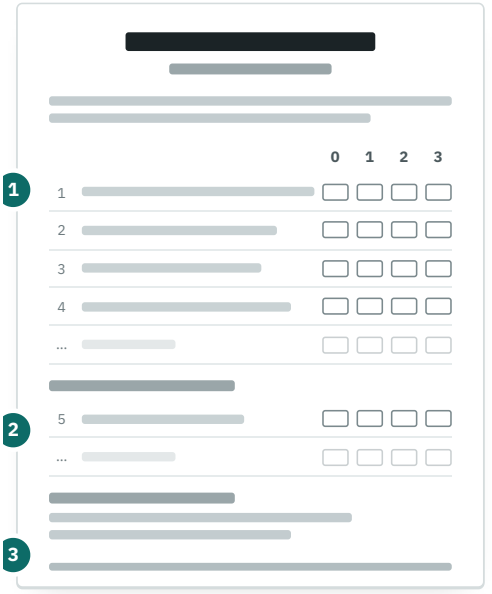
### The idea

Translating the DSM criteria into a form completed by parents and teachers, so that the physician or specialist counts symptoms from two sources rather than relying on a general impression, while also screening for the most prominent

co-occurring disorders (Wolraich et al., 2003).

### What it looks like

Two forms: one for parents and one for teachers. The first section lists the eighteen symptoms of the disorder in wording close to the DSM, each rated on a four-point scale from “Never” to “Very often.” Screening items follow for oppositional defiance, conduct disorder (on the parent form), and anxiety and depression. The final section contains performance items covering reading, mathematics, writing, relationships and following directions, each rated on a five-point scale from “Excellent” to “Problematic.”




1. The eighteen symptoms: nine for inattention, then nine for hyperactivity-impulsivity, on a scale from zero (Never) to three (Very often).
2. Screening items for co-occurring disorders, on the same scale.
3. Performance items on a five-point scale; these are used to infer Criterion D.

**Figure 10.7:** Illustrative drawing of the layout of the Vanderbilt teacher form. The gray lines do not represent the actual items.

### A scene from administration

The pediatric clinic hands Umm Faisal the parent form, along with the teacher form in a sealed envelope to deliver to the school. The teacher completes her form drawing on her written notes from the past two weeks as well as her

memory. When both forms come back, the physician counts the items rated "Often" or "Very often" on each, then looks at the performance items, and then at the screening items for co-occurring disorders (see Figure 10.8).

**Table 10.6: Vanderbilt scales at a glance.**

| | |
|---|---|
| **Purpose** | Rating ADHD symptoms and performance, screening for co-occurring disorders, and monitoring response to treatment |
| **Ages** | Primarily elementary school (about 6 to 12 years); the Saudi study included children aged 5 to 12 |
| **Edition and year** | Parent and teacher forms, with follow-up forms |
| **Publisher** | Available for clinical use as part of published guidance toolkits for pediatricians |
| **Qualification level** | Completed by parents and teachers; interpreted by a physician or qualified specialist |
| **Time** | About 10 minutes |
| **Score types** | The number of symptoms present on each list, the total score, and the number of performance items rated as problematic |
| **Cutoff scores** | Six symptoms present on a list, with impairment on at least one performance item, in line with the DSM criteria |
| **Reliability** | In published studies, including the Saudi study |
| **Classification accuracy** | In published studies |
| **Year of norm collection** | Based on symptom counts, without age norms |

| | |
|---|---|
| **Arabic version** | Yes. An Arabic version (**VADRS-A**), with parent and teacher forms, was studied with 1,812 participants in Saudi Arabia (504 parents and 1,308 teachers) rating children aged 5 to 12, under the supervision of the Saudi ADHD Society (Alqahtani et al., 2024) |

### How are the results interpreted?

Items rated "Often" or "Very often" count as symptoms present. The number of symptoms present on each list is then counted and compared with the six-symptom criterion, and impairment on the performance section is required. The two forms are then compared: do several symptoms appear in both settings? The scale translates the DSM criteria into a form, but the judgment remains with the specialist once it is combined with the developmental history, observation and the rest of the evidence. Clinical guidelines recommend collecting such scales from both the family and the teacher (Wolraich et al., 2019).

### Limitations

Counting symptoms does not rule out other explanations; an anxious or sleep-deprived student may meet the count. Because the scale does not compare the student with age norms, the question of whether symptoms are "inconsistent with developmental level" is left to the specialist's judgment, which matters for the youngest students in a class. The screening items for co-occurring disorders are few, so they serve only as a flag.

## Conners 4 Conners, Fourth Edition

Rating scale

### The idea

A detailed standardized rating scale of ADHD symptoms, co-occurring problems and their impact on functioning, which compares the student with same-age, same-sex peers using standard scores rather than relying on

symptom counts alone (Conners, 2022).

### What it looks like

Parent, teacher and self-report forms, in full and short versions, completed on paper or online. Alongside the symptom scales, it includes scales of impact on schoolwork, relationships and home life, as well as response validity indicators, such as an overly negative or overly positive impression, or inconsistency across responses.

### A scene from administration

An eighth-grade student is being assessed for a possible quiet presentation. He completes the self-report form online in a quiet room; his mathematics and Arabic teachers each complete the teacher form separately, and his mother completes the parent form. The specialist begins with the response validity indicators before looking at any score.

**Table 10.7: Conners, Fourth Edition (Conners 4) at a glance.**

| | |
|---|---|
| **Purpose** | Detailed assessment of ADHD symptoms, co-occurring problems and their impact, and monitoring of intervention |
| **Ages** | 6 to 18 years for the parent and teacher forms, with a self-report form for older youth |
| **Edition and year** | Fourth edition, 2022 |
| **Publisher** | Multi-Health Systems (MHS) |
| **Qualification level** | A professional trained in interpreting standardized rating scales, in accordance with the publisher's requirements |
| **Time** | Varies between the full and short versions |

| | |
|---|---|
| **Score types** | T-scores (mean 50, standard deviation 10) for each scale, and percentile ranks, that is, the percentage of peers in the norming sample who scored below the student; a percentile rank of 84 means the student scored higher than 84% of them |
| **Cutoff scores** | Descriptive bands for T-scores set out in the technical manual; the further a score rises above about 65, the greater the concern |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

On this scale, a higher score means more problems; a score of 70, for example, lies two standard deviations above the mean of the student's same-age, same-sex peers (see Chapter 2). The response validity indicators are read first, then the scales are compared across informants, and then the symptoms are linked to the impact scales: a high symptom score with little impact is not enough to meet Criterion D.

### Its limits

Its norms are American, and we are not aware of Arabic norms for it, so comparing an Arab student with them means comparing the student with another population. It is also a rating based on the informant's perceptions, so it is shaped by the informant's expectations and relationship with the student.

# ASEBA Achenbach System of Empirically Based Assessment: Child Behavior Checklist, Teacher's Report Form and Youth Self-Report

Broadband rating scale

## The idea

Rather than starting from a particular disorder, this system surveys a broad range of behavioral and emotional problems and groups them into syndromes derived statistically from data on thousands of children, and then into two broad dimensions: internalizing problems and externalizing problems. It offers nearly the same format to parents, teachers and students, which makes comparison across them straightforward (Achenbach & Rescorla, 2001).

## What it looks like

Three school-age forms: the **Child Behavior Checklist** (**Child Behavior Checklist, CBCL**) for parents, the **Teacher's Report Form** (**Teacher's Report Form, TRF**), and the **Youth Self-Report** (**Youth Self-Report, YSR**). Each has a section on competencies and functioning, followed by more than a hundred problem statements, each answered with 0 (not true), 1 (somewhat or sometimes true) or 2 (very true or often true), along with open-ended questions in which the respondent describes their concerns.

## A scene from administration

In Lamia's assessment (Case 4), her mother completes the Child Behavior Checklist, two teachers complete the Teacher's Report Form, and Lamia completes the Youth Self-Report with the specialist. The system is scored electronically, producing a single profile in which the eight syndromes from the three sources sit side by side.

**Table 10.8: Achenbach System, school-age forms (ASEBA) at a glance.**

| | |
|---|---|
| **Purpose** | Broadband rating of behavioral and emotional problems and competencies, and comparison across informants |
| **Ages** | 6 to 18 years for the parent and teacher forms and 11 to 18 years for the self-report; separate forms for preschoolers |
| **Edition and year** | School-age forms and manual, 2001 |
| **Publisher** | ASEBA, University of Vermont |
| **Qualification level** | Completed by parents, teachers and students; interpreted by a professional trained in standardized rating scales |
| **Time** | About 15 to 20 minutes per form |
| **Score types** | T-scores for eight syndromes (anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior), for internalizing, externalizing and total problems, and for DSM-oriented scales |
| **Cutoff scores** | Syndromes: 65 to 69 borderline range, 70 and above clinical range. Internalizing, externalizing and total problems: 60 to 63 borderline, 64 and above clinical |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual (the edition used here was published in 2001); the system also has multicultural norms specified by the publisher |

| | |
|---|---|
| **Arabic version** | We are not aware of an Arabic version standardized with published local norms at the time of writing; the publisher can be asked about licensed translations |

### How are the results interpreted?

The profile is read from general to specific: total problems, then internalizing and externalizing, then the syndromes driving them. Informants are then compared, as in the previous box. Elevated "withdrawn/depressed" and "somatic complaints" scores on the self-report and parent forms, with a smaller elevation on the teacher form, are a familiar picture of internalizing problems in adolescents. Its best-known alternative is the **Behavior Assessment System for Children**, Third Edition (**BASC-3**), which offers similar parent, teacher and student forms and is linked to the behavioral and emotional screening system mentioned in 10.7 (Reynolds & Kamphaus, 2015).

### Its limits

The syndromes are descriptive dimensions, so an elevated "attention problems" score calls for an ADHD assessment before any judgment is made. Its length is a burden for some families, and it may be affected by the respondent's reading level. Its norms are not local, so its scores should be read with caution for Arab students.

## RCADS Revised Child Anxiety and Depression Scale

Internalizing problems

### The idea

Broadband scales flag the presence of internalizing problems, but they do not distinguish precisely among types of anxiety and depression. That is what this scale offers: subscales corresponding to the most common anxiety and

depressive disorders in the DSM, completed by the student, with an accompanying parent form (Chorpita et al., 2000).

### What it looks like

Forty-seven short statements that the student answers on a four-point scale: "Never," "Sometimes," "Often" and "Always." They are distributed across six subscales: separation anxiety, social anxiety, generalized anxiety, panic disorder, obsessive-compulsive disorder, and major depression, in addition to a total anxiety score and a total internalizing score. Other scales serve the same purpose, including the Screen for Child Anxiety Related Emotional Disorders (**SCARED**) and the Children's Depression Inventory (**CDI-2**).

### A scene from administration

The specialist sits with an eighth-grade student and explains that there are no right or wrong answers, and that the answers describe how she usually feels. The student reads the statements herself and pauses at one containing a word she does not understand; the specialist explains the word without hinting at an answer. After scoring, the specialist asks about the elevated statements: "You said you often can't sleep. Tell me about that."

**Table 10.9: Revised Child Anxiety and Depression Scale (RCADS) at a glance.**

| | |
|---|---|
| **Purpose** | Rating symptoms of the most common anxiety and depressive disorders, screening within referred samples, and monitoring intervention |
| **Ages** | Students in roughly Grades 3 through 12 |
| **Edition and year** | Revised version, 2000, with a parent form and a later short version |
| **Publisher** | Available free of charge from its developers for noncommercial use |

| | |
|---|---|
| **Qualification level** | Completed by the student and parents; interpreted by a trained professional |
| **Time** | About 10 to 15 minutes |
| **Score types** | Raw scores, and T-scores by grade and sex for the subscales and total scales |
| **Cutoff scores** | T-scores of 65 and 70 are commonly used as the thresholds for the borderline and clinical ranges, respectively, following the developers' guidance |
| **Reliability** | In published studies |
| **Classification accuracy** | In published studies |
| **Year of norm collection** | In the developers' guidance; the original norms are American |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The subscales are read against one another: an elevation in depression alone differs from elevations in both generalized and social anxiety. They are then compared with the parent form. The most important thing the specialist does after scoring is to return to the elevated statements and ask about them; the scale guides the interview but does not replace it.

### Its limits

It depends on the student's reading ability and on the student's capacity to understand and express their feelings, and some students may underreport their symptoms for fear of stigma. It does not ask about everything a diagnosis

requires, such as duration, impact and the student's safety. Its norms are not local.

## CPT Computerized continuous performance tests

Supplement only

### The idea

Measuring sustained attention and impulsivity through a deliberately tedious computerized task that records responses to the millisecond, rather than relying on others' ratings.

### What it looks like

In continuous performance tests (**Continuous Performance Tests, CPT**), the student sits at a computer for about a quarter of an hour and is asked to press a button when a particular letter or shape appears and to withhold the response when anything else appears, or the reverse, depending on the test.

### A scene from administration

The student sits in a quiet room, practices the task for one minute, and then the test begins. The examiner leaves the room or sits behind the student, as the test instructions specify, and records any fidgeting, talking or fatigue observed during the task.

**Table 10.10: Computerized continuous performance tests (CPT) at a glance.**

| | |
|---|---|
| **Purpose** | A supplementary measure of sustained attention and impulsivity; a diagnosis should not rest on it alone |
| **Ages** | Vary by test; most start at school age |
| **Edition and year** | A family of commercial tests, each with its own edition |
| **Publisher** | Varies by test |

| | |
|---|---|
| **Qualification level** | A trained examiner administers it and a qualified professional interprets it, in accordance with the publisher's requirements |
| **Time** | Usually about 15 minutes |
| **Score types** | Standard scores or T-scores for omission errors, commission errors, reaction time and its variability |
| **Cutoff scores** | According to each test's manual |
| **Reliability** | In each test's technical manual |
| **Classification accuracy** | Limited, according to systematic reviews (see Limitations) |
| **Year of norm collection** | In each test's technical manual |
| **Arabic version** | The task is largely nonverbal, but we are not aware of published standardized Arabic norms for any of these tests at the time of writing |

### How are the results interpreted?

The program records omission errors, that is, failing to press when the target appears, which are associated with inattention; commission errors, that is, pressing when a nontarget appears, which are associated with impulsivity; and the variability of reaction time. The result is read as a hypothesis awaiting confirmation from other evidence.

⚠ **Limitations: Why are they not enough on their own?**

These tests appear objective because they are computerized, but systematic reviews have found that they do not discriminate accurately enough between children with and without the disorder, and that they do not reflect the child's behavior in the classroom (Hall et al., 2016). A child may perform well during a quiet quarter of an hour with an examiner yet struggle through a full school day. For this reason, they are used as a supplement to rating scales and observation.

## 10.9 When is a judgment made? Four cases

### Case 1: Faisal, 7 years 8 months

**Combined presentation, moderate severity**

**Background:** He is the student in the first scene. His teacher documented his behavior for two weeks and tried seating him near her desk, breaking tasks into smaller parts and a reinforcement system; he improved slightly, then slipped back. His age is roughly in the middle of his class. His mother says he "hasn't sat still since he started walking," and that his kindergarten teachers complained about the same thing.

| | Mother's rating | Teacher's rating |
|---|---|---|
| Inattention symptoms present | 7 of 9 | 8 of 9 |
| Hyperactivity symptoms present | 8 of 9 | 7 of 9 |
| Impaired performance | In relationships with siblings | In reading and relationships |
| Anxiety and depression items | Low | Low |

**Figure 10.8:** Hypothetical summary of the mother's and teacher's Vanderbilt ratings. The picture is consistent across home and school and exceeds the six-symptom criterion on both lists.

The team added that hearing and vision screening was normal, his sleep was regular, and an achievement test showed a mild weakness in reading consistent with his inattention rather than with a difficulty in phonological awareness. Direct classroom observation recorded him out of his seat in 60% of intervals using partial-interval recording (see Chapter 4), compared with 15% for two classmates observed for comparison.

| A ✓ | B ✓ | C ✓ | D ✓ | E ✓ |
|---|---|---|---|---|
| Both lists 6+ | Since kindergarten | Home and school | Clear impact | No better explanation |

**Decision:** The physician diagnosed Faisal with attention-deficit/hyperactivity disorder, combined presentation, of moderate severity: his symptom count exceeds the required threshold by one or two symptoms on each list, and his impairment is clear at home and at school without reaching a level that disrupts most aspects of his day (see Table 10.1). A classroom support plan, behavior management training for the family and monitoring of his reading skills are recommended.

**What evidence could change the decision?** If his reading weakness persisted after his attention in class improved, an assessment for a co-occurring learning disability would be warranted. And if kindergarten records showed that what his mother described had not been noticed there, Criterion B would need firmer confirmation before the diagnosis is settled.

## Case 2: Haya, 11 years

Anxiety explains the picture

**Background:** She is the student in the second scene. Her teacher requested an assessment for a possible predominantly inattentive presentation. But the specialist asked about timing and learned that Haya had performed well until the end of fifth grade, that her father had gone abroad for work the previous summer, and that every evening she asks her mother whether he is all right.

**Table 10.11: Evidence in Haya's case.**

| Source | Result | What it means |
|---|---|---|
| **Teacher's rating on an ADHD symptom scale** | 6 inattention symptoms | Appears, on the surface, to be met at school |
| **Mother's rating** | One symptom (she has been forgetting some of what she is asked to do for several months); she completes her homework at home with little help, and her mother recalls that she was attentive and organized before this year | Several symptoms do not appear at home, and there is no clear impact there |

| Source | Result | What it means |
|---|---|---|
| **Records from previous grades and her teachers' notes** | Good performance, with no notes about daydreaming or forgetfulness | No history of symptoms before this year |
| — | Emotional problems elevated on the mother's and teacher's forms | A clear indicator of anxiety |
| **Interview with Haya** | She says she is "thinking about Dad" during lessons and cannot sleep well | Her daydreaming stems from preoccupation with worry |

| A ? | B ✓ | C ✗ | D ✓ | E ✗ |
|---|---|---|---|---|
| From a single source | Only formally: no history before this year | Not present in two settings | Academic impact | Anxiety is a better explanation |

**Decision:** Haya is not diagnosed with attention-deficit/hyperactivity disorder. Criterion B is met in form, since her symptoms appeared before age twelve, but according to her mother and the records she has no history of them before this year. Two criteria rule out the diagnosis: the symptoms do not appear in two settings (C), and anxiety explains them better (E). She is referred to the school counselor and a psychologist for assessment of her anxiety and for support, and the family is informed.

**The lesson**

The question "When did this start?" reveals an acute onset tied to an event. An acute onset following sound functioning points toward another explanation, even when Criterion B is formally met. The criterion sets a maximum age of onset; it does not say that everything appearing before that age is developmental.

**What evidence could change the decision?** If her daydreaming and forgetfulness persisted after her anxiety eased with support, or if a more in-depth interview with her mother revealed symptoms at home that predated her father's departure, the possibility of attention-deficit/hyperactivity disorder would need to be reconsidered.

## Case 3: Saud, 6 years 1 month Not now

**Background:** His first-grade teacher referred him two months into the school year because he does not stay seated, cries when asked to write for a long time, and interrupts frequently. When the specialist reviewed his file, she found that he was among the youngest students in his class, nearly a year younger than the oldest, and that he had been admitted among the children who were less than ninety days short of turning six.

**Table 10.12: Evidence in Saud's case.**

| Source | Result | What it means |
|---|---|---|
| **Teacher's rating** | 6 hyperactivity symptoms | Compared with a class older than he is |

| Source | Result | What it means |
|---|---|---|
| **Mother's rating** | One symptom (very active during play); she describes him as "as lively as his brothers were at his age" and sees nothing at home that worries her | Several symptoms do not appear at home |
| **His former kindergarten teacher** | Noticed no particular problem | No clear history |
| **Comparative observation** | His behavior is close to that of the two youngest students in the class, and above the overall average | Consistent with his age |

**Decision:** Saud is not diagnosed at this time. His symptoms appear in only one setting and are consistent with his age when he is compared with the peers closest to him in age, which means the symptoms are not "inconsistent with developmental level" as Criterion A requires. Age-appropriate classroom accommodations are recommended, such as short movement breaks and shorter tasks, together with monitoring through structured observation until the end of the semester. If the symptoms persist or appear at home, he will be reassessed.

> **What evidence could change the decision?** If structured observation showed his activity and impulsivity remaining clearly above those of the peers closest to him in age through the end of the semester, despite the accommodations, or if his mother began to notice symptoms at home that disrupt his day, the team would move to a comprehensive evaluation.

## Case 4: Lamia, 14 years 3 months, eighth grade

**Eligible under the emotional disturbance category**

**Background:** Lamia was among the top students in elementary school. From the beginning of seventh grade, about fourteen months ago, she began to change: she is often absent because of a "stomachache" on school-day mornings, she has dropped the Qur'an memorization circle she used to love, she does not participate in lessons, and she spends recess alone. No one referred her, since she is "quiet and causes no trouble." But the universal behavior screening the school conducted in October placed her in the at-risk range for internalizing problems. She joined a ten-week support group with the school counselor, and the counselor contacted the family, but she improved only slightly. She was then referred for evaluation with her guardian's written consent.

**Table 10.13: Evidence in Lamia's case.**

| Source | Result | What it means |
|---|---|---|
| **School records** | 17 days absent in the current semester, compared with 3 days in sixth grade; her grade average fell from 94% to 72% | Clear educational impact and a change from her previous functioning |
| **Child Behavior Checklist (mother)** | Internalizing problems 72, externalizing 52; anxious/depressed 74, withdrawn/depressed 73, somatic complaints 70 | Internalizing problems in the clinical range; no externalizing problems |
| **Teacher's Report Form (two teachers)** | Internalizing problems 68 and 70 | They see the problems too, though to a lesser degree |

| Source | Result | What it means |
|---|---|---|
| **Youth Self-Report and the Revised Child Anxiety and Depression Scale** | Internalizing problems 73; major depression 76, generalized anxiety 69, social anxiety 66 | The highest scores come from Lamia herself |
| **The psychologist's interview with Lamia and her mother** | Sad mood most days for more than a year, loss of enjoyment, insomnia, fatigue, difficulty concentrating, and a feeling that she is "useless." As procedures require, the psychologist asked her about thoughts of self-harm, which she denied, and a monitoring plan was agreed with the family | A clear depressive picture with anxiety |
| **Pediatrician** | Physical examination and thyroid tests normal; no organic cause for the abdominal pain | Medical cause ruled out |
| **Hearing and vision** | Both normal | No sensory explanation |
| **Inattention symptom scale (teacher) and elementary school records** | 4 inattention symptoms at present, and no notes about them in elementary school | Recent concentration difficulties, explained by depression |
| **Rule violations** | No recorded violations and no behavior problems | No grounds for an exclusion based on "social maladjustment" |

After interviewing her, a child and adolescent psychiatrist diagnosed her with major depressive disorder with prominent anxiety symptoms and began a treatment plan. The school's assessment and diagnosis team then met, with her guardian present, to answer a different question: Does Lamia need special education services? The team applied the elements of the definition of emotional disturbance (10.5):

| Duration ✓ | Degree ✓ | Impact ✓ | Characteristics ✓ | Exclusion ✓ |
|---|---|---|---|---|
| More than a year | Clinical range from three sources | Absences and falling grades | The fourth and fifth | Not social maladjustment |

**Decision:** Lamia is eligible under the emotional disturbance category, which in Saudi regulations falls within the category of "behavioral and emotional disorders." Because the Special Education Organizational Guide does not spell out criteria for this category, the team documented its decision using the elements above and the psychiatrist's certified report. Her plan included a temporary reduction in homework with flexible deadlines, a plan for a gradual return to regular attendance, brief regular meetings with the school counselor, coordination with the treating provider, and a readministration of the anxiety and depression scale after eight weeks to monitor progress.

**The lesson**

Without universal screening, "quiet" Lamia might have gone without support until she dropped out of school. Without her self-report, her problems would have seemed milder than they are. Internalizing problems in adolescents need someone to look for them, and they require asking the student directly. Note, too, that a clinical diagnosis and eligibility for an educational category are two different decisions, each answering its own question and each made by those with the authority to make it.

**What evidence could change the decision?** If her attendance and grades improved steadily with treatment, and her anxiety and depression scale scores returned to the normal range on readministration, the team would review her need for special education services at the time of reevaluation, while maintaining counseling support.

## 10.10 Differential diagnosis and co-occurrence

### Similar conditions

Inattention, hyperactivity and challenging behavior are symptoms that appear in many conditions. These are the main questions to ask before reaching a judgment:

**Table 10.14: Differential diagnosis of ADHD and behavioral and emotional disorders.**

| Possible cause | Similarity | Distinctive features |
|---|---|---|
| **Relative age and immaturity** | Restlessness and weak self-control | Behavior is consistent with that of the peers closest in age, as in Saud's case |
| **Anxiety** | Daydreaming, fidgeting and difficulty concentrating | Daydreaming due to preoccupation with worries, and an onset linked to an event, as in Haya's case |
| **Depression** | Poor concentration and unfinished tasks, and irritability that may be read as defiance | Sadness or irritability, loss of interest in things previously enjoyed, and a clear change from previous functioning, as in Lamia's case |

| Possible cause | Similarity | Distinctive features |
|---|---|---|
| **Oppositional defiant disorder** | Refusing instructions and not completing tasks | The refusal is directed at the person making the request rather than at the task, and it appears even with short, easy tasks. It frequently co-occurs with ADHD |
| **—** | Temper outbursts and irritability | Severe, recurrent outbursts, persistently irritable mood between outbursts, and onset before age ten |
| **Learning disability** | Avoiding tasks and not completing them | Inattention appears only in the difficult subject, while the student attends well in others. The two may co-occur |
| **Hearing or visual impairment** | Does not respond or follow along | Improves when the sensory impairment is treated, which is why sensory screening precedes any diagnostic tool |
| **Insufficient sleep or sleep problems** | Fatigue, daydreaming and quick temper | A history of late or interrupted sleep, or snoring |
| **Trauma and family stress** | Hypervigilance, emotional reactivity and poor concentration | A history of difficult events, after which behavior may change noticeably |
| **Giftedness with an unsuitable curriculum** | Boredom, daydreaming and interrupting | Usually improves with enrichment and tasks matched to the student's level. But the disappearance of the behavior during a stimulating activity does not rule out the disorder (see the box in 10.1), and the two may co-occur |
| **Autism spectrum disorder** | Social difficulties and impulsivity | The social difficulty lies in reading social cues, not in impulsivity alone; both diagnoses may be given together |

## Co-occurring condition

The conditions in Table 10.14 are not always alternatives; many of them co-occur in the same student. Oppositional defiance frequently accompanies ADHD, anxiety and depression often occur together in adolescents, and published reviews report that a substantial proportion of students with learning disabilities also have ADHD, with estimates varying according to samples and criteria (DuPaul et al., 2013). Knowing that two disorders can co-occur is only a first step; each combination then calls for an assessment procedure that keeps one condition from masking or inflating the other:

- **With learning disabilities:** Achievement is measured with standardized tools in every ADHD assessment, and the team checks whether inattention appears across all subjects or only in the area of weakness. If treatment for attention begins, academic skills are measured again afterward.
- **With oppositional defiance:** Ask about symptoms in situations that involve no demands and no authority, such as free play or watching a program; inattention that appears there cannot be explained by defiance.
- **With anxiety and depression:** Establish when each began, and ask the student directly using a measure of internalizing problems, because poor concentration is a shared symptom that will be counted twice unless care is taken.
- **With autism spectrum disorder:** Read the attention and impulsivity items on rating scales with caution, because some may be elevated for reasons related to autism itself, such as not responding to one's name or absorption in a special interest; rely on observation across varied situations (see Chapter 9).
- **With giftedness:** Compare behavior in an ordinary lesson with a lesson that offers genuine challenge, gather ratings from two settings, and ask about developmental history before school (see Chapter 13).

Precise co-occurrence rates vary widely across studies depending on the source of the sample (clinical or community), the instrument and the criteria. For that reason we do not present them here as definitive figures, and we offer instead a practical rule: whenever one of these disorders is diagnosed, ask about the others.

## 10.11 In the Saudi and Arab context

**In the Saudi context: Who diagnoses, and who decides eligibility?**

The Special Education Organizational Guide lists "attention-deficit and hyperactivity disorders" and "behavioral and emotional disorders" as two separate special education categories (Ministry of Education, 1436 AH [Arabic]). The earlier Disability Welfare Law had counted "behavioral and emotional disorders" among disabilities (Disability Welfare Law, 1421 AH [Arabic]); it was later replaced by the Law on the Rights of Persons with Disabilities, which defines a person with a disability as someone with a "long-term disorder or impairment" in functions that include "psychological" functions (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic]). Table 10.15 compares this regulatory framework with the two diagnostic manuals and with U.S. law.

**Table 10.15: Attention-deficit/hyperactivity disorder and behavioral and emotional disorders in the diagnostic manuals, US law and Saudi regulations.**

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Category** | Separate diagnoses: attention-deficit/hyperactivity disorder, oppositional defiant disorder, conduct disorder, anxiety, depression and others | No separate category for ADHD, which is usually served under "other health impairment"; the "emotional disturbance" category covers emotional and behavioral conditions | Two separate categories: "attention-deficit and hyperactivity disorders" and "behavioral and emotional disorders" (Organizational Guide) |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Definition and criteria** | Detailed criteria specifying the number of symptoms, duration and age in DSM-5-TR; a description of features without a set number in ICD-11 | Five characteristics, over a long period of time and to a marked degree, adversely affecting educational performance; includes schizophrenia and excludes social maladjustment alone | For eligibility: "Hyperactivity, attention deficit or both, requiring intervention according to an accredited report." We found no detailed criteria in the Guide for emotional disorders, nor any counterpart to the exclusion of social maladjustment |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Who decides** | A qualified clinician, according to each country's licensing rules | A team of qualified professionals together with the parents | Assessment and diagnosis team: the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the school counselor, specialists relevant to the category, a physician where available, and the parent or guardian where possible. An "accredited report" is required, without naming who issues it |

| Element | DSM-5-TR and ICD-11 | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Instruments** | A comprehensive clinical evaluation, supported by rating scales from multiple sources | Multi-instrument evaluation; no single procedure may be relied upon | "Tools and measures approved by the Ministry of Education," "not relying solely on the results of a single scale," administered and interpreted "by qualified specialists." We found no published list of these instruments |
| **Consent and reevaluation** | Not applicable | Parental consent, and reevaluation at least every three years | Written consent from the guardian after notification, with no response within 14 days treated as consent; rediagnosis no earlier than 6 months and no later than 18 months after the initial diagnosis |

**Who issues the diagnosis?** We found no published text that specifies precisely who is authorized to diagnose attention-deficit/hyperactivity disorder or emotional disorders in Saudi Arabia. What we did find is that the Organizational Guide requires an "accredited report" for eligibility without naming who issues it (Ministry of Education, 1436 AH [Arabic]),

and that the professional licensing examination for psychologists assesses their readiness to conduct psychological assessment "to identify/diagnose" psychological disorders (Saudi Commission for Health Specialties, 2024 [Arabic]). This chapter therefore uses a single formulation: **the clinical diagnosis is made by a health professional licensed by the Saudi Commission for Health Specialties, within the scope of that license**, such as a pediatrician, a psychiatrist or a clinical psychologist, **and the school's assessment team decides eligibility for services**. Readers should consult any subsequent regulations issued by the Commission and the Ministry of Education.

**Prevalence:** A meta-analysis of fourteen Saudi studies estimated the prevalence of attention-deficit/hyperactivity disorder at about 12.4%, with a confidence interval of 5.4% to 26% (Aljadani et al., 2023). At the time of writing, we found no published, peer-reviewed national estimate of the prevalence of emotional disorders among school students in Saudi Arabia.

**A question for reflection:** If Lamia's case came before a team applying the text of the Organizational Guide to the letter, what evidence would it rely on for the eligibility decision, given that the text does not spell out the category's criteria? And what is the risk of leaving the definition of the category to the judgment of each team?

## Available Arabic instruments

Table 10.16 brings together the Arabic versions of this chapter's instruments that we found published at the time of writing. It makes no claim to completeness, and it is no substitute for verifying the licensing and norms of any instrument before using it in a decision.

**Table 10.16: Published Arabic versions of the instruments in this chapter.**

| Instrument | Study | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| — | Alyahri & Goodman (2006) | Yemen; children aged 5 to 12, clinical sample (87) and community sample (100) | Evidence of the Arabic version's validity in distinguishing between the two samples | Small sample from a single country; no published Saudi norms |
| **Vanderbilt scales** | Alqahtani et al. (2024) | Saudi Arabia; 1,812 participants (504 parents and 1,308 teachers) rating children aged 5 to 12 | Psychometric properties of an Arabic version, with parent and teacher forms, in a Saudi sample | For the elementary grades; consult the study for the scores it reports and its limitations |
| **Conners scales** | At the time of writing, we know of no standardized Arabic version published in a peer-reviewed journal | | | |
| **Achenbach System and Behavior Assessment System for Children** | At the time of writing, we know of no published standardized Arabic version with local norms | | | |
| **Anxiety and depression scales** | At the time of writing, we know of no published standardized Arabic version of the Revised Child Anxiety and Depression Scale | | | |

| Instrument | Study | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| **Universal behavior screening systems** | At the time of writing we know of no published standardized Arabic version | | | |

The gap here is wider for internalizing problems than for attention-deficit/hyperactivity disorder. A Saudi elementary school student can be rated on an Arabic rating scale studied in a large Saudi sample, but for the anxious or depressed adolescent we know of no Arabic self-report instrument standardized with local norms. Until this gap is closed, the report should state the source of each instrument and its norms, and the student's own report and interview should carry more weight than the standard score alone.

## Conclusion

The central argument we have sought to set out in this chapter is that judgment in these categories goes beyond the mere presence of a behavior: it turns on the number of symptoms, when they first appeared, the range of settings in which they show themselves and their effect on functioning, and on the absence of a more likely explanation. It follows that assessment begins with sensory checks and universal screening, rests on rating scales from multiple informants, compares the student with the peers closest in age, treats the student's own report as essential for internalizing problems, and keeps computerized tests in a merely supplementary role. It also follows that the clinical diagnosis, made by a licensed health professional, must be kept distinct from the eligibility decision, made by the assessment team.

Some of the chapter's questions, however, remain contested or unsettled. The exclusion of social maladjustment alone from the emotional disturbance category in US law is a matter of debate; the ICD-11 does not require a set

number of symptoms as the DSM does; and Saudi regulations treat emotional disturbance as a separate category without published detailed criteria, leaving it to the team to state the criteria on which it based its judgment.

## 10.12 Summary, key terms and activities

### Chapter summary

- Attention-deficit/hyperactivity disorder is diagnosed on the basis of at least six symptoms in one or both lists, several of which appeared before age 12; several symptoms are present in two settings, they interfere with functioning, and no other disorder better explains them. The diagnosis specifies the presentation and the severity, and the ICD-11 does not require a set number of symptoms.
- A learning disability is not among the exclusionary criteria, and autism has not been an exclusion since the fifth edition of the DSM.
- The youngest students in a class are more likely to be misdiagnosed, according to studies from more than one country, so the student should be compared with the peers closest in age. The quiet presentation and internalizing problems are often overlooked, particularly in girls and adolescents.
- Oppositional defiant disorder, conduct disorder, disruptive mood dysregulation disorder, anxiety, depression and posttraumatic stress disorder each have distinct criteria, and in children and adolescents they take particular forms, such as irritable rather than sad mood in depression.
- The emotional disturbance category in US law requires a long period of time, a marked degree and an educational impact; it includes schizophrenia and excludes social maladjustment alone, an

exclusion that remains a matter of debate. Saudi regulations treat the category as separate, without published detailed criteria.

- Assessment begins with sensory checks and universal screening and rests on rating scales from multiple informants; the student's own report is essential for internalizing problems. Computerized tests serve only as a supplement.
- Clinical diagnosis falls within the remit of a licensed health professional, and the eligibility decision within that of the assessment team.

## Glossary of terms

| | |
|---|---|
| **Attention-Deficit/Hyperactivity Disorder (ADHD)** | A neurodevelopmental disorder marked by a persistent pattern of inattention, hyperactivity-impulsivity, or both. |
| **Executive functions** | Capacities that direct behavior toward a goal, such as planning, working memory and self-control. |
| **Presentation** | The current picture of symptoms: predominantly inattentive presentation, predominantly hyperactive-impulsive presentation, or combined presentation. |
| **In partial remission** | A specifier used when full criteria were previously met but have not been fully met over the past six months, while impairment in functioning persists. |
| **Relative age effect** | The greater likelihood that the youngest students in a class will be diagnosed because they are |

| | |
|---|---|
| | compared with older classmates. |
| **Externalizing problems** | Problems directed toward others, such as aggression, defiance and rule-breaking. |
| **Internalizing problems** | Problems directed inward, such as anxiety, sadness, withdrawal and somatic complaints. |
| **Oppositional defiant disorder** | A persistent pattern of angry or irritable mood, argumentativeness, defiance and vindictiveness. |
| **Conduct disorder** | A repetitive and persistent pattern of violating the rights of others or major age-appropriate social norms or rules. |
| **Limited prosocial emotions** | A specifier for conduct disorder describing lack of remorse, lack of empathy, unconcern about performance and shallow affect. |
| **Disruptive mood dysregulation disorder** | Severe, recurrent temper outbursts with persistently irritable mood between them, beginning before age 10. |
| **Emotional disturbance** | An educational category in US law for students whose emotional or behavioral characteristics impede their learning over a long period of time and to a marked degree. |
| **Social maladjustment** | A term left undefined in US regulations; students so described are excluded from the emotional disturbance category unless they are also found to have an emotional disturbance. |

| | |
|---|---|
| **Universal behavior screening** | Periodically rating every student in a class with a brief instrument to find those who need support or evaluation, including students who have not been referred. |
| **Informant discrepancy** | Differences among parents', teachers' and students' ratings of the same problems, read as information about context rather than measurement error. |
| **Continuous performance test** | A computerized task that measures sustained attention and impulsivity through responses to a sequence of stimuli. |
| **Functional behavioral assessment** Functional Behavioral Assessment (FBA) | Gathering information on what precedes and follows a behavior in order to determine the function it serves for the student. |

T-score, sensitivity and specificity, and cut-off score are terms defined in Chapter 2, and partial-interval recording is defined in Chapter 4.

## Review questions

1. A fifth grader's teacher rated seven inattention symptoms and two hyperactivity symptoms, and his mother rated three inattention symptoms, along with marked difficulty completing homework. Is Criterion C met? What is the likely presentation? And what information is needed to determine severity? **(Objective 1)**
2. Identify two differences between the DSM and the ICD-11 covered in this chapter, and explain how each would affect a report you receive about a student. **(Objective 1)**
3. In light of the first-grade admission rule in Saudi Arabia, how large might the age gap be between the youngest and the oldest student in a single class?

Propose a procedure in your school that would reduce misdiagnosis arising from this. **(Objective 2)**

4. An eighth-grade girl's grades have slipped; she forgets her homework and seems distracted. What questions would you ask before considering attention-deficit/hyperactivity disorder? And when does a late appearance of difficulties not rule out the diagnosis? **(Objectives 2 and 6)**
5. Classify the following as externalizing or internalizing problems: refusing instructions, stomachaches on school-day mornings, frequent fighting, silence at school while speaking at home, and running away from school. Then explain the difference between a student who has severe outbursts three times a week with irritable mood between them and a student who argues with and defies his teachers without severe outbursts, and why the two diagnoses are not given together. **(Objective 3)**
6. A secondary school student skips school and deliberately breaks rules with his friends, and shows no clear emotional symptoms. How does the definition of emotional disturbance in US law treat his case? What is your critique of this exclusion? And what do Saudi regulations say? **(Objective 4)**
7. A student performed within normal limits on a computerized continuous performance test, while his family's and teachers' ratings are very high. How would you interpret this? **(Objective 5)**
8. A mother rated her daughter's internalizing problems in the normal range on the Child Behavior Checklist, while the student rated herself in the clinical range on the self-report. How would you interpret this? Why would you not average the two scores? And which instrument would you add next? **(Objective 5)**
9. Why do hearing and vision checks and screening come at the second step of the decision pathway, before any diagnostic instrument? Name three explanations ruled out at the third step, with a question or instrument for each. **(Objective 6)**
10. Both Haya and Lamia show inattention and falling grades. Why did the first case end in a referral for anxiety assessment, and the second in eligibility

under the emotional disturbance category? Identify the evidence that made the difference. (Objective 7)

## Case for application

### Walid, 15 years, tenth grade

**The story:** The vice principal for student affairs referred him after two fights this semester and repeated complaints from his teachers: he argues, refuses instructions, loses his temper quickly and raises his voice. His family moved to the city a year ago. With his guardian's consent, the team gathered the following data over one month.

**Table 10.17: Data in Walid's case.**

| Source | Result |
|---|---|
| **Father** | Severe temper outbursts at home three or four times a week, sometimes breaking his belongings, "since he was about eight"; between outbursts he is "upset and irritable most of the day" |
| **Mother** | Over the past four months he has drifted away from his old friends and sleeps a great deal; he once told her, "Nobody understands me" |
| **Strengths and Difficulties Questionnaire (teacher)** | Conduct problems in the high range; hyperactivity-inattention in the borderline range |
| **Strengths and Difficulties Questionnaire (self-report)** | Emotional problems in the high range; conduct problems in the borderline range |

| Source | Result |
|---|---|
| **Inattention symptom scale (mathematics teacher)** | Five inattention symptoms and three hyperactivity symptoms |
| **Third-grade record** | A single note: “Doesn’t finish his work; needs reminding” |
| **Behavior records** | No theft, no running away from school, no cruelty to animals; regular attendance |
| **Hearing and vision** | Both normal on this year’s school screening |

**Tasks:** (1) Which disorders do you consider, and on which criterion does each depend? Why are oppositional defiant disorder and disruptive mood dysregulation disorder not diagnosed together? (2) What evidence is missing before any judgment can be made, whom would you ask, and with which instrument? (3) If the emotional picture is confirmed, would Walid be eligible for the emotional disturbance category under the US definition, or would he be considered “socially maladjusted”? Give your reasons, then explain what Saudi regulations say. **(Objectives 3, 4 and 7)**

## Learning activities

Every activity carried out with students is conducted with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a specialist; its results are not used in any decision concerning the student.

1. **Comparative observation:** Observe a student referred for hyperactivity, together with two students of the same sex who are closest to him in age, for ten minutes using partial-interval recording, and compare the percentages. *Desk-based alternative:* Design the observation form, write operational

definitions for two behaviors, and try it out on a published video of a classroom. **(Objectives 2 and 6)**

2. **Classroom screening:** With the school administration's permission, and after informing families according to its procedures, have the teacher form of the Strengths and Difficulties Questionnaire completed for an entire class; then identify the students with elevated emotional problems who were not among those you had been watching. Pass anything that calls for follow-up to the school counselor through the official channel, without informing any family of a result or discussing names. *Desk-based alternative:* Write ten brief hypothetical student profiles, rate them with the questionnaire, and compare whom the screening "caught" with whom you would have referred on impression alone. **(Objective 5)**
3. **A map of explanations:** List ten different reasons a student might not pay attention in class, and for each one specify a question or instrument that would reveal it and where it falls in the decision pathway. **(Objective 6)**
4. **Comparing systems:** Apply the columns of Table 10.15 to the cases of Lamia and Walid, and write for each: What is the decision under each system? What evidence does it rest on? And what does the Saudi text leave unresolved? **(Objective 4)**

## Chapter references

Sabq. (2026, March 8). [News report on first-grade admission criteria for the 1448 AH school year, citing the Ministry of Education] [In Arabic]. https://sabq.org/article/gXkENsW

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Disability Welfare Law, Royal Decree No. (M/37) (23/9/1421 AH) [In Arabic]. https://www.un.org/development/desa/disabilities/wp-content/uploads/sites/15/2019/11/Saudi-Arabic_Disability-Law.pdf

Saudi Commission for Health Specialties. (2024). *Saudi Psychology Licensure Examination (SPSLE): Applicant guide*. https://scfhs.org.sa/sites/default/files/2024-03/SPSLE%20Applicant%20Guide%202024.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (2025, March 20). [News release on first-grade admission for the 1447 AH school year] [In Arabic]. https://www.moe.gov.sa/ar/mediacenter/MOEnews/Pages/news1_20032025.aspx

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Aljadani, A. H., Alshammari, T. S., Sadaqir, R. I., Alrashede, N. O. E., Aldajani, B. M., Almehmadi, S. A., Altuhayni, A. S., & Abouzed, M. A. (2023). Prevalence and risk factors of attention deficit–hyperactivity disorder in the Saudi population: A systematic review and meta-analysis. *Saudi Journal of Medicine & Medical Sciences, 11*(2), 126–134. https://doi.org/10.4103/sjmms.sjmms_528_22

Alqahtani, M. M. J., Al Saud, N. M., Alsharef, N. M., Alsalhi, S. M., Al-Hifthy, E. H., AlHadi, A. N., Ad-Dab'bagh, Y., Alenazi, F. A., Alotaibi, B. M., Alsaeed, S. M., Arnout, B. A., ALQasem, L., Alhossein, A., Alqahtani, Y. J., AlGhamdi, S. A., Alrahili, N., Varnham, J., & Asiri, S. A. (2024). Psychometric properties of the Arabic Vanderbilt Children's ADHD Diagnostic Rating Scale (VADRS-A) in a Saudi population sample. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 12*(1), 72–83. https://doi.org/10.2478/sjcapp-2024-0008

Alyahri, A., & Goodman, R. (2006). Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment. *Eastern Mediterranean Health Journal, 12*(Suppl. 2), S138–S146.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy, 38*(8), 835–855.

Conners, C. K. (2022). *Conners 4th Edition*. Multi-Health Systems.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900. https://doi.org/10.1037/a0038498

DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications for DSM-5 and assessment and treatment. *Journal of Learning Disabilities, 46*(1), 43–51.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581–586.

Hall, C. L., Valentine, A. Z., Groom, M. J., Walker, G. M., Sayal, K., Daley, D., & Hollis, C. (2016). The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: A systematic review. *European Child & Adolescent Psychiatry, 25*(7), 677–699. https://doi.org/10.1007/s00787-015-0798-x

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), 20 U.S.C. § 1400 et seq. (2004).

Layton, T. J., Barnett, M. L., Hicks, T. R., & Jena, A. B. (2018). Attention deficit–hyperactivity disorder and month of school enrollment. *New England Journal of Medicine, 379*(22), 2122–2130. https://doi.org/10.1056/NEJMoa1806828

Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). Routledge.

Morrow, R. L., Garland, E. J., Wright, J. M., Maclure, M., Taylor, S., & Dormuth, C. R. (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *CMAJ, 184*(7), 755–762.

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry, 164*(6), 942–948.

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56*(3), 345–365.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children* (3rd ed.). Pearson.

U.S. Department of Education. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final rule (34 CFR Parts 300 and 301). *Federal Register, 71*(156).

Wolraich, M. L., Hagan, J. F., Jr., Allan, C., Chan, E., Davison, D., Earls, M., Evans, S. W., Flinn, S. K., Froehlich, T., Frost, J., Holbrook, J. R., Lehmann, C. U., Lessin, H. R., Okechukwu, K., Pierce, K. L., Winner, J. D., Zurhellen, W., & Subcommittee on Children and Adolescents With Attention-Deficit/Hyperactive Disorder. (2019). Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics, 144*(4), Article e20192528. https://doi.org/10.1542/peds.2019-2528

Wolraich, M. L., Lambert, W., Doffing, M. A., Bickman, L., Simmons, T., & Worley, K. (2003). Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *Journal of Pediatric Psychology, 28*(8), 559–568.

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

Chapter 11

# 11 Sensory and multiple disabilities and communication disorders

In these categories, much of the diagnostic work falls to equipment and medical specialists, from the audiometer and the letter chart to the device that measures pressure in the middle ear. Yet the educational question remains the team's to answer: what does this number mean for the student's learning? And how do we assess the student in a way that reaches past the disability to the ability behind it? This chapter takes up these questions and dwells on the hardest of them, which arises when several disabilities converge in a single child, nearly every standardized instrument becomes unusable, and careful functional assessment is all that remains.

**Three messages to the counseling office**

The second-grade teacher: "Khalid daydreams a lot, and he only follows instructions if I repeat them. He often asks the boy next to him what I said. The odd thing is that in some lessons he is fully present, and in others it's as if he's in another world. Should we refer him for an ADHD evaluation?"

> The first-grade teacher: “Rahaf holds the book very close, tires quickly when reading, and confuses letters that look alike.”
>
> The kindergarten teacher: “Abdulrahman is six, and he says ‘galam’ (گلم) instead of ‘qalam’ (قلم, pen) and ‘tamaka’ (تمكة) instead of ‘samaka’ (سمكة, fish). Does he have an articulation disorder?”

The three messages raise questions that this chapter sets out to answer. Could what looks like inattention be a hearing loss that face-to-face conversation with the child fails to reveal? How do we assess vision in a way that serves learning? And how do we tell a child’s dialect apart from a disorder in his speech? We then move on to harder questions that the messages did not raise: How do we assess the ability of a deaf student who communicates in sign language? And how can a judgment be reached about a child who cannot see, cannot speak and cannot hold a pencil?

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Read an audiogram, determining the degree and type of loss by comparing air conduction with bone conduction, and infer which speech sounds the child is missing.
2. Classify results for distance visual acuity and visual field according to the international classification and the Saudi regulations, and specify what functional vision assessment and learning media assessment add to the medical report.
3. Identify, for each communication disorder, its distinguishing feature, the specialist who assesses it and an appropriate assessment procedure.

4. Apply the criteria for developmental language disorder to a child's data, and distinguish it from dialectal difference and second-language learning.
5. Select instruments for assessing the ability of a student who is deaf, blind or has a motor impairment, and determine in a given scenario whether an accommodation invalidates norm-based interpretation of the score.
6. Plan a functional assessment for a child with multiple disabilities or deafblindness for whom standardized instruments are unsuitable.
7. Judge whether the evidence in a new case is sufficient to warrant a judgment, and justify that judgment through the steps of the decision pathway, including ruling out sensory impairment before assigning any other category.

## Introduction

In sensory disabilities and communication disorders, equipment and medical specialists carry a large share of the diagnosis. Yet the importance of these categories extends well beyond their own borders, because weak hearing, vision or language can mimic any other category. A student who follows instructions only when they are repeated, as in one of the messages at the start of the chapter, may be referred for an attention evaluation when the root of the problem lies in his hearing. The educational question remains the team's to answer: what does the medical number mean for the student's learning? And how can the student be assessed in a way that reaches past the disability to the ability behind it, especially when disabilities converge and standardized instruments are, more often than not, unusable?

We open the chapter by explaining why these categories matter for all the others (Section 11.1), then present hearing impairment (Section 11.2) and visual impairment (Section 11.3), from measurement to functional assessment. Next we address communication disorders and the assessment of need for augmentative and alternative communication (Section 11.4), developmental language disorder and how it presents in Arabic (Section 11.5), and the

difference between difference and disorder (Section 11.6), followed by multiple disabilities, deafblindness, physical and health impairments, and brain injury (Section 11.7). We then turn to warning signs and decision pathways (Section 11.8), instrument profiles (Section 11.9), four cases (Section 11.10), and differential diagnosis and co-occurrence (Section 11.11), and we close with the Saudi and Arab context (Section 11.12).

## 11.1 Why do these categories matter for every category?

Hearing and vision screening appears in the decision pathway of almost every category in special education. The reason is that a hearing or vision impairment can resemble any other category: it may look like inattention, as we will see in Khalid's case in this chapter, or like a learning disability, or autism, or even intellectual disability if the student is assessed with instruments that do not suit his senses. Sensory disabilities are therefore the **gateway of exclusion** in every other diagnosis: ruling out hearing and vision impairment is a step that precedes any diagnostic instrument, and we do not judge a child's ability until we are sure that he hears and sees what we present to him.

This chapter brings together three groups of categories joined by a single feature: the disability lies in the very **channels of reception and expression** on which assessment instruments depend. Hearing and visual impairments affect what reaches the child; communication disorders affect what comes out of the child; and multiple and physical disabilities affect both, as well as the ability to point, grasp and write. In all of these cases, the first question in assessment becomes: does the instrument measure what it was designed to measure in this child, or does it measure the disability?

**Common misconceptions**

- **"The child passed newborn hearing screening, so his hearing is fine."** Hearing loss may emerge after birth, may be acquired, after meningitis for example, or may be a fluctuating conductive loss that comes and goes with ear infections. A screening result is a snapshot of a single day.
- **"A hearing aid restores normal hearing."** A hearing aid amplifies sound but does not fully restore its clarity, and its benefit is weaker in noise and at a distance.
- **"Glasses solve vision problems."** Glasses correct refractive errors; they do not repair retinal disease, cerebral visual impairment or a restricted visual field.
- **"Sign language is simplified gesture."** Sign languages are complete natural languages with their own vocabulary and grammar, and they differ from one country to another.
- **"A child who does not speak does not understand."** The absence of speech may be motor or sensory in origin, and many of these children understand far more than they are able to express.

## 11.2 Hearing impairment

### The audiogram

The audiologist measures the softest sound the child can hear at different frequencies, from low-pitched to high-pitched sounds, in each ear separately, and plots the results on a chart called an **audiogram** (**Audiogram**). The horizontal axis represents frequency in hertz, from low on the left to high on

the right. The vertical axis represents sound intensity in decibels, with louder sounds toward the bottom. The lower the marks fall, the louder a sound the child needs in order to hear, and thus the more severe the hearing loss.

Hearing is measured in two ways. In **air conduction** (**Air conduction**), sound is delivered through headphones and travels through the outer and middle ear to the inner ear. In **bone conduction** (**Bone conduction**), a small vibrator is placed behind the ear, and sound reaches the inner ear through the bones of the skull, bypassing the outer and middle ear. Comparing the two is what reveals where the problem lies: if air conduction is reduced while bone conduction is intact, the problem lies in the pathway of sound to the inner ear; if both are reduced, the problem lies in the inner ear or the nerve, as in the left ear in Figure 11.1.

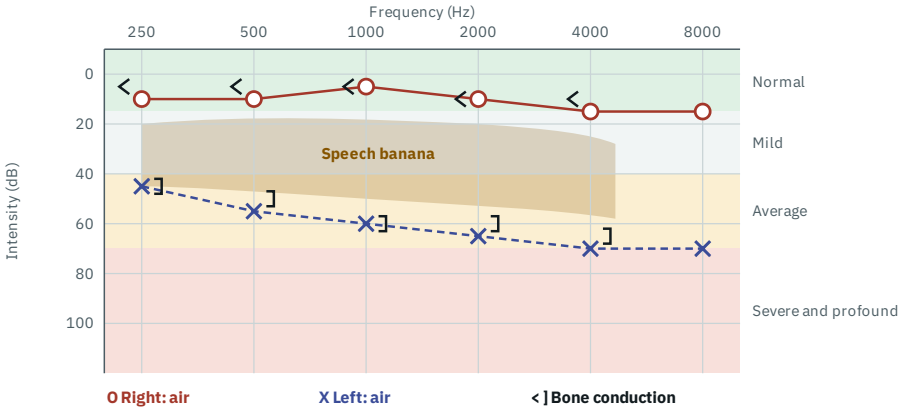



**Figure 11.1:** A hypothetical audiogram for Khalid. The right-ear marks (circles) fall between 5 and 15 dB, indicating normal hearing. The left-ear marks (× symbols) slope from 45 dB at the low frequencies to 70 dB at the high frequencies, averaging about 60 dB across the main speech frequencies, which indicates a moderate-to-severe loss. The left-ear bone-conduction marks (], measured with the right ear masked by noise so that it could not respond in its place) lie close to the air-conduction marks, with no gap between them: a **unilateral sensorineural** loss. Most of the shaded area labeled "Speech banana" lies above the left-ear marks, so Khalid can barely hear speech in that ear and relies on his right ear alone. Had the bone-conduction marks been normal, with a gap between them and the air-conduction marks, the loss would have been conductive.

The shaded area labeled “Speech banana” is the range in which the sounds of ordinary speech fall. Any speech sound that lies above the child’s marks, that is, softer than them, goes unheard. This explains why some children hear “that someone is talking” without understanding the words: they hear the loud, low-pitched sounds such as vowels, but miss many of the soft, high-pitched sounds such as /s/, /f/ and /θ/.

## Degrees and types of hearing loss

Degree is usually determined by the average of the thresholds at the main speech frequencies in each ear. Table 11.1 presents a classification commonly used in assessing children (Clark, 1981).

**Table 11.1: Degrees of hearing loss and what a child may miss at each (Clark, 1981).**

| Score | Approximate range (dB) | What the child may miss without a hearing aid |
|---|---|---|
| **Slight** | 16–25 | Some faint and distant speech; tires in noise |
| **Mild** | 26–40 | Much of faint or distant speech, and high-pitched consonant sounds |
| **Average** | 41–55 | Most speech at normal conversational level |
| **Moderately severe** | 56–70 | Nearly all ordinary speech; needs a hearing aid and language intervention |
| **Severe** | 71–90 | Does not hear speech, even loud speech, without amplification |
| **Profound** | 91 or more | May perceive some loud sounds as vibration; relies on vision or on a cochlear implant |

Classifications differ across sources, and the differences are substantial. In its world report on hearing, the World Health Organization adopts another classification based on the average of four frequencies (500, 1000, 2000 and 4000

Hz) in the **better ear** (World Health Organization [WHO], 2021), shown in Table 11.2. A child whose average threshold is 40 dB thus has a "mild" loss in Clark's classification but a "moderate" loss in the WHO classification. Every report should therefore state the classification it has adopted, and the numbers should be read alongside the descriptive labels.

**Table 11.2: Degrees of hearing loss in the World Health Organization classification (WHO, 2021).**

| Score | Average threshold in the better ear (dB) |
|---|---|
| **Normal hearing** | Less than 20 |
| **Mild** | 20 to less than 35 |
| **Average** | 35 to less than 50 |
| **Moderately severe** | 50 to less than 65 |
| **Severe** | 65 to less than 80 |
| **Profound** | 80 to less than 95 |
| **Complete or total** | 95 or more |
| **Unilateral** | Less than 20 in the better ear, and 35 or more in the worse ear |

The audiologist also determines the **type** of loss by comparing the two kinds of conduction:

- **Conductive** (**Conductive**): bone conduction is normal and air conduction is reduced, with a gap between them. It originates in the outer or middle ear. Among its most common causes in children is fluid collecting behind the eardrum with middle ear infection, and it is often temporary or **fluctuating**: it comes with a bout of infection and eases afterward, and it is treated medically.
- **Sensorineural** (**Sensorineural**): both air and bone conduction are reduced by roughly the same amount, as in the left ear in Figure 11.1. It originates in

the inner ear or the auditory nerve and is usually permanent.

- **Mixed** (**Mixed**): both are reduced, with a gap between them, so the loss combines the two types.

## Presentations missing from the standard audiogram

Three presentations of hearing loss can easily escape a team that asks only whether the child can hear or not.

- **Unilateral hearing loss** (**Unilateral hearing loss**): the child hears well face to face in a quiet room and so passes many everyday observations. But he struggles in noise, has difficulty telling where a sound is coming from, and misses speech coming from the side of the weaker ear. Even so, the claim that "one ear is enough" is mistaken: these children are at greater risk than their peers for difficulties in language and achievement, and they need monitoring and classroom accommodations, as we will see in Khalid's case (11.10).
- **Auditory neuropathy spectrum disorder** (**Auditory neuropathy spectrum disorder**): the inner ear works, so otoacoustic emissions appear normal, but the signal is not transmitted to the brain in an orderly way, and the auditory brainstem response is absent or abnormal. The distinguishing educational sign is that the child's understanding of speech is far worse than the teacher would expect from the audiogram, and that hearing may fluctuate from day to day. Because some screening programs rely on otoacoustic emissions alone, these children may pass the screening.
- **Fluctuating hearing loss** (**Fluctuating hearing loss**): its best-known form is the conductive loss that accompanies recurrent middle ear infections. It is among the conditions that most closely resemble inattention, because the child is "present" one week and "absent" the next, and because a single test on a good day may come back normal.

**For the specialist: The debate over central auditory processing disorder**

The team is sometimes referred a child whose audiogram is normal but who is said to have **central auditory processing disorder** (**Central auditory processing disorder**): difficulty understanding speech in noise, distinguishing similar sounds or following long instructions. The American Speech-Language-Hearing Association defines it as a deficit in the central nervous system's processing of auditory information that is not attributable to higher-order language or cognitive deficits, while acknowledging that it may co-occur with other disorders (American Speech-Language-Hearing Association [ASHA], 2005). The concept, however, is widely contested. Most of its tests depend on language, attention and memory, so their results are hard to separate from weaknesses in those abilities; there is no agreed reference test; and its symptoms overlap with those of ADHD and developmental language disorder. The practical conclusion for the team is this: the explanation should be considered only after peripheral hearing has been fully assessed and language and attention have been evaluated, and it should then be treated as a description of a pattern of difficulty that merits classroom accommodations, added to whatever other diagnoses are established.

## Screening and diagnosis

- **Newborn screening:** quick tests that require no response from the infant, such as measuring **otoacoustic emissions** from the inner ear (**Otoacoustic emissions, OAE**) or the **auditory brainstem response** (**Auditory brainstem response, ABR**). The Joint Committee on Infant Hearing (**Joint Committee on Infant Hearing, JCIH**) recommends the 1-3-6 principle: screening by the end of the first month, diagnosis by the end of the third month and intervention by the end of the sixth month. Its most recent position statement adds a more ambitious target for systems that have already achieved this: 1-2-3, that is, diagnosis by the end of the second month and intervention by the end of the third (JCIH, 2019). A well-known study found that children whose hearing loss was identified and whose intervention began before six months of age had clearly better language than those identified later (Yoshinaga-Itano et al., 1998). Section 11.12 describes the national newborn hearing screening program in the Kingdom.
- **School screening:** the student is presented with tones at 20 dB at 1000, 2000 and 4000 Hz in each ear. If the student fails to hear any of them, the screening is repeated, and if the tone is missed again, the student is referred to an audiologist (ASHA, 1997).
- **Diagnosis:** this is the audiologist's task, using pure-tone audiometry by air and bone conduction, speech audiometry for detection and discrimination, middle ear pressure measurement (**Tympanometry**) and other procedures. The first profile in Section 11.9 presents these procedures.

## What educational assessment adds

An audiogram yields a number, but it does not tell us how the child hears in a classroom with thirty students and an air conditioner. The educational team therefore completes the picture by assessing **functional hearing** (**Functional hearing**), that is, how the child actually uses his hearing in class; by assessing language in the mode through which the child communicates, whether spoken

Arabic or sign language; and by determining the most suitable mode of instruction. The child's cognitive ability is measured with nonverbal instruments, or instructions are given in the communication mode he has mastered, rather than through verbal tests delivered in a language he cannot hear well (see the box in 11.9).

## The perspective of the Deaf community and sign language

Many deaf people see themselves as members of a linguistic and cultural community with its own language, not as individuals who lack hearing. Whether or not the team adopts this perspective, it has three consequences for assessment. First, **Saudi Sign Language** is a complete language, so a child who is fluent in it has an intact first language, and his language development should be judged in that language before his spoken or written Arabic. Second, assessment should be conducted in the child's stronger language, by an examiner fluent in it or with the help of a qualified interpreter, with an explicit statement in the report that interpretation alters the standardized conditions of administration. Third, the greatest danger to a deaf child is **language deprivation**, which is more harmful than deafness itself: reaching school age without a fluent language, whether spoken or signed, because no one gave the child access to a language he could perceive in full. Such a child may look on tests as if he has an intellectual disability, when what he actually has is deprivation of language.

At the time of writing, we know of no published standardized instrument for assessing Saudi Sign Language. Its assessment therefore relies on nonstandardized sources: a sample of the child's narrative in sign language, evaluated by a teacher or a deaf adult fluent in it; comparison with what is known about language development in general; and questions to the family about the language in which the child was raised and when his exposure to it began.

# 11.3 Visual impairment

## Visual acuity

**Visual acuity** (**Visual acuity**) is the sharpness of vision. It is measured with a chart of letters or symbols of decreasing size and expressed as a fraction. A value of 6/6 means that the person sees at six meters what a person with normal vision sees at six meters. A value of 6/60 means that the person must come as close as six meters to see what a person with normal vision sees from sixty meters. In systems that use feet instead of meters, 6/6 is written as 20/20. With young children and those who do not know letters, charts with simple symbols are used, such as LEA symbols, a circle, a square, an apple and a house, or four-letter charts (**HOTV**) on which the child points to the matching letter on a card in front of him without having to name it.

This is **distance acuity**. **Near acuity** (**Near acuity**) is measured at reading distance and is not always equal to distance acuity: some children see the board but tire when reading a book, and others the reverse. Nor does acuity reveal **contrast sensitivity** (**Contrast sensitivity**), the ability to distinguish an object from its background when their colors are similar, such as gray print on white paper or a stair step the same color as the floor. A student's acuity may be acceptable on a black-on-white chart, yet he stumbles over a faded photocopied worksheet.

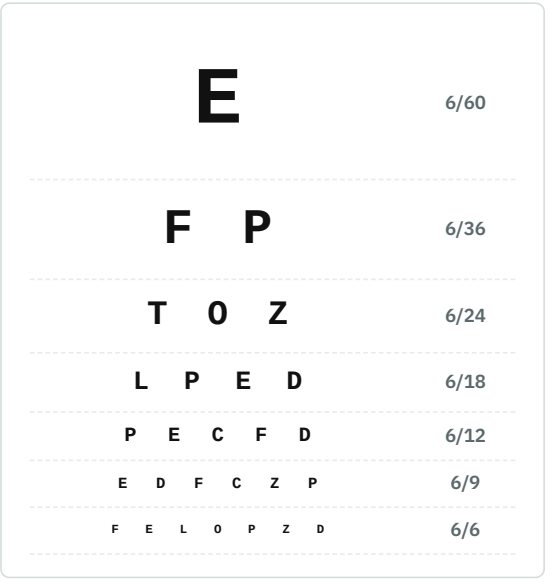



**Figure 11.2:** An illustration of a visual acuity chart, with the categories of distance vision impairment in the International Classification of Diseases, Eleventh Revision (ICD-11) (World Health Organization, 2019). Acuity is measured in the better eye with correction. A person whose visual field in the better eye is restricted to 10 degrees or less is also classified as blind, whatever their acuity. This differs from the Saudi regulations, under which a student is considered educationally blind if acuity is below 6/60 or the visual field is narrower than 20 degrees (see 11.12).

## Visual field

The **visual field** (**Visual field**) is the area a person can see without moving the eyes. Visual acuity may be good at the center of vision while the field is restricted, so the student sees what is directly in front of him clearly, as if looking through a tube, but misses what lies to either side. Such a student may read the small letters on the chart yet bump into classmates in the hallway and fail to see the whole board.



**Normal field**
Sees the whole scene, out to both sides

**Restricted field**
Sees only the center and misses what lies to either side

**Figure 11.3:** One scene viewed with two visual fields. Acuity alone does not reveal a restricted field, which is why each is measured.

## Functional vision assessment, learning media assessment, and orientation and mobility

The ophthalmologist or optometrist provides acuity, field and the medical diagnosis. The educational assessments are carried out by the teacher of students with visual impairments and other specialists:

- **Functional vision assessment** (**Functional vision assessment**): shows how the student actually uses vision in the classroom, the hallways and the playground: what print size the student reads comfortably, from what distance the board is visible, what lighting and contrast suit the student, and how long the student can sustain visual work before tiring.
- **Learning media assessment** (**Learning media assessment**): determines the most suitable medium for the student to learn reading and writing: regular print with magnification devices, large print, braille, a combination of media, or auditory media. The decision rests on observing how the student explores the environment, primarily through vision or through touch, and on the student's efficiency and speed in reading with each medium (Koenig & Holbrook, 1995).

- **Orientation and mobility assessment** (**Orientation and mobility**): conducted by an orientation and mobility specialist, it shows how the student determines where he is and moves about safely and independently: in hallways and on stairs, in dim lighting and when crossing the street, and whether he needs training with the white cane.

Assessment of a student with a visual impairment extends beyond the general curriculum. Specialists in educating this population have formulated what is known as the **expanded core curriculum** (**Expanded core curriculum**): skills that sighted children usually learn through incidental observation but that a blind or low-vision student needs to be taught explicitly. It comprises nine areas: compensatory and access skills such as braille, orientation and mobility, social interaction, independent living, recreation and leisure, career education, assistive technology, sensory efficiency, and self-determination. Each area requires an assessment of the student's present level, because a plan that overlooks them teaches the student the curriculum while leaving him unable to access it.

**Cerebral visual impairment**

Some children have healthy eyes, but the brain does not process what the eye sees in a typical way. This is called **cerebral or cortical visual impairment** (**Cerebral/Cortical visual impairment, CVI**), and it is one of the common causes of visual impairment among children with multiple disabilities. It appears in characteristic patterns, such as a preference for a particular color, difficulty seeing objects in a visually cluttered environment and delayed visual response. Detecting it requires a specialized functional assessment that complements the conventional eye examination (Roman-Lantzy, 2018). We will meet it in Salman's case (11.7).

## 11.4 Communication disorders

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**), classifies communication disorders into several types (American Psychiatric Association [APA], 2022). Under its implementing regulations, the U.S. Individuals with Disabilities Education Act (**IDEA**) adds **voice disorders** to the category of "speech or language impairment"; these do not appear in the diagnostic manual because they are not neurodevelopmental disorders (U.S. Department of Education, 2006, 34 CFR §300.8(c)(11)). All of these disorders are assessed by the speech-language pathologist:

**Table 11.3: Types of communication disorders and their distinguishing features.**

| Disorder | Distinctive features | Example |
|---|---|---|
| **Language disorder** | Persistent difficulty in acquiring and using language, in vocabulary, sentence structure or connected discourse, in comprehension or expression (see 11.5) | A six-year-old who speaks in two-word sentences and does not understand questions beginning with "why" |
| **Speech sound disorder** | Persistent difficulty producing speech sounds that is inappropriate for the child's age and interferes with the intelligibility of speech | A five-year-old who substitutes so many sounds that only his family can understand him |
| **Childhood-onset fluency disorder** | Repetitions or prolongations of sounds and syllables, or blocks before words, that interfere with communication; commonly known as stuttering | "B-b-b-baba," or a long silent block before a word |

| Disorder | Distinctive features | Example |
|---|---|---|
| **Social (pragmatic) communication disorder (Social [pragmatic] communication disorder)** | Difficulty using language in social situations, without restricted and repetitive patterns of behavior, which is what distinguishes it from autism spectrum disorder | Does not know when to start or end a conversation and does not grasp hints |
| **Voice disorders** | Deviation in voice quality, pitch, loudness or resonance from what suits the child's age and sex, such as persistent hoarseness or the excessive nasality that accompanies cleft palate | A student whose voice has been hoarse for months and gives out by the end of the school day |

In assessment, the speech-language pathologist uses standardized language tests such as the Clinical Evaluation of Language Fundamentals, Fifth Edition (**CELF-5**) (Wiig et al., 2013), presented in the third profile in 11.9. The pathologist also uses picture-based articulation tests in which the child names pictures covering the sounds of the language in their various positions in the word: initial, medial and final. A **language sample** (**Language sample analysis**) drawn from the child's natural speech is analyzed, yielding measures such as mean length of utterance. For stuttering, the percentage of stuttered syllables in a speech sample is calculated, and a standardized instrument may be used to rate severity, such as the Stuttering Severity Instrument, Fourth Edition (**SSI-4**) (Riley, 2009). For voice disorders, voice therapy begins only after an ear, nose and throat physician has examined the larynx, because persistent hoarseness may have an organic cause that requires medical treatment.

**Comparison with ICD-11**

The International Classification of Diseases, Eleventh Revision (**ICD-11**), was adopted in 2019 and came into effect on 1 January 2022. It places speech and language disorders among the neurodevelopmental disorders under the heading "developmental speech or language disorders," distinguishing **developmental speech sound disorder**, **developmental speech fluency disorder** and **developmental language disorder** (**Developmental language disorder**). The most important difference from the American manual is that pragmatic impairment in ICD-11 is a subtype of developmental language disorder, "with impairment of mainly pragmatic language," alongside other subtypes, including one that affects both receptive and expressive language and one that affects mainly expressive language. The classification states explicitly that the language in question may be spoken or signed (World Health Organization, 2024). Hearing and visual impairments, by contrast, are classified in ICD-11 as health conditions in other chapters, outside the neurodevelopmental disorders.

**Arabic is not a translation of English**

Arabic differs from English in its sounds, the order in which they are acquired and its grammatical structures. An articulation or language test translated from English does not measure Arabic sounds that do not exist in English, nor does it reflect their order of acquisition. Researchers have studied the order of acquisition of Arabic consonants among Jordanian children, first at preschool age and then at older ages (Amayreh & Dyson, 1998; Amayreh, 2003). Such studies, along with instruments built on Arabic, should inform judgments about a child's articulation, bearing in mind that the child's dialect may differ from that of the study sample.

### Assessing the need for augmentative and alternative communication

For some children, speech is not enough to communicate, either because they do not speak or because their speech is unintelligible to anyone outside the family: the child with cerebral palsy affecting the speech muscles, the nonspeaking autistic child, the child with multiple disabilities. For these children, **augmentative and alternative communication** (**Augmentative and alternative communication, AAC**) is assessed; it includes signs, pictures, symbol boards and speech-generating devices. This assessment does not require the child first to prove "readiness" by reaching a particular cognitive or linguistic level; every child whose speech does not meet his communicative needs is a candidate. The assessment answers practical questions: How can the child make a selection, by touch, by eye gaze, or with a switch pressed with the head? What can the child see, and what size of symbol can he discriminate? Does he understand that a picture stands for an object? With whom does he communicate, and what does he want to talk about? One or more systems are then tried in real situations and judged by what the child achieves with them. The evidence indicates that alternative communication often supports the development of speech, contrary to the common fear that it hinders it.

## 11.5 Developmental language disorder

Children whose language was delayed for no apparent reason used to be described in many ways: "language delay," "developmental aphasia" and "specific language impairment" (**Specific language impairment**). The last of these usually required the child's language to be clearly below his nonverbal intelligence. An international panel of specialists then came together in a consensus study known as **CATALISE** and settled on a unified term: **developmental language disorder** (**Developmental language disorder, DLD**) (Bishop et al., 2017). The international classification (ICD-11) has adopted the same term.

## What has changed?

- **Language disorder or developmental language disorder?** "Language disorder" is used for any child with a language difficulty that is likely to persist beyond early childhood and that affects communication or learning. If the difficulty is associated with a known condition, such as sensorineural hearing loss, autism, intellectual disability or a genetic syndrome, the term is "language disorder associated with…." The term "developmental language disorder" is reserved for children whose difficulty is not associated with any of these conditions (Bishop et al., 2017).
- **A discrepancy with nonverbal intelligence is not required.** A child whose nonverbal intelligence is in the low average range, above the level of intellectual disability, can be diagnosed with developmental language disorder. A British population study found that about 7% of children at school entry have the disorder, and that children with low nonverbal intelligence did not differ from others in the severity or impact of their language difficulties (Norbury et al., 2016). Requiring a discrepancy had denied children services without justification.
- **Co-occurrence does not rule out the diagnosis.** The presence of ADHD, dyslexia or developmental coordination disorder does not preclude a diagnosis of developmental language disorder.

## How does it present in Arabic?

Describing this disorder in Arabic requires more than translating what is known about it in English. Arabic is a derivational language rich in morphology, and difficulty in it surfaces in places specific to the language. Table 11.4 presents areas that the speech-language pathologist would do well to examine in the child's language sample and in structured tasks, judging them against the standard of the child's own dialect.

**Table 11.4: Areas to examine in a child's Arabic when developmental language disorder is suspected.**

| Area | What is observed | Example prepared by the author |
|---|---|---|
| **Morphological patterns** | Difficulty deriving words from their roots according to their patterns, and reliance on a single form that the child repeats | Says "الرجل الذي يكتب" ("the man who writes") and does not use "الكاتب" ("the writer") at an age when his peers use it |
| **Plural forms** | Overgeneralizing the sound masculine or feminine plural to words that take a broken plural in his dialect, and continuing to do so after peers have moved past it | "كتابات" instead of "كتب" ("books"), and "ولدين" ("two boys," dual) for more than two |
| **Attached pronouns** | Omitting or confusing them, or using the noun in place of the pronoun | "أعطى أحمد الكرة أحمد" ("Ahmad gave the ball to Ahmad") instead of "أعطاه الكرة" ("he gave him the ball") |
| **Agreement** | Failure of the verb or adjective to agree with its subject or the noun it modifies in gender and number | "البنت راح" ("the girl went," with a masculine verb), "الأولاد يلعب" ("the boys plays") |
| **Connected discourse** | Short, disjointed stories, few connectives, and difficulty retelling a story heard in its correct sequence | Retells a five-event story in two unconnected sentences |
| **Comprehension** | Difficulty understanding long sentences, passive constructions and complex questions, sometimes masked by reliance on context | Carries out the first of two instructions and leaves the second |

## The dilemma of diglossia

The Arab child speaks a dialect at home and encounters Modern Standard Arabic at school and in books, and the examiner's or the test's dialect may differ from the child's own. When the child is measured in Standard Arabic or in another dialect, his performance appears lower than it actually is; this is **diglossia** (**Diglossia**). Matters become more complicated still with a child raised in a language other than Arabic. Assessment therefore rests on three principles: language is measured first in the child's own dialect; if the child is bilingual, his other language is examined as well, because a true disorder appears in both languages; and tasks less influenced by prior language experience are used, such as dynamic assessment (see 11.6 and Chapter 5) and nonword repetition.

## Its role in differential diagnosis

Developmental language disorder lies at the heart of differential diagnosis for several categories. **With intellectual disability**, the deficit is general, encompassing nonverbal ability and adaptive behavior, whereas in developmental language disorder nonverbal ability and adaptive behavior outside the language domain are clearly stronger. **With weak reading comprehension**, many students who read fluently but do not understand have a weakness in oral language itself, so listening comprehension is measured before a reading difficulty is concluded. **With autism spectrum disorder**, the two may resemble each other in weak language and social difficulties, but the social difficulties in developmental language disorder stem from the language weakness, are not accompanied by restricted and repetitive patterns of behavior, and leave nonverbal communication and the desire to interact intact. **With hearing loss**, the loss is ruled out first, including fluctuating loss that a single test may miss.

## 11.6 Difference versus disorder

Among the most important things a teacher should know is that a child who speaks in his own dialect does not have a disorder. Children learn to talk from their families and pronounce words as the people around them do. A child who says "گلم" (galam, "pen") in a community where everyone pronounces the qaf as a g is speaking entirely correctly by the standard of his dialect. A disorder, by contrast, is measured as a departure from the pattern of the child's own dialect. Among the patterns that warrant evaluation is the replacement of **fricatives** (**Fricatives**), sounds produced with a continuous stream of turbulent air, such as s, sh and z, with **stops** (**Stops**), sounds in which the airflow is blocked and then released in a single burst, such as t and d. This substitution is called **stopping** (**Stopping**).

**Dialectal difference: not a disorder**

قلم → گلم (qalam → galam, "pen")

Pronouncing the qaf as a g is a feature of many dialects across the Arabian Peninsula. The child learns Modern Standard Arabic at school, and his pronunciation is not treated.

**A pattern that warrants evaluation**

سمكة → تمكة (samaka → tamaka, "fish")

Replacing a fricative such as s with a stop such as t is a pattern that fades with development. Because the child's dialect does not pronounce s as t, its persistence across an entire class of sounds, together with poor speech intelligibility, warrants evaluation.

**Figure 11.4:** A child's pronunciation is judged by the standard of his dialect and community.

The same rule applies to a student learning Arabic as a second language. His errors in sentence structure may simply be a natural stage in learning a new language. To tell the two apart, we ask about his first language: does he speak it with fluency appropriate for his age? A genuine language disorder appears in both languages, whereas difficulty learning a second language appears in

the new language alone. Here **dynamic assessment**, presented in Chapter 5, is useful: the child is taught a new language skill during the assessment session itself, and the examiner observes how quickly he learns it. A child who learns quickly once taught is most likely showing limited experience with the language, whereas one who learns very slowly despite instruction warrants a more in-depth evaluation.

## 11.7 Multiple disabilities, deafblindness, physical and health impairments, and brain injury

In these categories, standardized instruments recede to the margins, because most of them assume a child who can see, hear, point, hold a pencil and sit still for a full hour. Functional assessment, repeated observation in natural settings and curriculum-based measurement become the primary sources of evidence. Table 11.5 presents the definitions of these categories in U.S. law and the focus of assessment in each. Saudi regulations specify corresponding categories, which are presented in Section 11.12.

**Table 11.5: Categories of multiple, physical and health impairments and brain injury in U.S. law, and the focus of assessment in each.**

| Category | Definition in U.S. law (abridged) | Focus of assessment |
|---|---|---|
| **Deafblindness (Deaf-blindness)** | Concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational needs that they cannot be accommodated in programs designed solely for deaf children or solely for blind children | Communication: how the child receives and expresses information, and a joint functional assessment of his remaining hearing and vision |

| Category | Definition in U.S. law (abridged) | Focus of assessment |
|---|---|---|
| **Multiple disabilities** | Concomitant impairments, such as intellectual disability with blindness or with an orthopedic impairment, the combination of which causes such severe educational needs that they cannot be accommodated in a program designed for one of the impairments alone. Does not include deafblindness | Functional skills in daily routines, communication, and the means of access to learning |
| **Orthopedic impairment** | A severe orthopedic impairment that adversely affects educational performance, including impairments caused by a congenital anomaly, by disease, or by other causes such as cerebral palsy and amputation | Access: seating, writing and mobility, and adapting assessment instruments so that motor impairment is not measured in place of the intended ability |
| **Other health impairment** | Limited strength, vitality or alertness due to chronic or acute health problems, such as asthma, diabetes, epilepsy, heart conditions, sickle cell anemia, and attention-deficit/hyperactivity disorder, that adversely affects educational performance | The effect of the condition and its treatment on attention, energy and attendance, and distinguishing between instruction the student has missed and what he is unable to do |
| **Traumatic brain injury** | An acquired injury to the brain caused by an external physical force, resulting in functional or psychosocial impairment that adversely affects educational performance. Does not include congenital or degenerative injuries or those induced by birth trauma | Comparison with the student's level before the injury, and tracking rapid change in the first months |

Source: IDEA (2004) and its implementing regulations (U.S. Department of Education, 2006, 34 CFR §300.8(c)).

## Assessing the child with deafblindness

A child with deafblindness is seldom both totally deaf and totally blind; most of these children have residual hearing, residual vision, or both. Assessment therefore begins with **a joint functional assessment of hearing and vision**: what the child sees, at what distance and in what lighting; what he hears and under what conditions; and how he uses the two senses together. This assessment is carried out over several days in familiar settings, because the child's responses are affected by fatigue, medication and location.

Next comes the **communication assessment**, the centerpiece of the entire evaluation. Many of these children are at the **pre-symbolic communication** stage: they communicate through body movements, changes in facial expression, crying, and taking an adult's hand and guiding it, before words, signs or pictures have developed. The assessment asks: does the child act intentionally to influence others? Does he use real objects as symbols, such as holding a spoon to ask for food? Does he understand **tactile communication** (**Tactile communication**), such as signs made with his hand resting on the signer's hand so that he can feel them, or consistent touch cues on his body that announce what is about to happen? Among the descriptive tools used here is the **Communication Matrix**, which describes seven levels of communication, from unintentional behavior to language, without yielding a standard score. Family members and others who know the child well take part in this assessment, because they are best able to read his subtle signals.

## Orthopedic impairment: when does an accommodation invalidate the norms?

A child with cerebral palsy may understand the question and know the answer, yet be unable to point to it quickly, arrange the pieces or draw the figure. If a standardized test is administered to him as it stands, it measures his hands, not his mind. The response mode is therefore adapted: pointing with eye gaze instead of a finger, selecting an option with a switch, or saying the answer for the examiner to record, along with comfortable, supportive seating that stabilizes the body.

**For the specialist: accommodations that preserve the norms and accommodations that invalidate them**

Norms assume that the student was given the same task under the same conditions as the norming sample. The decisive question, then, is whether the accommodation changes **what the test measures** or only **how the student accesses it**. If the test measures reasoning, changing the way the student indicates an answer usually leaves the construct intact, and the score can be interpreted against the norms, with a note in the report. But if timing is removed from a test of speed, if drawing is replaced with selecting an option on a test of visual-motor integration, or if the items of a reading test are read aloud to the student, the construct itself has changed and the standard score becomes meaningless. In that case the result is written as a description of what the student did under the accommodated conditions and is not converted into a standard score or a percentile rank. (The percentile rank is the percentage of the student's peers in the norming sample who scored below him, so a rank of 84 means that he outperformed 84% of them; it is precisely this comparison that has lost its basis here.) The same holds for converting a test into braille or large print: an enlarged item may become easier or harder, and braille reading is slower than print reading, so item timing and difficulty change. Every accommodation is named in the report, along with the examiner's judgment of its effect on the norms.

## Health impairments: attention and attendance

Students with chronic illnesses are often referred for "poor attention" or "low achievement." The first question here is whether what we are seeing is an effect of the illness or its treatment, or a separate difficulty. **Absence seizures** appear as repeated lapses lasting a few seconds that a teacher may mistake for

daydreaming, and some antiepileptic medications can affect alertness and memory. Low blood sugar in a student with diabetes weakens concentration at particular times of day. Frequent absences for treatment or hospitalization, as may occur with sickle cell anemia, leave gaps in instruction, not in ability. Assessment therefore takes three things into account: it is scheduled at a time when the student is in his best health; the physician is asked about the expected effect of medication; and the student's achievement is compared with the instruction he has actually received, alongside comparison with his class, with measurement repeated once his condition stabilizes.

## Acquired brain injury and distinguishing it from intellectual disability

A student who was learning like his peers and then, after a car accident or a fall, saw his performance decline differs from a student with intellectual disability, even if their scores are similar. The difference is established through three lines of evidence. **The first is history:** the injured student has a record of performance before the injury, in grades, reports and work samples, that serves as a baseline for comparison. **The second is the pattern of abilities:** injury usually leaves an uneven profile, such as slow processing and weak new learning and attention, while knowledge acquired before the injury remains intact, whereas the deficit in intellectual disability is more general. **The third is the trajectory of change:** the injured student's performance may improve rapidly in the first months, so assessment is repeated at intervals before any lasting judgment is made. The criterion of onset during the developmental period is not decisive on its own, because a child injured in childhood is injured during the developmental period; for this reason the diagnostic manual permits a diagnosis of intellectual disability in such cases if its criteria are met, together with a diagnosis of a neurocognitive disorder (APA, 2022). The educational difference here is that this student's needs change quickly, so his plan is reviewed more often than those of other students.

## A case of multiple disabilities: Salman, age 8

**A judgment without a standard score**

**The story:** Salman has spastic quadriplegic cerebral palsy affecting his arms, his legs and the muscles of speech; he does not speak and uses a wheelchair. He also has epilepsy controlled by medication, and his mother says that he "sometimes sees and sometimes doesn't." He was referred to the assessment and diagnosis team to determine his category and services before he started school. A psychologist at a center he had attended earlier had tried to administer a nonverbal intelligence test, stopped after a few minutes and wrote: "Did not respond to the items; score cannot be computed."

The team decided not to try again with another test, and instead to gather evidence over four weeks in settings familiar to Salman, at home and at the center, with his mother taking part.

**Table 11.6: Evidence in Salman's case.**

| Source | Result |
|---|---|
| **Hearing (auditory brainstem response)** | Normal in both ears; this measure requires no response from Salman |
| **Eye examination** | Both eyes are structurally intact, and the examination does not explain his weak visual responses |
| **—** | He looks at shiny red objects more than at others, responds a few seconds after they are presented, loses them when they are placed against a cluttered background, and sees best what is presented on his left side. The pattern is consistent with cerebral visual impairment |

| Source | Result |
|---|---|
| **Physical therapist and occupational therapist** | He can turn his head to the right voluntarily, and after brief training he presses a large switch mounted on his wheelchair with his cheek |
| **Communication assessment** | He smiles and moves his legs when he hears his brother's voice, cries when an activity he enjoys stops, and calms down when it resumes. By the third week he had begun pressing the switch to restart a song the specialist had paused; that is, he acts intentionally to influence others. He has not yet shown that he understands pictures as symbols |
| **Interview with his mother and observation of routines** | He depends on others for eating, dressing and getting around. He knows his daily routine: he calms down when placed in his car seat and anticipates his bath when he hears running water |
| **Neurologist** | Seizures have been controlled for a year; the current dose is not expected to affect his alertness much, and morning assessment is preferable |

**How the team read the evidence:** No standardized intelligence instrument is suitable for Salman. Verbal instruments require speech, and nonverbal ones require vision and hand movement, and he has neither to the degree they assume. Any score computed for him would measure his motor and visual impairments rather than his ability. The team therefore made no judgment about the level of intellectual disability. The international classification provides, for a case like this, the category "disorder of intellectual development, provisional" (**Disorder of intellectual development, provisional**), used when there is evidence of intellectual impairment but valid assessment is not possible because of a sensory or motor impairment (World Health Organization, 2024).

**Decision:** Salman is found eligible for special education services under the category of multiple disabilities, and the report documents the established impairments with their evidence: orthopedic impairment, cerebral visual impairment, and communication at the pre-symbolic stage. The report states that the estimate of his intellectual ability is deferred and that no IQ score computed for him should be relied upon. Augmentative communication begins with a single switch and consistent touch cues for activities, and materials are visually adapted to his pattern: a single bright color, a simple background, and presentation on his left side. His communication is to be reassessed in six months.

**The lesson of the case:** The absence of a standard score does not mean the absence of evidence. The judgment here rested on repeated, documented observations from multiple sources in natural settings, each answering an educational question: how does Salman reach the world, and how does he act on it? Had the team settled for the phrase "did not respond to the items," Salman would have been left without a plan.

> **What evidence could change the decision?** If, at reassessment in six months, Salman showed a consistent, intentional choice between two pictures or two symbols using the switch, it would become possible to move to tasks measuring symbolic understanding with a response mode suited to him. If his antiepileptic medication changed or his seizures returned, his visual and communicative responses would need to be reinterpreted in that light.

# 11.8 Warning signs and decision pathways

## Warning signs

Teachers and families are the first to notice. Table 11.7 brings together signs that call for referral to a specialist; their value lies in raising the alert. Any one of them warrants a hearing or vision examination, regardless of the result of earlier screening.

**Table 11.7: Warning signs in hearing, vision and communication.**

| Area | Signs at home or in class | Refer to |
|---|---|---|
| **Hearing** | Often asks for things to be repeated; watches classmates to know what to do; understands better when he can see the speaker's face; turns up the television; mispronounces and misspells high-pitched sounds; has recurrent ear infections; or performs inconsistently from one week to the next | An audiologist, and an ear, nose and throat specialist when an infection is suspected |
| **Vision** | Holds the book close to his face; tilts his head; closes one eye; rubs his eyes; bumps into objects or stumbles on stairs; tires quickly when reading; copies from the board with many errors; or one eye appears to turn | An ophthalmologist, then a teacher of students with visual impairments for a functional assessment |
| **Language** | Does not combine two words into a phrase by age two; at school age speaks in shorter, simpler sentences than his peers; does not understand complex instructions; or cannot recount what happened to him in order | A speech-language pathologist, after a hearing examination |

| Area | Signs at home or in class | Refer to |
|---|---|---|
| **Speech, fluency and voice** | Speech is unintelligible to people outside the family after age four; stutters with facial tension or avoids speaking; or hoarseness persists for weeks | A speech-language pathologist, and an ear, nose and throat specialist for hoarseness |
| **At any age** | Loss of words or skills the child had already acquired | Urgent referral to a physician |

**For the teacher: three observations in class**

- **Change the channel and note the difference:** Give the student a spoken instruction while you stand behind him, then a similar instruction while he can see your face, then a written instruction. If he carries out the instruction clearly better when he sees your face or reads it, hearing should be checked before attention.
- **Record when and where:** If you find yourself saying "sometimes he hears and sometimes he doesn't," write down the days, where the speaker was in relation to the student, and the noise level. Fluctuation linked to colds, and a difference between one side of the student and the other, are two patterns an audiologist knows well.
- **Enlarge the print once:** Make an enlarged copy of a page from the student's textbook, and compare his errors and fatigue on the two pages. If his errors on similar-looking letters disappear, the root of the problem lies in vision, and his reading itself is most likely sound.

The value of these observations is that they turn a general complaint into information the specialist can use; the diagnosis itself remains the specialist's responsibility.

## Decision pathways

Figures 11.5, 11.6 and 11.7 present the decision pathway for each of the three main categories. A single principle runs through them: medical or audiological measurement identifies the impairment, educational assessment determines its impact, and the educational category is assigned only when an impact on learning has been established. In the pathway for communication disorders, a hearing examination comes at the second step as a prerequisite for any language instrument.

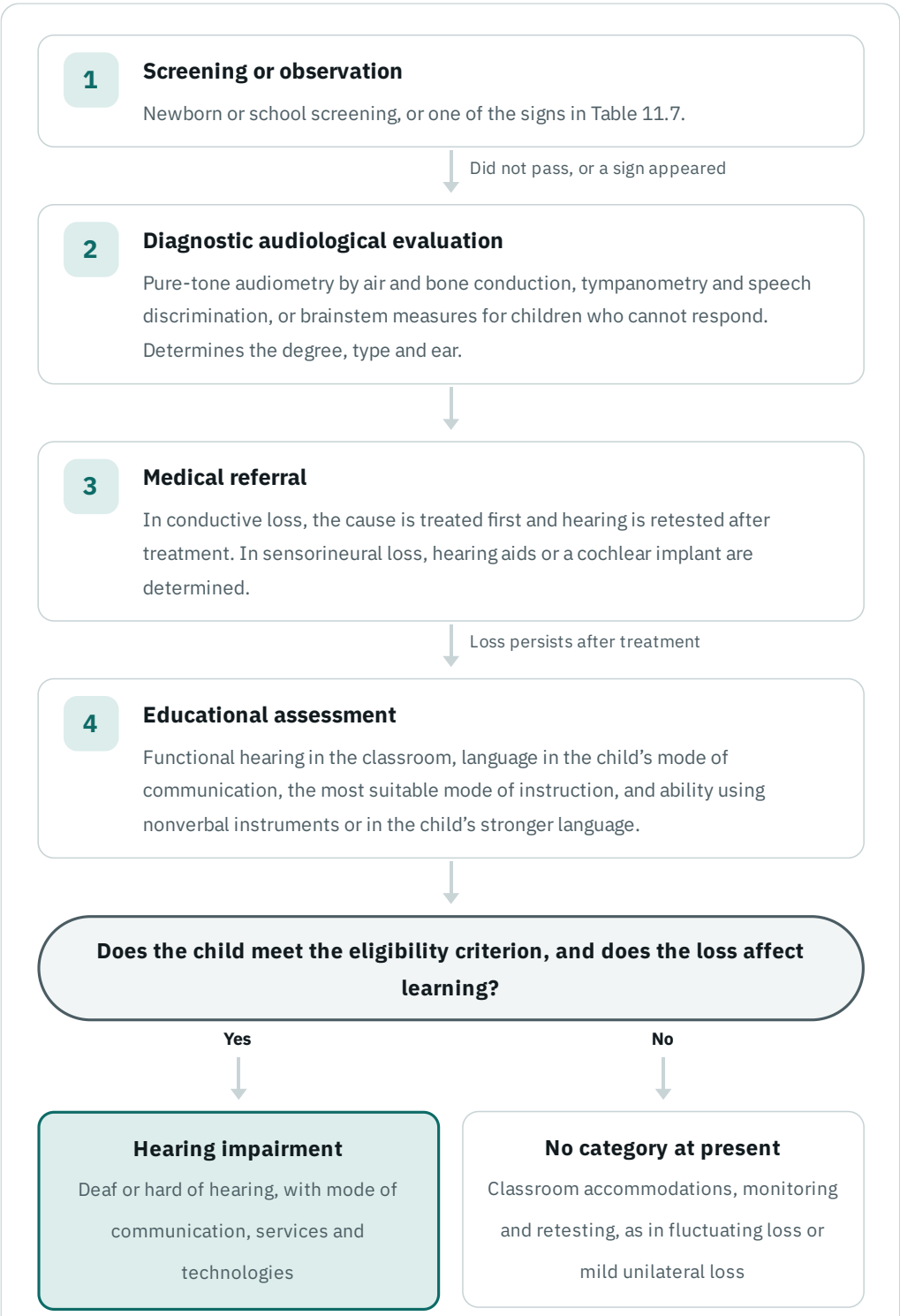

1
Screening or observation
Newborn or school screening, or one of the signs in Table 11.7.
Did not pass, or a sign appeared
2
Diagnostic audiological evaluation
Pure-tone audiometry by air and bone conduction, tympanometry and speech discrimination, or brainstem measures for children who cannot respond. Determines the degree, type and ear.
3
Medical referral
In conductive loss, the cause is treated first and hearing is retested after treatment. In sensorineural loss, hearing aids or a cochlear implant are determined.
Loss persists after treatment
4
Educational assessment
Functional hearing in the classroom, language in the child's mode of communication, the most suitable mode of instruction, and ability using nonverbal instruments or in the child's stronger language.
Does the child meet the eligibility criterion, and does the loss affect learning?
Yes
No
Hearing impairment
Deaf or hard of hearing, with mode of communication, services and technologies
No category at present
Classroom accommodations, monitoring and retesting, as in fluctuating loss or mild unilateral loss

**Figure 11.5:** Decision pathway for hearing impairment. Under Saudi regulations, the eligibility criterion is measured in the better ear with hearing aids (see 11.12).

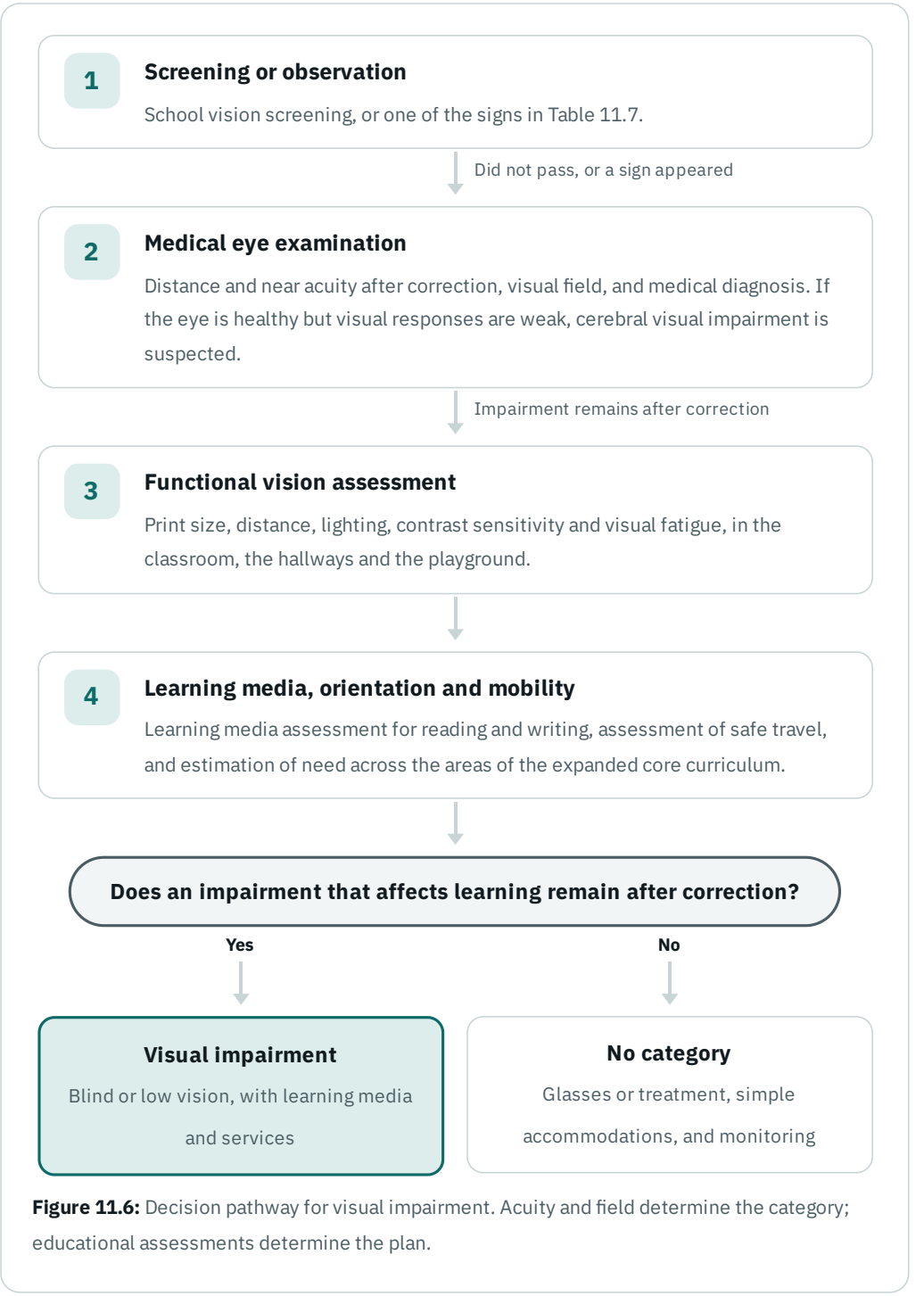



**Figure 11.6:** Decision pathway for visual impairment. Acuity and field determine the category; educational assessments determine the plan.

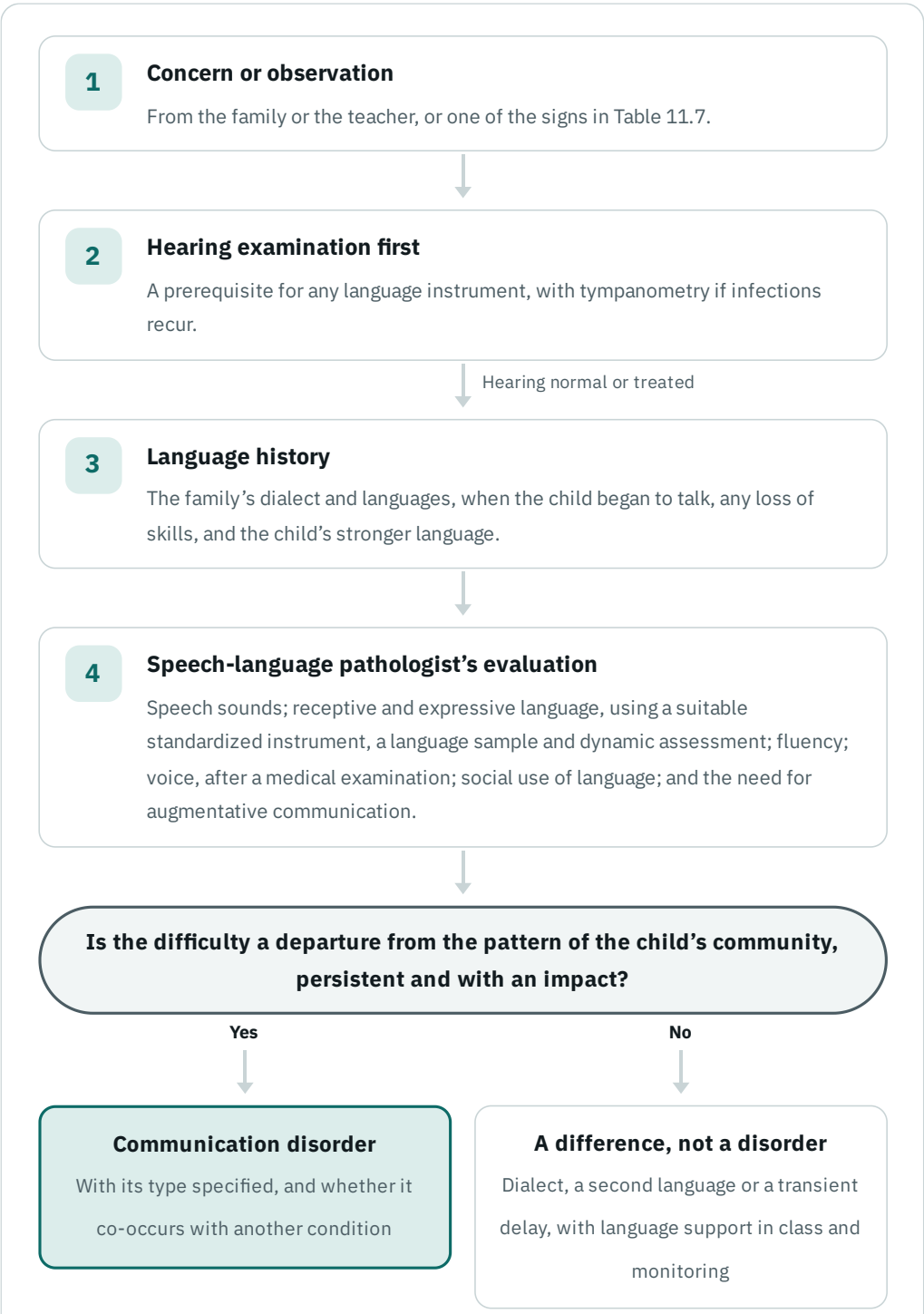

1
Concern or observation
From the family or the teacher, or one of the signs in Table 11.7.
2
Hearing examination first
A prerequisite for any language instrument, with tympanometry if infections recur.
Hearing normal or treated
3
Language history
The family's dialect and languages, when the child began to talk, any loss of skills, and the child's stronger language.
4
Speech-language pathologist's evaluation
Speech sounds; receptive and expressive language, using a suitable standardized instrument, a language sample and dynamic assessment; fluency; voice, after a medical examination; social use of language; and the need for augmentative communication.
Is the difficulty a departure from the pattern of the child's community, persistent and with an impact?
Yes
No
Communication disorder
With its type specified, and whether it co-occurs with another condition
A difference, not a disorder
Dialect, a second language or a transient delay, with language support in class and monitoring

**Figure 11.7:** Decision pathway for communication disorders. The hearing examination at the second step is never skipped, however obvious the cause may seem.

## 11.9 Instrument profiles

The three profiles present core procedures in these categories: audiometry, functional vision assessment together with learning media assessment, and a comprehensive language test. The first two are professional procedures without published norms, so some fields in their tables differ from those in the test profiles. Where we were unable to verify a figure in the technical manual itself, we have written "in the technical manual." This book does not publish the items of any test; the examples were prepared by the author.

## PTA · OAE · ABR Audiometry: pure tones, tympanometry and objective measures

Hearing impairment

### The idea

Pure-tone audiometry (**Pure-tone audiometry, PTA**) determines the softest sound the child can hear at each frequency, by air and bone conduction, and thus yields the degree and type of loss. Tympanometry shows whether the eardrum moves normally, and so indicates fluid behind it. Otoacoustic emissions and the auditory brainstem response, for their part, are **objective** measures that require no response from the child; they are used with newborns and with children who cannot respond.

### What it looks like

A sound-treated booth, an audiometer, headphones, and a bone-conduction vibrator. The response method changes with age: an older infant is rewarded with a lighted toy when he turns toward the sound, a preschooler drops a block into a basket each time he hears a sound, and an older child raises his hand or presses a button. In tympanometry, a small probe is placed at the entrance of the ear canal for a few seconds.

### A scene from administration

The audiologist tells a six-year-old: "Every time you hear the beep, even if it's very soft, put the bead in the box." She begins with a clearly audible tone at 1000 Hz, then lowers it by 10 dB each time he responds and raises it by 5 dB each time he does not, until she reaches the softest sound he responds to on most trials. She notices that the child has begun dropping the bead whenever she looks at him, responding to her glance rather than to the sound, so she varies the timing of the presentations to keep him from guessing.

**Table 11.8: Profile of pure-tone audiometry, tympanometry and objective measures.**

| | |
|---|---|
| **Purpose** | To determine the presence, degree, type and ear of hearing loss, and to monitor it |
| **Ages** | Objective measures from birth; behavioral testing with age-appropriate methods from about 6 months, and by the conventional method from about age 5 |
| **Edition and year** | A standard clinical procedure governed by professional guidelines (ASHA, 1997) |
| **Publisher** | Not applicable |
| **Qualification level** | A licensed audiologist; school screening is conducted by trained personnel under professional supervision |
| **Time** | About half an hour to an hour for a full evaluation; longer with young children |

| | |
|---|---|
| **Score types** | Thresholds in decibels for each frequency and ear, the pure-tone average, the word recognition percentage, and the tympanogram type (normal; flat, usually indicating fluid; or shifted toward negative pressure) |
| **Cutoff scores** | 20 dB for screening; degrees of loss according to the adopted classification (Tables 11.1 and 11.2); educational eligibility limits according to the regulations (see 11.12) |
| **Reliability** | Thresholds typically vary within 5 dB between two successive tests; reliability is lower with young children and when the child cooperates poorly |
| **Classification accuracy** | High for determining degree and type in a cooperative child; a single test does not detect fluctuating loss |
| **Year of norm collection** | Not applicable; the child is compared with reference thresholds for normal hearing |
| **Arabic version** | Pure-tone testing requires no language. Speech discrimination testing, however, requires equivalent Arabic word lists; at the time of writing we had not verified any recent, published, standardized Saudi lists, so the lists used and their source should be ascertained |

### How are the results interpreted?

The air-conduction marks are read to determine the degree, then compared with bone conduction to determine the type: a gap between the two with normal bone conduction indicates conductive loss, whereas the two converging at a reduced level indicates sensorineural loss. The tympanogram is then read: a flat tympanogram with conductive loss suggests fluid behind the eardrum. Finally, the thresholds are compared with the speech banana to see what the child is missing, and the word recognition score is compared with what the audiogram predicts; if it is much worse, auditory neuropathy spectrum disorder is considered.

### Its limits

It measures what the child hears in a quiet room; how well he understands in a noisy classroom is another matter. It gives a picture of a single day and therefore does not detect fluctuation unless testing is repeated. Behavioral testing depends on the child's cooperation and attention, so it may be inaccurate with young children or children with multiple disabilities, in which case objective measures are decisive.

## FVA · LMA Functional vision assessment and learning media assessment

Visual impairment

### The idea

The medical report answers the question: what is wrong with the eye? These two assessments answer two other questions: how does the student use his vision in his daily life, and by what means does he read and write? They require no norms, because the student is compared with the demands of his environment rather than with his peers (Koenig & Holbrook, 1995).

### What it looks like

A collection of tools and situations rather than a single test booklet: near and distance acuity charts, including LEA Symbols and HOTV charts for young children; contrast sensitivity charts; texts in graded print sizes; light sources; and observation checklists for the classroom, hallway and playground. In learning media assessment, reading rate and accuracy are recorded in regular print, in large print, and in braille for students who are learning it, and the way the student explores new objects is observed.

### A scene from administration

A teacher gives a third-grade student three cards bearing the same text in three print sizes, asks her to read each for one minute, and records the words read correctly. She then repeats this with lighting from the side. Next, she

accompanies the student at recess and notes where she stops, where she touches the wall, and whether she recognizes her friends from a distance or only after they speak.

**Table 11.9: Profile of functional vision assessment and learning media assessment.**

| | |
|---|---|
| **Purpose** | To determine the student's visual needs for learning, reading and writing media, and accommodations |
| **Ages** | All ages; LEA Symbols and HOTV charts are used with young children from about age three |
| **Edition and year** | Multiple professional procedures; learning media assessment has a published guide (Koenig & Holbrook, 1995) |
| **Publisher** | Not applicable to the procedure; the learning media assessment guide is published by the Texas School for the Blind and Visually Impaired |
| **Qualification level** | A teacher of students with visual impairments, together with an orientation and mobility specialist |
| **Time** | Several sessions in different settings; learning media assessment is repeated annually or whenever vision changes |
| **Score types** | Descriptive: the most suitable print size, distance and lighting; reading rate and accuracy in each medium; and the duration of visual work before fatigue |
| **Cutoff scores** | None; the decision rests on comparing efficiency across media and against classroom demands |
| **Reliability** | Depends on the observer's expertise and the range of settings; it is therefore repeated on different days |
| **Classification accuracy** | Not designed for classification; the category is determined by acuity and field |

| | |
|---|---|
| **Year of norm collection** | Not applicable |
| **Arabic version** | LEA Symbols do not depend on language. At the time of writing we know of no published Arabic guide to learning media assessment, so graded Arabic texts are prepared locally, bearing in mind that equivalent print size in Arabic script differs from that in Latin script |

### How are the results interpreted?

The aim is to find the medium that gives the student the best combination of accuracy, speed and comfort over sustained work. A student who reads large print at an acceptable rate without tiring has large print as her medium. A student whose visual condition is deteriorating, or whose reading rate plateaus at a level insufficient for classroom demands no matter how much the print is enlarged, may need braille as a primary or supplementary medium, even if he has some vision.

### Its limits

There are no norms against which to compare the student, so its quality depends on the quality of observation. It is affected by fatigue, lighting and the day of assessment. It does not replace a medical examination, and on its own it is insufficient for cerebral visual impairment, which requires a functional assessment tailored to its patterns.

## CELF-5 Clinical Evaluation of Language Fundamentals, Fifth Edition

Language

### The idea

A comprehensive language battery that measures receptive and expressive language in vocabulary, sentence structure and language memory. It yields a Core Language Score, which answers the first question: does the child have a language disorder? It then provides detailed indexes that answer the second question: in which aspects of language does the difficulty lie? (Wiig et al., 2013).

### What it looks like

An easel that presents pictures to the child and carries the instructions for the examiner, a record form, and a rating scale booklet. Its task types include pointing to the picture that matches a sentence the child hears, following increasingly long directions, completing a sentence about a picture, and repeating sentences accurately. It includes a pragmatics rating form completed by the examiner and another completed by a teacher or parent.

### A scene from administration

The examiner says to a seven-year-old: "Touch the cat after you touch the big ball." The girl touches the cat first. The examiner records the error and notes that the girl carried out all the short directions correctly but erred whenever a direction involved temporal order or a condition.

**Table 11.10: Profile of the Clinical Evaluation of Language Fundamentals (CELF-5).**

| | |
|---|---|
| **Purpose** | To determine the presence of a language disorder and describe language strengths and weaknesses |
| **Ages** | 5 to 21 years 11 months |
| **Edition and year** | Fifth, 2013 |
| **Publisher** | Pearson |
| **Qualification level** | A speech-language pathologist or qualified examiner, in accordance with the publisher's requirements |

| | |
|---|---|
| **Time** | About 30 to 45 minutes for the core tests; longer for a full evaluation |
| **Score types** | Scaled scores for subtests (mean 10, SD 3), and a Core Language Score and indexes for receptive language, expressive language and others (mean 100, SD 15) |
| **Cutoff scores** | A suggested cut-off score for the Core Language Score is given in the technical manual |
| **Reliability** | In the technical manual |
| **Classification accuracy** | In the technical manual |
| **Year of norm collection** | In the technical manual |
| **Arabic version** | We know of no published standardized Arabic version at the time of writing. Translating its items does not produce a valid instrument, because the structures being measured are English |

### How are the results interpreted?

The Core Language Score is read first against the cut-off score; then the indexes are compared with one another: is reception weaker than expression, or the reverse? Does the difficulty lie in content and meaning or in structure? The result is then compared with a language sample from the child's natural speech; if the two pictures agree, the judgment is strengthened.

### Its limits

It is built on English and its norms are American, so it is unsuitable for Arab children except in rare cases of a child educated in English, and even then it is interpreted with caution. In the absence of a standardized Arabic alternative,

the speech-language pathologist relies on the language sample, nonstandardized tasks built on Arabic, and dynamic assessment, and states this explicitly in the report.

### Assessing ability in deaf or blind students

Deaf or blind students are sometimes referred for an assessment of intellectual ability, either for placement or because a co-occurring disability is suspected. The rule is: **ability is measured through the channel that works**, and the limits of each accommodation are stated in the report.

**Table 11.11: Options for assessing ability in deaf or blind students, and their limits.**

| Student | Option | What to watch for |
|---|---|---|
| **Deaf or hard of hearing** | The Leiter International Performance Scale, Third Edition (**Leiter-3**) (Roid et al., 2013), whose instructions are given entirely through gesture; the Universal Nonverbal Intelligence Test, Second Edition (**UNIT2**) (Bracken & McCallum, 2016); and the Nonverbal Index of the Wechsler Intelligence Scale for Children, Fifth Edition (**WISC-V**), made up of subtests that require no verbal response (Wechsler, 2014) | "Nonverbal" does not mean "language-free": a child with language deprivation may perform poorly even on nonverbal tasks. Presenting the instructions in sign language through an interpreter is a change in administration conditions that must be noted in the report. We know of no published standardized Arabic versions of the first two scales |
| **Blind or severely visually impaired** | Verbal tests presented orally, such as the verbal comprehension and auditory working memory subtests of intelligence scales | They measure one aspect of ability, so no full-scale score is computed from them. Some items assume visual experience, such as questions about the colors of objects, so each item is reviewed |

| Student | Option | What to watch for |
|---|---|---|
| **Low vision, reading enlarged print** | Enlarging or converting visual materials | Enlargement may change an item's difficulty and timing, and braille is slower than print; if the construct changes, the result is interpreted descriptively, apart from the norms (see 11.7) |

This matter is all the more important in the Kingdom, because the Regulations require, for admission of a deaf or hard-of-hearing pupil to programs for these students, an IQ of no less than 75 on the Wechsler scale or 73 on the Stanford-Binet (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), without specifying which edition or which part is to be administered. The known Saudi standardization of the Wechsler scale is of its older revised edition (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024), and two Gulf versions of the Stanford-Binet Fifth Edition are in use (Mohammed, 2016 [Arabic]); we found no published study of the validity of any of them with deaf students. The score on which the admission decision rests should therefore be computed from the nonverbal part, and read alongside evidence of the student's ability in other settings.

## 11.10 When is a judgment made? Four cases

### Case 1: Khalid, age 7

**Hearing loss in one ear: no judgment about attention for now**

**The story:** Khalid is the student in the first letter. His second-grade teacher noticed that he drifted off during lessons, followed spoken instructions only after they were repeated, and watched his classmates to copy them, and that he was "present" in some lessons and "absent" in others. The teacher suspected an attention problem. When the specialist requested teacher ratings, she noticed that the complaints were strongest from the Arabic teacher, who explains while standing by the window to Khalid's left, and weakest from the mathematics teacher, who stands at the board to his right, writes a great deal and explains with visual examples. In a quiet one-to-one session, with the specialist seated facing him, his attention was good and his questions were to the point. His mother reported that he turns his head when spoken to, always holds the phone to his right ear, and cannot tell where someone is calling him from in the market. Khalid had passed newborn hearing screening and had never been screened at school. The team therefore postponed the attention assessment and requested hearing screening first.

**Table 11.12: Evidence in Khalid's case.**

| Source | Result |
|---|---|
| **School screening** | Passed in the right ear; heard none of the tones in the left ear, including on retest |
| **Audiological evaluation** | As in Figure 11.1: the right ear is normal, and the left shows a moderate-to-severe sensorineural loss. Tympanograms are normal in both ears, so there is no fluid. Speech discrimination in noise is clearly weaker when speech comes from his left side |
| **Ear, nose and throat physician** | No medically treatable cause. The loss most likely developed after birth, since he passed newborn screening. Annual testing is requested, because some unilateral losses change over time |

| Source | Result |
|---|---|
| **Classroom observation** | He sits near the window, with the air conditioner on his right and the Arabic teacher on his left. He often turns to the classmate beside him after spoken instructions |
| **Attention ratings** | High from the Arabic teacher, low from the mathematics teacher, and few concerns at home |
| **Speech-language pathologist (brief screening)** | His vocabulary and sentence structure are age-appropriate. His spelling errors cluster in soft, high-pitched sounds such as s, f and th |

**Decision:** No judgment of ADHD is made for now. Nor does Khalid meet the "hard of hearing" criterion in the Saudi regulations, because that criterion is measured in the **better ear**, and his right ear is normal. Still, the loss is permanent and its effect in the classroom is evident. It is worth noting here that the US legal definition of hearing impairment requires only an effect on learning, without stipulating a degree or both ears (U.S. Department of Education, 2006, 34 CFR §300.8(c)(5)). The team recommends, immediately and without waiting for any decision, that he sit at the front of the classroom with his right ear toward the teacher and away from sources of noise, that he be able to see the speaker's face, that important instructions be written down, and that his understanding be checked by asking him what he has been asked to do rather than "Do you understand?" The audiologist will weigh the benefit of a wireless hearing system (**Remote microphone system**) that carries the teacher's voice to his good ear. His attention will then be reconsidered after several months of these accommodations: if symptoms of inattention persist in two settings under adequate listening conditions, the team will proceed with an evaluation (see Chapter 10).

**The lesson of the case:** What looked like inattention had an auditory explanation. Khalid does not follow instructions because he does not hear them in full when they come from his left or through noise, and he copies his classmates to find out what is expected. The clue that alerted the team was the way his performance varied with where the speaker stood and how material was presented. A child who "hears well" in face-to-face conversation may have a severe loss in one ear that only testing each ear separately will reveal.

> **What evidence could change the decision?** If symptoms of inattention persisted at home and at school after months of suitable seating and the hearing system, the team would go ahead with an attention evaluation. If annual testing showed a decline in the right ear, his standing under the "hard of hearing" criterion and his services would both change.

## Case 2: Rahaf, age 6 Moderate low vision

**The story:** Rahaf's teacher noticed that she held the book close to her face, tilted her head and grew tired after a few minutes of reading. Her acuity was poor on the school vision screening, so she was referred to an ophthalmologist, who diagnosed a retinal disease. Her best-corrected distance acuity with glasses was 6/36 in the better eye, that is, moderate low vision, and her visual field was normal.

**Table 11.13: Results of Rahaf's educational assessments.**

| Educational assessment | Result |
|---|---|
| — | She reads comfortably in large print about three times the size of her textbook's font, at a distance of 15 centimeters, with side lighting. She cannot see the board from her seat. She struggles with faded photocopied worksheets, which indicates low contrast sensitivity |
| — | She explores her surroundings mainly through vision, reads enlarged print at a reasonable speed for her age, and does not tire of it as she does of regular print |
| **Orientation and mobility** | She moves around the school safely in daylight, hesitates on the dimly lit stairs and touches the handrail |
| **Early reading assessment** | Her errors on similar-looking letters almost disappear when the print is enlarged, so the difficulty lies in seeing, not in reading |

**Decision:** Visual impairment with moderate low vision, which falls within the "low vision" range in the Saudi regulations (6/24 to 6/60). Her primary learning medium is enlarged print, together with magnification devices, a copy of the board content on her desk and high-contrast worksheets. The stairs are to be lit, and she will be trained to move on them. Her learning medium will be reassessed annually, because her retinal condition may change.

> **What evidence could change the decision?** If her acuity declined or her field narrowed as the retinal disease progressed, or if her reading speed in large print leveled off below what her class demands, the learning medium would need to be reassessed and braille considered as a supporting medium.

## Case 3: Abdulrahman, age 6

**Sound dialect, and a pattern that warrants treatment**

**The story:** He is the child the kindergarten teacher described in the third message: he says “galam” (گلم) instead of “qalam” (قلم) and “tamaka” (تمكة) instead of “samaka” (سمكة). His hearing was tested first and found to be normal. The speech-language pathologist then evaluated him, asked the family about their own pronunciation and listened to a sample of his conversation with his mother.

**Table 11.14: Observations in Abdulrahman’s case and their interpretation.**

| Observation | Interpretation |
|---|---|
| **He pronounces /q/ as /g/ in every word** | As his family does. A dialectal difference, not a disorder |
| **He replaces s, sh and z with t or d, in every position in the word** | This does not occur in his family’s dialect, and it affects an entire class of sounds |
| **His vocabulary and sentences are age-appropriate, and his comprehension is good** | Language is intact; the problem lies in speech sounds alone |
| **Unfamiliar listeners cannot understand about a third of his words** | A clear effect on communication |
| **Hearing is normal** | Hearing loss is ruled out as a cause of this pattern |

**How the specialist read the evidence:** Studies of the acquisition of Arabic consonants indicate that mastery of some fricatives comes late in Arabic and may extend beyond school entry (Amayreh & Dyson, 1998; Amayreh, 2003). At the time of writing, we found no published Arabic norm specifying the age at which stopping disappears. Had

Abdulrahman's error involved a single late-mastered sound, watchful waiting and follow-up would have been the better course. What tipped the judgment was three things taken together: the substitution covers an entire class of sounds, it reduces his intelligibility to unfamiliar listeners at an age when he is expected to be understood, and his family's dialect does not explain it.

**Decision:** Speech sound disorder, manifested in the replacement of fricatives with plosives. Speech therapy sessions are recommended. His pronunciation of /q/ is not treated; he will learn Standard Arabic at school as his classmates do.

> **What evidence could change the decision?** If samples of his speech in other settings showed that the substitution is limited to one or two late-mastered sounds, or that unfamiliar listeners understand most of what he says, monitoring would be preferable to treatment.

## Case 4: Mashael, age 14 Deaf, with no additional learning disability

**The story:** Mashael is in the second year of middle school. She has been deaf since birth, with a profound bilateral sensorineural loss, and wears two hearing aids of limited benefit. Her parents are deaf, so she grew up with Saudi Sign Language from birth. This year she moved to a new school, and her Arabic teacher noticed that her reading and writing were at roughly a fourth-grade level. The teacher asked: does she have a learning disability, or low intellectual ability?

**Table 11.15: Evidence in Mashael's case.**

| Source | Result |
|---|---|
| **Language history** | Full exposure to sign language from birth; she learned written Arabic at school as a second language, accessed through reading |
| **Nonverbal intelligence test with gestural instructions, administered by a psychologist with a teacher fluent in sign language present** | Average (a score of 108 on American norms, read with caution in the absence of Arabic norms) |
| **Sign language narrative sample, evaluated by a deaf teacher fluent in the language** | She retold a picture story in precise sequence and rich detail, using advanced signed constructions, at a level appropriate for her age |
| **Mathematics** | At grade level in arithmetic, weaker on written word problems |
| **Arabic writing** | Errors in attached pronouns, prepositions and agreement, which are common among learners of Arabic as a second language. Her vocabulary is good in the topics she has been taught |
| **Previous school's report** | Her reading improved markedly when the teacher explained texts to her in sign language before she read them |

**How the team read the evidence:** Had Mashael been assessed with a verbal intelligence test in Arabic, her score would have come out low, and the team would have concluded that she has an intellectual disability. Had her language development been judged by her Arabic writing alone, they would have concluded that she has a language

disorder. Yet her first language is intact and rich, her nonverbal ability is average, and her weakness is confined to her second language and improves when the text is explained to her in her first. This is the pattern of a second-language learner, not of a learning disability.

**Decision:** Mashael remains in the hearing impairment category, with no added diagnosis of learning disability or intellectual disability. Explicit instruction in Arabic reading and writing that builds on sign language is recommended, along with test instructions delivered in sign language and an interpreter or a teacher fluent in it during lessons. Should she show difficulty in sign language itself, or in arithmetic despite full access to instruction, that would be grounds for reconsidering a co-occurring learning disability.

**The lesson of the case:** A deaf student's language development is measured in the student's stronger language. Delay in written Arabic alone proves neither a learning disability nor any other disability, because it is to be expected of anyone who learns Arabic as a second language without hearing it.

> **What evidence could change the decision?** If her Arabic reading made no meaningful progress after a sufficient period of explicit instruction building on sign language, or if a second signed narrative sample revealed a weakness absent from the first, the team would reconsider a co-occurring learning disability.

# 11.11 Differential diagnosis and co-occurrence

## Similar conditions

Sensory disabilities and communication disorders resemble many categories, and many categories resemble them. Table 11.16 brings together the conditions that recur most often in referrals, and the evidence that distinguishes them.

**Table 11.16: Differential diagnosis in sensory disabilities and communication disorders.**

| What it looks like | What it may be | Distinguishing evidence |
|---|---|---|
| **Inattention** | Hearing loss, including fluctuating and unilateral loss | Performance improves when the student can see the speaker's face or the instruction is written, and fluctuates with bouts of ear infection. Hearing is tested before any attention measure |
| **Difficulty in reading or spelling** | Low vision or hearing loss | Visual errors disappear when print is enlarged; auditory errors concentrate on the sounds the child cannot hear |
| **Autism spectrum disorder** | Hearing loss, deafblindness or developmental language disorder | Intact nonverbal communication and a desire to interact, with improvement once the sensory barrier is removed. Restricted and repetitive patterns are absent in developmental language disorder |
| **Intellectual disability** | Sensory or motor disability, language deprivation, developmental language disorder or brain injury | Performance on a tool suited to the functioning channel, adaptive behavior, nonverbal ability and pre-injury performance |

| What it looks like | What it may be | Distinguishing evidence |
|---|---|---|
| **Central auditory processing disorder** | Inattention, developmental language disorder or undetected hearing loss | Full audiological evaluation, followed by assessment of language and attention, before this explanation is accepted (see 11.2) |
| **Language disorder** | Dialectal difference or second-language learning | Judgment against the standard of the child's dialect, assessment of the child's other language, and dynamic assessment |
| **Communication disorder (does not speak at school)** | Selective mutism | Speaks fluently at home, as family recordings confirm. The diagnostic manual classifies it among the anxiety disorders |
| **Low vision originating in the eye** | Cerebral visual impairment | The eye examination is normal or does not explain the impairment, with distinctive patterns on functional assessment |

## Co-occurring condition

Many causes of deafness and blindness, such as extremely preterm birth, certain infections before and after birth, and genetic syndromes, also affect the brain, so other disabilities occur alongside the sensory disability in the same child. Surveys report that a considerable proportion of deaf and hard-of-hearing children have an additional disability, and that visual impairment is common among children with cerebral palsy and multiple disabilities, although estimates vary with samples and definitions. Developmental language disorder, for its part, frequently co-occurs with dyslexia, ADHD and developmental coordination disorder. Each combination carries its own guidance for assessment:

- **Sensory disability with suspected intellectual disability:** A judgment of intellectual disability requires both a tool suited to the functioning channel and adaptive behavior that supports it; where this is not possible, the judgment is deferred, as in Salman's case.
- **Hearing impairment with a suspected learning disability:** Evidence is sought from outside the second language, such as difficulty in sign language itself, or in arithmetic despite full access to instruction, as in Mashael's case.
- **Sensory disability with suspected autism:** The sensory loss is addressed first, with a hearing aid, glasses or accommodations, and the social symptoms are then reconsidered. Any autism measure used with a deaf or blind child is interpreted with great caution, because its items assume intact hearing and vision.
- **Developmental language disorder with reading difficulty:** Listening comprehension and decoding are each measured separately; if both are weak, the two problems co-occur.
- **Developmental language disorder with inattention:** Instructions for attention measures are given in simple language, and the team observes whether the student's attention also falters on nonlinguistic tasks; if it falters only on language tasks, language is the more likely explanation.
- **Stuttering with social anxiety,** especially in adolescents: the students themselves are asked about the situations they avoid and how they feel in them, because avoidance may disrupt their lives more than the stuttering itself.

## 11.12 In the Saudi and Arab context

**In the Saudi context: How are this chapter's categories defined?**

The Special Education Organizational Guide adopts eleven categories, six of which concern this chapter: hearing impairment, visual impairment, language and speech disorders, multiple impairments, physical and health impairments, and deafblindness (Ministry of Education, 1436 AH [Arabic]). The Regulations for Special Education Institutes and Programs set the levels of hearing and vision on which admission is based (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). Table 11.17 compares these regulations with the two diagnostic classifications and with US law.

**Table 11.17: This chapter's categories in the diagnostic classifications, US law and the Saudi regulations.**

| Category | DSM-5-TR, ICD-11 and the World Health Organization | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Hearing impairment** | Not a disorder in the diagnostic manual. The World Health Organization grades severity by the four-frequency average in the better ear, with loss beginning at 20 decibels | Two categories: deafness, when the child cannot process linguistic information through hearing even with amplification; and hearing impairment, whether permanent or fluctuating. No numbers are given, and an effect on educational performance is required | Deaf: a loss of 70 decibels or more; hard of hearing: 35 to 69 decibels; in the better ear after the use of hearing aids, as reported by an accredited specialist body. Admission to their programs requires an IQ of no less than 75 on the Wechsler scale or 73 on the Stanford-Binet (the Regulations, Article 4) |

| Category | DSM-5-TR, ICD-11 and the World Health Organization | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Visual impairment** | Categories of acuity in the better eye; blindness is worse than 3/60, or a field no more than 10 degrees in radius | An impairment in vision that, even with correction, affects educational performance, including partial sight and blindness. No numbers are given | Blind: acuity below 6/60 in the better eye after correction, or a visual field of less than 20 degrees; low vision: 6/24 to 6/60 (the Regulations; the Organizational Guide) |
| **Communication disorders** | Four disorders in DSM-5-TR; in ICD-11, developmental language disorder, with pragmatic impairment as one of its subtypes | Speech or language impairment: stuttering, impaired articulation, a language impairment or a voice impairment | A category of "language and speech disorders." In the sources we were able to verify, we found no published eligibility criteria for it |

| Category | DSM-5-TR, ICD-11 and the World Health Organization | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Multiple, physical and health impairments, and brain injury** | No umbrella category; each condition is diagnosed on its own, and ICD-11 allows a "provisional" category of intellectual disability when assessment is not feasible | Five separate categories: deaf-blindness, multiple disabilities, orthopedic impairment, other health impairment and traumatic brain injury | Three categories: multiple impairments, deafblindness, and physical and health impairments. There is no separate category for brain injury. We have not verified the text of their definitions and detailed criteria |
| **Who decides** | The audiologist and ophthalmologist for measurement, and the speech-language pathologist for communication | A team of qualified professionals together with the parents | The school assessment team, after the guardian's written consent, together with a report from the specialist body for hearing and vision (the Organizational Guide) |

| Category | DSM-5-TR, ICD-11 and the World Health Organization | US law (IDEA) | Saudi regulations |
|---|---|---|---|
| **Reevaluation** | Depending on the condition | At least every three years, unless the parents and the agency agree otherwise | The diagnosis is repeated no sooner than 6 months and no later than 18 months after the initial diagnosis, and thereafter as needed (the Organizational Guide) |

Two points emerge from the table. First, the Saudi regulations classify a student as educationally blind within wider limits than the international classification: a student with an acuity of 4/60 has "severe low vision" under the international classification but is "blind" under the regulations. This makes educational sense, since students in this range usually need the media used by blind learners. Second, the hearing criterion is measured with hearing aids, so it captures the disability that remains after intervention, whereas the audiologist measures the underlying loss unaided. The team must therefore be sure which of the two measurements a report refers to. The Law on the Rights of Persons with Disabilities affirms the right to "modern measurement and diagnostic tools" and to early intervention programs and early detection of disability (Law on the Rights of Persons with Disabilities, 1445 AH [Arabic], Arts. 8 and 9).

**Newborn hearing screening in the Kingdom:** The National Newborn Hearing Screening Program has operated since 2016. In 2021, 199,034 newborns were screened, a coverage rate of 92.6%, and 1.87% did not pass

and were referred (Alothman et al., 2024). A systematic review of thirteen Saudi studies identified six challenges facing the program, among them low awareness, loss of some children to follow-up after referral, and weak data-entry and tracking systems (Alanazi & Nicholson, 2025). For the school team, this means that a child's having been screened at birth does not always mean that diagnostic follow-up took place if the child did not pass, so the question should be asked explicitly. At the time of writing, we found no published official Ministry of Health page on the program that we could consult, and no published text establishing a national hearing and vision screening program at school entry or its protocol.

**A question for reflection:** Khalid has a permanent, moderate-to-severe sensorineural loss in one ear, and his other ear is normal. Does he meet the "hard of hearing" criterion in the Saudi regulations? If not, where will he obtain the accommodations he needs? And should unilateral loss be named explicitly in the regulations, as the World Health Organization's classification does?

## Available Arabic instruments

Table 11.18 gathers the Arabic instruments and studies in this chapter's areas that we found published at the time of writing. It makes no claim to completeness, nor does it replace verifying the licensing and norms of any instrument before it is used in a decision.

**Table 11.18: Published Arabic instruments and studies in sensory disabilities and communication.**

| Area | Study or instrument | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| **Acquisition of Arabic consonants** | Amayreh & Dyson (1998) | Jordan; preschool children | Order of consonant acquisition and approximate ages of mastery | A descriptive study, not a standardized test, and the sample's dialect is Jordanian |
| **Completing the consonant inventory** | Amayreh (2003) | Jordan; older children | The ages by which the late sounds are mastered | A descriptive study, and the sample's dialect is Jordanian |
| **Intelligence with a nonverbal component** | Two Gulf versions of the Stanford-Binet Intelligence Scales, Fifth Edition (Mohammed, 2016 [Arabic]) | Used in the Kingdom; we have not verified the two norming samples | Half of the scale is nonverbal | We found no published study of their validity with deaf students |
| **Purely nonverbal intelligence** | At the time of writing, we know of no published standardized Arabic version of the Leiter scale or the Comprehensive Test of Nonverbal Intelligence | | | |
| **Comprehensive language tests** | At the time of writing, we know of no published standardized Arabic version of the Clinical Evaluation of Language Fundamentals or any equivalent | | | |

| Area | Study or instrument | Country and sample | What it offers | Limitations |
|---|---|---|---|---|
| **Saudi Sign Language** | At the time of writing, we know of no published standardized instrument for assessing it | | | |
| **Speech discrimination testing** | We have not verified any recent published standardized Saudi word lists | | | |
| **Learning media assessment for blind students** | At the time of writing, we know of no published Arabic guide | | | |

The gap is clear: most of what the team needs in these categories, from a standardized Arabic language test, to a tool for assessing Saudi Sign Language, to Arabic norms for sound acquisition in the dialects of the Kingdom, is not available in published form. Until that gap is filled, assessment relies on language samples, dynamic assessment and tools that do not depend on language, with every report stating the source and limits of each piece of evidence.

## Conclusion

The central argument we have sought to develop in this chapter is that the number produced by a device or a medical examination is where educational judgment begins, not where it ends. The audiogram and vision measurements establish the degree and type of loss, but what this means for the student's learning calls for functional assessment and learning media assessment, and a deaf student's language development is measured in the student's stronger language. It follows that hearing and vision screening should precede any diagnostic tool; that the team should watch for forms that are easily missed, such as fluctuating and unilateral loss and auditory neuropathy; that neither

dialect nor second-language learning should be treated as a disorder; and that the report should name the classification it has adopted for degrees of hearing and vision.

The hardest judgments we have discussed remain those in multiple disabilities and deafblindness, where standardized tools usually fall away. Judgment then rests on repeated functional and communicative assessment in natural settings, and no degree of intellectual disability is assigned when it cannot be validly measured. The line between an accommodation that preserves the norms and one that invalidates them remains a matter of judgment in each case: what changes the mode of access usually preserves them, and what changes the construct invalidates them, in which case the result is reported descriptively.

## 11.13 Summary, key terms and activities

### Chapter summary

- Hearing and vision screening is the gateway of exclusion in every diagnosis and precedes any diagnostic tool, because impairment in either can resemble any other category.
- The audiogram shows the softest sound a child hears at each frequency, and comparing air and bone conduction determines the type of loss. Fluctuating and unilateral loss and auditory neuropathy are forms the team can easily miss.
- Classifications of degrees of hearing and vision differ across sources, and between the international classification and the Saudi regulations, so the report names the classification it has adopted.
- Vision is measured through distance and near acuity, visual field and contrast sensitivity, complemented by functional vision assessment,

learning media assessment, orientation and mobility, and the areas of the expanded core curriculum.

- Communication disorders are of five types, including voice disorders, and are assessed by the speech-language pathologist after hearing has been tested. Developmental language disorder does not require a discrepancy with nonverbal intelligence.
- Dialect is not a disorder, second-language learning is not a language disorder, and a deaf student's language development is measured in the student's stronger language.
- In multiple disabilities and deafblindness, standardized tools usually fall away, so judgment rests on repeated functional and communicative assessment in natural settings, and no degree of intellectual disability is assigned when it cannot be validly measured.
- An accommodation that changes the mode of access usually preserves the norms; one that changes the construct invalidates them, and the result is then reported descriptively.

## Glossary of terms

| | |
|---|---|
| **Audiogram** | A chart showing the softest sound a person hears at different frequencies in each ear. |
| **Decibel** | The unit for measuring sound intensity; the higher the value, the louder the sound. |
| **Bone conduction** | Transmission of sound to the inner ear through the bones of the skull, bypassing the outer and middle ear. |

| | |
|---|---|
| **Conductive hearing loss** | Loss arising in the outer or middle ear, shown by a gap between air and bone conduction; it may be temporary or fluctuating. |
| **Sensorineural hearing loss** | Loss arising in the inner ear or the auditory nerve, usually permanent. |
| **Auditory neuropathy spectrum disorder** | A disorder in which the inner ear functions but the signal is not transmitted to the brain in an orderly way, so speech understanding is poorer than the audiogram would predict. |
| **Tympanometry** | A measurement of eardrum movement; a flat tracing usually indicates fluid behind the eardrum. |
| **Saudi Sign Language** | A complete natural language used by deaf people in the Kingdom, with its own vocabulary and grammar. |
| **Language deprivation** | A child's reaching an advanced age without a fluent language, spoken or signed, because full access to a language was never provided. |
| **Visual acuity** | Clarity of sight, expressed as a fraction such as 6/6 and measured at distance and near. |
| **Contrast sensitivity** | The ability to distinguish an object from its background when their colors are close. |
| **Visual field** | The area a person can see without moving the eyes. |
| **Functional vision assessment** | An assessment of how a student actually uses vision in learning situations and daily life. |

| | |
|---|---|
| **Learning media assessment** | An assessment that determines the most suitable medium for a student with low vision or blindness to learn reading and writing. |
| **Expanded core curriculum** | Skills that a student with visual impairment needs to be taught explicitly, in nine areas, including braille, orientation and mobility, and independent living. |
| **Cerebral visual impairment** | Impairment in the brain's processing of what the eye sees, with the eye itself intact or nearly so. |
| **Developmental language disorder** | A persistent language difficulty with functional impact, not associated with a known condition such as hearing loss, autism or intellectual disability, and not requiring a discrepancy with nonverbal intelligence. |
| **Speech sound disorder** | Persistent difficulty producing speech sounds that is not appropriate for age and interferes with the intelligibility of the child's speech. |
| **Stopping** | Replacing a fricative sound, such as s, with a plosive sound, such as t. |
| **Voice disorders** | A departure in the quality, pitch, loudness or resonance of the voice from what suits the speaker's age and sex. |
| **Augmentative and alternative communication** | Means that supplement or replace speech, such as signs, pictures and speech-generating devices. |

| | |
|---|---|
| **Tactile communication** | Communication that relies on touch, such as signs the child feels with a hand placed over the speaker's hand. |
| **Pre-symbolic communication** | A stage at which the child communicates through body movements, facial expressions and actions, before words, signs and pictures. |
| **Dialectal difference** | Pronunciation that matches the child's dialect and community; it is not considered a disorder even when it differs from Standard Arabic. |

Dynamic assessment, accommodation, cut-off score and norming sample are terms defined in Chapters 2, 3 and 5.

## Review questions

1. On an audiogram, the air-conduction marks for both ears fall at about 35 decibels across all frequencies, and the bone-conduction marks at 5 decibels. What are the degree and type of loss? What does the child miss in speech? How does this differ from the left ear in Khalid's audiogram, and from its effect in the classroom? **(Objective 1)**
2. A child's average threshold in the better ear is 38 decibels. Classify the loss using Clark's classification and the World Health Organization's classification, then state whether the child meets the "hard of hearing" criterion in the Saudi regulations, and what information is missing to decide. **(Objectives 1 and 2)**
3. A student has an acuity of 6/9 but bumps into objects, cannot see the whole board and struggles with faded worksheets. Which two measurements would you request? And in which category might the student fall if the field narrowed to 15 degrees, under the Saudi regulations and the international classification? **(Objective 2)**

4. Why is the ophthalmologist's report not enough to plan the education of a student with low vision? Name three things that educational assessments add. **(Objective 2)**
5. A girl whose voice has been hoarse for three months, a boy who stutters and avoids answering orally, and a child whom only his family can understand. For each, identify the likely type of disorder, who assesses it and the first assessment step. **(Objective 3)**
6. A seven-year-old has clearly weak language, a nonverbal IQ of 84 and age-appropriate adaptive behavior outside language. Can he be diagnosed with developmental language disorder? Justify your answer using what the CATALISE consensus concluded. **(Objective 4)**
7. A five-year-old who grew up speaking Urdu joined an Arabic-medium kindergarten six months ago and makes frequent errors in Arabic sentence structure. How would you distinguish normal language learning from a language disorder? **(Objective 4)**
8. A student with cerebral palsy responds on an intelligence test by looking at the answer choice instead of pointing with her finger; on another test, time limits were removed for her on tasks that measure processing speed. In which of the two cases can the score be interpreted against the norms, and why? **(Objective 5)**
9. Propose a four-step assessment plan for a four-year-old girl with partial deafblindness who neither speaks nor signs. Who takes part? Where is the assessment carried out? What is recorded? **(Objective 6)**
10. Choose another special education category, such as learning disabilities or autism spectrum disorder, and show at which step in its decision pathway hearing and vision screening must appear, and what happens to the validity of the remaining tools if it is delayed. **(Objective 7)**

## Case for application

### Hadi, age 16, first year of secondary school

**The story:** Eight months ago, Hadi sustained a moderate brain injury in a car accident. He spent weeks in the hospital and then returned to school. His teachers say he is "no longer the Hadi we knew": he is slow, forgets what was explained the day before, tires before midday and is quick to anger. One teacher suggested moving him to a program for intellectual disability. With his family's consent, the team gathered the following data.

**Table 11.19: Data in Hadi's case.**

| Source | Result |
|---|---|
| **Pre-injury achievement record** | Grade averages between 85% and 92% in middle school, with no previous referral |
| **Individual intelligence test (seventh month after the injury)** | Full-scale score 74; verbal comprehension 94, reasoning 88, working memory 70, processing speed 62 |
| **Adaptive behavior scale (mother's ratings)** | Low average, with weakness in managing his time and his medications |
| **Hearing and vision** | Hearing is normal; the eye examination is normal, but he reports double vision when tired |
| **Medical report** | Marked improvement between the third and sixth months, with some further improvement expected |
| **Classroom observation** | He answers well in oral discussion early in the day, and his performance drops clearly after the fourth period |

**Tasks:** (1) Does the evidence support a diagnosis of intellectual disability? Justify your answer with at least three pieces of evidence. (2) Under which category could Hadi be served in the Saudi regulations, given that there is no separate category for brain injury? And what would you need to verify to do so? (3) When would you reassess his ability, and why? (4) Propose three accommodations that need not wait for the decision. **(Objectives 5, 6 and 7)**

## Learning activities

Any activity carried out with children or families is conducted with the consent of the family and the school administration, without any data that could identify the child, and under the supervision of a specialist, and its results are not used in any decision concerning the child.

1. **Reading an audiogram:** Draw three hypothetical audiograms for three children: a mild conductive loss, a moderate bilateral sensorineural loss and a severe sensorineural loss in one ear. Describe what each child hears in class, and propose three classroom accommodations. *Output:* three audiograms with a paragraph for each; about one hour. **(Objective 1)**
2. **Simulating a narrow field and low contrast:** Hold a paper tube in front of your eye and try to find a book on a shelf, then read a page that has been photocopied faintly. Write what you have learned about the needs of a student with a restricted field or low contrast sensitivity, noting that a simulation does not convey the real experience of disability. *Output:* one page; about 30 minutes. **(Objective 2)**
3. **Dialect checklist:** Record five minutes of a child's conversation with the family, and identify the sounds the child pronounces in the family dialect and the sounds that warrant evaluation. *Desk-based alternative:* List ten dialect features in your region that a teacher from another region might mistake for articulation errors, together with their Standard Arabic equivalents. *Output:* a table; about one hour. **(Objective 4)**

4. **A communication profile for a nonspeaking child:** Choose Salman's case or a similar hypothetical case, and write a one-page profile that answers the four augmentative and alternative communication assessment questions in 11.4 and proposes a first means of communication to try. *Output:* one page; about one hour. **(Objective 6)**
5. **Comparing systems:** Apply the Saudi regulations column of Table 11.17 to the cases of Khalid, Rahaf, Mashael and Salman, and write for each case: Would the regulations have reached the same decision? On what evidence would it rest? What does the text leave unresolved? *Output:* a table with commentary; about one hour. **(Objectives 2 and 7)**

## Chapter references

Mohammed, H. A. A. (2016). [Differences between the two Gulf versions of the Stanford-Binet Intelligence Scales, Fifth Edition, in assessing and diagnosing intelligence and mental abilities among persons with disabilities in the Northern Borders Region] [In Arabic]. *Jazan University Journal of Human Sciences, 5*(Suppl.), 236–254. https://search.mandumah.com/Record/1103082

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alanazi, A. A., & Nicholson, N. (2025). Challenges of newborn hearing screening programs in Saudi Arabia: A systematic review. *Audiology Research, 15*(2), Article 34. https://doi.org/10.3390/audiolres15020034

Alothman, N., Elbeltagy, R., & Mulla, R. (2024). Universal newborn hearing screening program in Saudi Arabia: Current insight. *Journal of Otology, 19*(1), 35–39. https://doi.org/10.1016/j.joto.2024.01.002

Amayreh, M. M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research, 46*(3), 517–529.

Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research, 41*(3), 642–653. https://doi.org/10.1044/jslhr.4103.642

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*.

American Speech-Language-Hearing Association. (2005). *(Central) auditory processing disorders* [Technical report].

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080.

Bracken, B. A., & McCallum, R. S. (2016). *Universal Nonverbal Intelligence Test* (2nd ed.). PRO-ED.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *ASHA, 23*(7), 493–500.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), 20 U.S.C. § 1400 et seq. (2004).

Joint Committee on Infant Hearing. (2019). Year 2019 position statement: Principles and guidelines for early hearing detection and intervention programs. *Journal of Early Hearing Detection and Intervention, 4*(2), 1–44.

Koenig, A. J., & Holbrook, M. C. (1995). *Learning media assessment of students with visual impairments: A resource guide for teachers* (2nd ed.). Texas School for the Blind and Visually Impaired.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257.

Riley, G. D. (2009). *Stuttering Severity Instrument* (4th ed.). PRO-ED.

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale* (3rd ed.). Stoelting.

Roman-Lantzy, C. (2018). *Cortical visual impairment: An approach to assessment and intervention* (2nd ed.). AFB Press.

U.S. Department of Education. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final rule (34 CFR Parts 300 and 301). *Federal Register, 71*(156).

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals* (5th ed.). Pearson.

World Health Organization. (2019). *World report on vision*.

World Health Organization. (2021). *World report on hearing*.

World Health Organization. (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural or neurodevelopmental disorders*.

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998). Language of early- and later-identified children with hearing loss. *Pediatrics, 102*(5), 1161–1171.

# 12 Chapter 12 Giftedness

Where the inquiry into disability categories turns to what impedes learning, the inquiry into giftedness turns to what lies beyond what the classroom offers. And because giftedness does not force itself on our attention as disability does, the greatest danger in identifying it is overlooking a gifted student whom no one noticed, a more serious error than misidentifying one.

**Scene from the teachers' room**

The fourth-grade teachers were asked to nominate students for the gifted program. Each teacher nominated her highest-scoring students: the organized ones who raise their hands and hand in their homework on time. None of them nominated Jana, because she "daydreams a lot, and her spelling grades are average," even though her teacher the previous year had thought she might be gifted.

But the science teacher remembered that two months earlier, in an open-ended project, Jana had designed an experiment to test which types of soil hold water the longest, and had recorded her results in a table that nobody had asked her for.

Is Jana gifted? Are the nominated students gifted? Perhaps, and perhaps not: high grades and good conduct are not giftedness, and daydreaming and average grades do not rule it out. In this chapter, we set out how to answer that question with evidence rather than impressions.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Compare four models of giftedness (Renzulli, Gagné, Subotnik and colleagues, and Pfeiffer) in terms of what each looks for, its tools, its threshold, and whom it may miss.
2. Determine, in a given identification program, which features are consistent with the principles of fair identification and which violate them, and propose a modification for each violation.
3. Design opportunities to display giftedness that precede nomination, and determine, in a given situation, how their absence affects teacher nomination.
4. Interpret the results of ability and creative thinking measures for a gifted student, including the difference between the General Ability Index and the Cognitive Proficiency Index, the ceiling effect, and the results of above-level testing.
5. Apply the admission criteria in Saudi regulations and the result of the Mawhiba measure to a given student, and critique them in light of the principles of identification.
6. Judge whether the evidence in a new case suffices for a determination of giftedness and the choice of service, justify that judgment, and distinguish the gifted student from the high achiever and from the student with a disorder.

# Introduction

Identifying gifted learners differs from diagnosing disability both in the direction of the question and in the nature of the costliest error. The inquiry into disability categories is directed at what impedes learning; the inquiry into giftedness is directed at what exceeds what the classroom provides. Because giftedness does not catch a teacher's attention the way disability does, the greater risk is that a gifted student will be missed, whether because he or she was never given an opportunity to show ability or because the student did not fit the picture of the organized, high-scoring pupil, as happened to Jana in the opening scene. Who is found and who is missed depends on the definition of giftedness adopted, on the instruments and their ceilings, and on the admission rules.

We open the chapter by defining giftedness and its models (Section 12.1), then present the principles of identification and the construction of local norms (Section 12.2), followed by nomination, the opportunity to display, and nomination and observation tools (Section 12.3). We then take up ability measures and their limits, the ceiling effect, and above-level testing (Section 12.4), followed by creativity, achievement, and performance (Section 12.5), instrument profiles (Section 12.6), and identification in the Saudi context and the decision pathway (Section 12.7). We apply all of this to three cases (Section 12.8) and close with what resembles giftedness and what conceals it (Section 12.9).

**Identification, not diagnosis: what sets the concept apart in giftedness**

In the field of gifted education, the term **identification** (**Identification**) denotes a comprehensive educational process whose purpose is to determine which students need enrichment or acceleration services beyond what the regular classroom offers. Giftedness is not a disorder whose symptoms are sought out; it is high potential or performance for which we seek the service that will develop it. The process therefore starts from the program's definition of giftedness and from the services it provides: the instrument is chosen according to the type of service, multiple criteria are used instead of a single score, and the student is compared with local norms that take into account the opportunities he or she has had (NAGC, 2019b; Pfeiffer, 2015). Its outcome is a decision to place the student in a program, open to review, unlike a permanent classification that follows the student, because giftedness develops and emerges at different times (Lohman & Korb, 2006). Its most common error is **omission**: missing the student who was never nominated in the first place, which happens more often than placing a student who does not need the program (Card & Giuliano, 2016).

With disabilities, by contrast, identification is bound up with determining **eligibility** for special education services: the questions are whether the student has a disability in one of the specified categories and whether it affects his or her learning to a degree that calls for those services. In many categories, such as autism spectrum disorder and attention-deficit/hyperactivity disorder, this runs alongside a clinical diagnosis made against the criteria of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2022) or the International Classification of Diseases (**ICD**), issued by a qualified professional and sometimes the basis for treatment outside school. The question in disability is "What impedes this student's learning, and does he or she

qualify for services?"; in giftedness it is "What does this student need that the classroom does not provide?" For this reason, we prefer the term "identification" for giftedness and use "diagnosis" when we are speaking of a disorder, as when distinguishing giftedness from conditions that resemble it in Section 12.9.

## 12.1 What is giftedness?

Definitions of giftedness abound, and no single agreed definition unites them. This multiplicity has practical consequences, because the definition a program adopts determines whom it looks for and with which tools. A program that defines giftedness as a high IQ will administer an intelligence test; one that defines it as creativity will look for creative products; and one that defines it as excellence in a particular domain will measure that domain. The most prominent conceptions follow.

### Giftedness as high general intellectual ability

The scientific study of the gifted began with Lewis Terman, who launched his longitudinal study of gifted children in the 1920s and defined the gifted child by a very high score on an intelligence test: an IQ of roughly 135–140 or above on the Stanford–Binet for most of his sample, that is, approximately the top 1% (Terman, 1925). General intellectual ability remains part of most definitions, but relying on it alone has been criticized for overlooking giftedness in specific domains, creativity, and motivation.

### Renzulli's Three-Ring Conception

In his Three-Ring Conception, Joseph Renzulli proposed that gifted behavior emerges when three traits come together: above-average (though not necessarily exceptional) ability, task commitment, and creativity (Renzulli,

1978). What matters in this model is that it speaks of **gifted behavior** that appears in certain situations, rather than a fixed trait that a student either has or lacks.

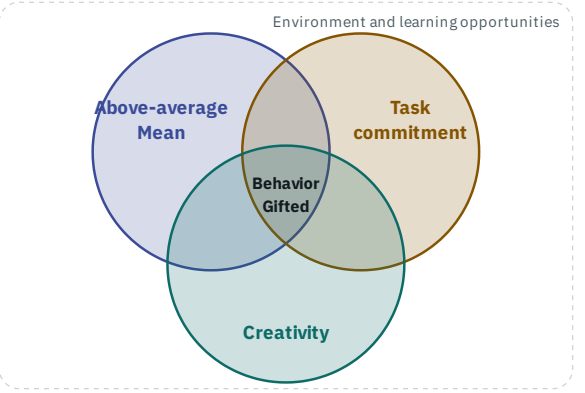



**Figure 12.1:** The Three-Ring Conception. Gifted behavior emerges where ability, commitment, and creativity intersect, surrounded by an environment that either provides or withholds opportunities for it to appear.

Renzulli distinguishes between two kinds of giftedness. The first is **schoolhouse giftedness** (**Schoolhouse giftedness**), proficiency in learning lessons and taking tests, which ability and achievement tests detect well. The second is **creative-productive giftedness** (**Creative-productive giftedness**), the capacity to produce original ideas and products for a real audience, which tests alone do not reveal (Renzulli, 1986). This distinction bears directly on identification. In the Schoolwide Enrichment Model associated with the theory, a **talent pool** (**Talent pool**) of roughly 15–20% of students is formed, instead of selecting a small number by a cutoff score, by combining ability and achievement scores with teacher ratings. Students who display gifted behavior during enrichment activities may then move to a higher level of service, even if they were not in the pool to begin with (Renzulli & Reis, 2014). This is what the phrase “above-average, though not necessarily exceptional, ability” means.

## Gagné's model: from gifts to talents

In his Differentiated Model of Giftedness and Talent (**Differentiated Model of Giftedness and Talent, DMGT**), François Gagné distinguishes between two words often used as though they meant the same thing: **gifts** (**Gifts**), high natural abilities not yet refined, and **talents** (**Talents**), mastered skills in a particular field that have developed from those gifts through learning and practice. Gagné later revised his model in a second version (DMGT 2.0), changing three things that bear directly on identification (Gagné, 2009, 2013).

**First**, gifts now span six domains: intellectual, creative, social, perceptual, muscular, and motor control. Intellectual ability is thus one domain among six, and a student may be gifted in a domain that an intelligence test does not measure at all. **Second**, the **developmental process** became a separate component with three dimensions: the activities the student gains access to, along with their content and format; the investment he or she makes in time, effort, and money; and progress, with its stages, pace, and turning points. The process is surrounded by intrapersonal catalysts, such as physical and mental traits and goal management (awareness, motivation, and volition), and by environmental catalysts, such as the milieu, influential individuals, and available services including enrichment and acceleration. **Third**, **chance** reaches beyond environmental factors to affect every component: the gifts themselves through heredity, the family and place into which a child is born, and the opportunities and people the child happens to encounter.

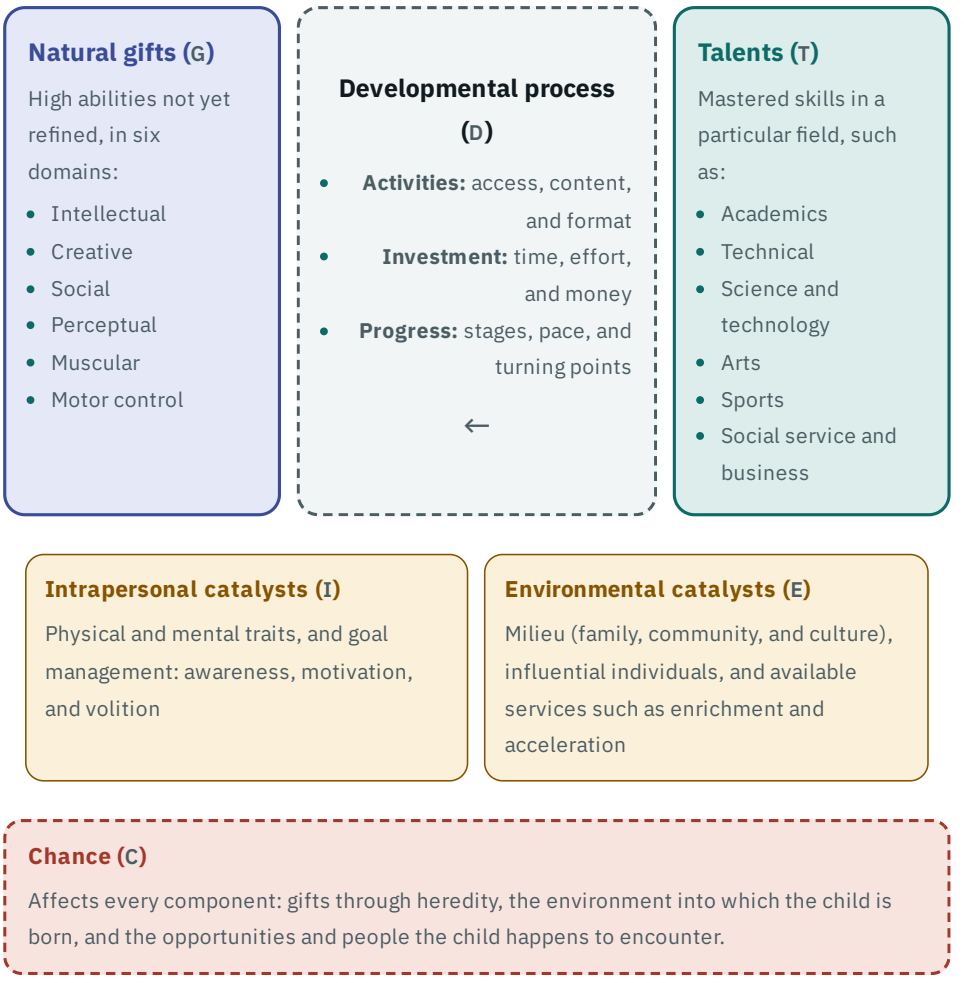



**Figure 12.2:** The Differentiated Model of Giftedness and Talent in its second version, simplified by the author (after Gagné, 2009, 2013). Gifts are transformed into talents through a developmental process with its own dimensions, which intrapersonal and environmental catalysts speed up or slow down; chance touches every component, as the dashed frame around them indicates. The list of fields in the right column is illustrative.

The second version adds something missing from most definitions: a **numerical threshold**. For Gagné, the gifted are those who fall within the top 10% of their peers in a domain of gifts or a field of talent. Those above the threshold are then divided into five levels, each ten times rarer than the one before.

**Table 12.1: Levels of giftedness in the Differentiated Model (Gagné, 2009) and their approximate equivalents on an intelligence scale with a mean of 100 and a standard deviation of 15.**

| Level | Rarity among peers | Approximate IQ equivalent |
|---|---|---|
| **Mild** | 1 in 10 (top 10%) | About 120 |
| **Average** | 1 in 100 | About 135 |
| **High** | 1 in 1,000 | About 145 |
| **Exceptionally** | 1 in 10,000 | About 155 |
| **Extremely** | 1 in 100,000 | About 165 |

This threshold gives the reader a reference point for reading admission rules. An IQ of 115 corresponds roughly to the top 16%, 120 to the top 10% (Gagné's threshold), 125 to the top 5%, and 130 to the top 2%. Terman's cutoff falls at Gagné's "moderately" gifted level. When a rule requires a particular score, it is implicitly choosing one of these levels, and it would do well to state that choice and its rationale (we will see in Section 12.7 that Saudi regulations have used more than one cutoff). Note that the numerical equivalents in the table are approximate and lose precision the higher we go, because norming samples include only very small numbers at the upper extreme.

The model has a further implication for identification: a student in an environment with limited opportunities may have high potential that has not yet turned into outstanding performance. If we look only for demonstrated performance, we will miss this student. Identification therefore needs instruments that reveal potential alongside achievement.

## Giftedness as talent development

Recent theories increasingly view giftedness as a developmental process that extends across the lifespan. Its clearest expression is the **talent development megamodel** (**Talent development megamodel**) proposed by Rena Subotnik

and her colleagues (Subotnik et al., 2011) and elaborated in a later review (Worrell et al., 2019). The model rests on five ideas, each with implications for identification:

- **Giftedness is a trajectory within a specific domain.** The trajectory begins with potential, then moves to competency, then expertise, and finally eminence or professional creativity. Identification looks for something different at each stage: potential at the outset, then demonstrated achievement as the student advances.
- **Each domain has its own timing.** The starting point, peak, and end of the trajectory vary from one domain to another. Mathematics, music, and sports begin early, and outstanding students in them can be identified in elementary school. Domains such as psychology, law, and philosophy begin later, and aptitude for them emerges only after a student has studied them. Both the age of identification and the type of instrument, in other words, depend on the domain.
- **Ability is necessary and malleable.** General ability and domain-specific abilities are a prerequisite, yet they grow and are shaped by schooling and opportunity.
- **Psychosocial skills are decisive.** These include persistence, tolerance of criticism, stress management, confidence, and the ability to learn from mentors. Their importance grows as the student advances along the trajectory, until at the higher stages they are what separates people of comparable ability.
- **Opportunities must be offered, and they must be taken.** Schools and families must provide opportunities, and the student must seize them. The aim of the model is for the talented individual to reach, in adulthood, eminent achievement or a creative contribution to his or her field; what counts is what the person attains as an adult, more than a "gifted" label carried in childhood.

This is consistent with the definition of the National Association for Gifted Children (**National Association for Gifted Children, NAGC**), which describes gifted individuals as those who perform, or have the capability to perform, at higher levels compared with others of the same age, experience, and environment in one or more domains (NAGC, 2019a). Note two phrases in this definition: "have the capability to perform," meaning potential alongside achievement, and "environment," meaning that the comparison takes account of the opportunities available.

## Pfeiffer's tripartite model

Steven Pfeiffer proposed a model closer to assessment practice than to theory building, one that views giftedness through three lenses, each with its own tools (Pfeiffer, 2015, 2018). **The first lens, high intelligence**, is revealed by individual and group intelligence tests and suits programs that offer general academic challenge. **The second lens, outstanding accomplishment**, is revealed by achievement tests, products, and performance in the field, and suits subject acceleration or domain-specific programs. **The third lens, potential to excel**, concerns the student who has not yet had the opportunities; it is revealed by rating scales, observation over time, and dynamic assessment, and it suits a program that provides opportunities and then revisits the decision. The model's value lies in reminding the team that a student may be gifted through one lens and invisible through the others, and that the third lens is the least stable of the three, so that its decision should be treated as provisional and subject to review.

Table 12.2 summarizes the four models from the standpoint of identification. Its "instruments" column applies the principle of aligning identification with services, which is taken up in the next section.

**Table 12.2: Four models of giftedness from the standpoint of identification.**

| Model | What it looks for | Typical instruments | Threshold | Whom it may miss |
|---|---|---|---|---|
| **Renzulli: the three rings** | Gifted behavior that emerges when above-average ability, task commitment, and creativity converge | Ability and achievement tests, teacher rating scales, and products and projects observed during enrichment | A talent pool of about 15–20%, with flexible entry for those who display the behavior | Students given no situations in which to display it, and those of exceptional ability who need acceleration beyond enrichment |
| **Gagné: the Differentiated Model** | High gifts in six domains, and talents in specific fields | Ability tests for gifts; achievement tests and performance assessment for talents | The top 10% of peers, in five levels | Students near the threshold, whose outcome is decided by measurement error, and those whose gifts lie in a domain the available instruments do not measure |

| Model | What it looks for | Typical instruments | Threshold | Whom it may miss |
|---|---|---|---|---|
| **Subotnik and colleagues: talent development** | The student's position on the trajectory of a specific domain, and his or her psychosocial skills | Domain-specific instruments, above-level testing, and performance before experts in the field | Varies by stage and domain: potential first, then achievement | Students who lacked early opportunities in an early-starting domain, and able students with weak psychosocial skills |
| **Pfeiffer: the three lenses** | High intelligence, outstanding accomplishment, or potential to excel | Intelligence tests; achievement tests and products; rating scales, observation, and dynamic assessment | Set for each lens according to the program | Students who appear through no lens because the instruments do not suit their language or disability, and those judged through the third lens whose decision is never reviewed |

**Common misconceptions**

- **"A gifted student excels in every subject."** Many gifted students are outstanding in one area and average in others, and some have a difficulty or disability alongside their giftedness; these are twice-exceptional learners (Chapter 13 discusses them in detail).
- **"A gifted student will find the way alone."** Gifts do not turn into talents without opportunities and appropriate instruction, and a gifted student who finds no challenge may lose motivation.
- **"A gifted student is well behaved and well liked by teachers."** Some gifted students challenge authority, ask awkward questions, and grow bored with repetition, so they come across as disruptive or indifferent.

## 12.2 Principles of identifying gifted learners

Most of the literature and professional standards agree on a set of principles for fair identification, among them those set out in the Pre-K–Grade 12 Gifted Programming Standards issued by the same association (NAGC, 2019b):

**Table 12.3: Principles of fair identification of gifted learners and what they mean in practice.**

| Principle | What it means in practice |
|---|---|
| **Aligning identification with services** | Instruments are chosen according to what the program will provide. A mathematics acceleration program, for example, needs a measure of mathematical ability specifically. |
| **Multiple sources** | The decision rests on evidence from ability, achievement, observation, and performance taken together, rather than on a single score. |

| Principle | What it means in practice |
|---|---|
| **Universal screening** | All students at a given grade level are assessed, rather than waiting for nominations, because nomination misses many. A study in a large U.S. school district found that moving from nomination to universal screening markedly increased the representation of students from low-income and minority families in gifted programs (Card & Giuliano, 2016). |
| **Fair instruments** | Instruments that do not disadvantage a student because of language, background, or disability, including nonverbal tests within their limits (see Section 12.4) and **local norms** (**Local norms**), that is, comparing the student with peers in the same school or district rather than with the national norming sample alone. |
| **Ongoing identification** | Doors remain open across grade levels, because giftedness develops and emerges at different times. A longitudinal study that followed elementary students found that a large proportion of those who exceeded a high cutoff in one year no longer did so two or three years later, while others crossed it later on (Lohman & Korb, 2006). A label assigned once and never reviewed errs in both directions. |

## How are local norms constructed?

The idea behind local norms is simple: we ask where the student stands among those who have had similar opportunities to learn, rather than among a national sample that includes far wealthier schools. Peters and colleagues found that computing norms at the school level increases the representation of underrepresented groups in gifted programs (Peters et al., 2019). Applying them, however, requires decisions made in advance and announced openly, which the following box summarizes.

**Six questions before using local norms**

1. **Who is the reference group?** Students in the same grade who share the same opportunities to learn: the school, or a group of similarly situated schools within the education directorate, such as its rural schools. Schools with very different opportunities should not be pooled into one group, as this reintroduces the very problem local norms were meant to solve.
2. **What is the minimum size?** There is no agreed figure, but the upper ranks require large numbers. In a class of 20 students, the "top 5%" is a single student. A percentile rank is the percentage of the student's peers in the reference group or norming sample who scored lower than the student; a percentile rank of 84 means the student outperformed 84% of them. In a class of this size, a percentile rank of 95 does not distinguish the first student from the second, and a change in one or two answers is enough to change the judgment. If the group numbers only in the tens, it is better to enlarge it or to combine years.
3. **School or directorate?** School norms come closer to the question "Who needs more than his or her class provides?", while directorate norms are more statistically stable. A common compromise is to compute norms for a group of similarly situated schools, or for the school with several years combined.
4. **Should data be pooled across years?** Yes, when classes are small, provided that the instrument and administration conditions remain constant and that the report says so.
5. **What are their limits?** The label does not transfer between schools: a student identified in one school might not be identified in another. The absolute level of identified students also varies from school to school, so services must start from the student's actual level as the evidence

shows it. Moreover, local norms reveal able students in less advantaged settings, but they do not address the cause of the gap.

6. **How do they relate to excellence gaps? Excellence gaps** (**Excellence gaps**) are differences between groups in the proportions reaching advanced levels of performance (Plucker & Peters, 2016). Local norms are a first step toward closing them, because they reveal who merits services; closing the gaps themselves requires services that make up for the opportunities that were missing.

## 12.3 Nomination and the opportunity to display

Teacher nomination remains the most widespread route into identification. It has its uses, since the teacher sees the student over a long period and in many situations. But it has well-known limits: teachers tend to nominate high-scoring, well-behaved students who fit the image of the "good student," and to overlook those who are quiet, those who differ linguistically or culturally, and those whose ability is masked by a difficulty. Nomination may even be influenced by the teacher as well as by the student: a large U.S. study found that the likelihood of a student being identified as gifted varied with the background of the student's teacher, even when students' test scores were equal (Grissom & Redding, 2016). In Saudi Arabia, Alfaiz examined the accuracy of parent and teacher nomination in identifying gifted students (Alfaiz, 2026), a question on which local practitioners will need to follow the evidence.

### The opportunity to display

In his research on the visibility of gifted students for nomination, the author of this book proposes a framework he calls the **opportunity to display** (**Opportunity to display**), which locates a reason for the shortcomings of nomination deeper than teacher bias: **a teacher cannot see what the**

**classroom gives the student no opportunity to show**. A classroom built on memorization and short answers does not reveal a student's ability to solve an open-ended problem, pose an original question, or build a project. Many nomination checklists ask about behaviors, such as curiosity, unusual questions, and original thinking, that appear only in activities that allow for them. When such activities are absent, what might be called the teacher's **inference load** increases: the teacher is forced to infer giftedness from indirect signs such as grades and good conduct, and so falls back on nominating high achievers.

The framework is close to well-established ideas in the literature; what it adds is to make opportunity a **first step** in the identification pathway, where it had been a circumstance mentioned when interpreting results. An instrument such as the Teacher's Observation of Potential in Students links observation to providing rich learning activities over several weeks (Coleman et al., 2010). Local norms compare the student with those who have had similar opportunities (Peters et al., 2019). And Pfeiffer's "potential to excel" lens concerns students who have not yet had the opportunities (Pfeiffer, 2015). What is new here is the question that comes before all of them: did the classroom offer anything to observe in the first place?

**What this means in practice**

Before we ask teachers to nominate, we need to give students opportunities to show what they can do: open-ended activities, problems without a single solution, and independent projects, even if only for a few weeks. Observation and nomination come afterward. In Jana's scene at the start of the chapter, no one noticed her ability until the open-ended science project gave her an opportunity to display it.

**For the teacher: indicators worth noticing**

These are situations to watch for and record; when you see them, note the date and context, without waiting for grades:

- Grasps a new idea after a brief presentation and grows restless when it is repeated.
- Asks questions that go beyond the lesson, or connects two subjects no one else has connected.
- Reaches a solution by a route other than the one taught, and can explain that route.
- Holds a deep, sustained interest in a topic outside the curriculum and knows more about it than would be expected at his or her age.
- Uses vocabulary or reasoning beyond his or her years, or understands complex humor before peers do.
- Produces outside school what does not show up in it: a project, a program, drawings, or writing.
- Shows clear improvement in performance and behavior when given a harder task.

If you notice these indicators in a student whose grades are average or whose behavior is disruptive, that is all the more reason to record them.

## Nomination and observation tools

**Table 12.4: Common nomination and observation tools in identifying gifted learners.**

| Instrument | The idea | Distinguishing features |
|---|---|---|
| **Teacher's Observation of Potential in Students (TOPS)** | The teacher observes students over several weeks during rich learning activities and records behaviors in nine domains, such as learning quickly, curiosity and creativity, advanced reasoning, spatial ability, and leadership (Coleman et al., 2010) | Observation is tied to providing activities that offer an opportunity to display; designed for young students and for students from less advantaged backgrounds |
| **Traits, Aptitudes, and Behaviors (TABs)** | Ten core attributes of giftedness observed by the teacher: motivation, interests, communication skills, problem solving, memory, inquiry, insight, reasoning, imagination and creativity, and humor (Frasier et al., 1995) | Describes how each attribute may manifest differently in students from different cultures |
| **Gifted Rating Scales (GRS)** | A standardized teacher rating scale measuring intellectual ability, academic ability, creativity, artistic talent, leadership, and motivation (Pfeiffer & Jarosewich, 2003) | Has norms and T-scores with a mean of 50 and a standard deviation of 10, so the student is compared with peers (see its profile in Section 12.6) |
| **Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS)** | A set of teacher rating scales across multiple areas, such as learning, creativity, motivation, leadership, and the arts (Renzulli et al., 2010) | Flexible; the program can select the scales that suit it. Lacks national norms, so its scores are read against local norms that the program builds |

At the time of writing, we found no published Saudi norming study of any of these four tools to which we could refer readers. Anyone using an Arabic version of one of them should verify its source and the evidence its developers provide for its validity and reliability in the local setting. Whatever the instrument, teacher ratings remain one source among several, to be combined with measures of ability, achievement, and performance.

## 12.4 Ability measures and their ceilings

### Group and individual tests

In universal screening, group ability tests are typically administered to an entire class, on paper or online, such as the Cognitive Abilities Test (**CogAT**), which measures verbal, quantitative, and nonverbal reasoning, and the Naglieri Nonverbal Ability Test, Third Edition (**NNAT3**), which relies on figural matrices and is therefore little affected by language. Students who exceed the **screening cutoff**, the score above which a student is referred for more in-depth assessment, are then evaluated with an individual intelligence test such as the Wechsler Intelligence Scale for Children or the Stanford–Binet (Chapter 3 presents these instruments in detail, and the Cognitive Abilities Test profile appears in Section 12.6).

### Limits of nonverbal tests

Nonverbal tests are often presented as the answer to fairness for linguistically diverse students and those from less advantaged settings. There is some truth in this, but the literature points to three limits (Lohman, 2005). **First**, nonverbal reasoning predicts academic achievement less well than verbal and quantitative reasoning do, and academic programs need students who will succeed in their subject matter. **Second**, such tests sometimes narrow group differences, though their effect on this is weaker than commonly assumed. **Third**, relying on them alone may exclude students who genuinely excel in

verbal or quantitative reasoning, since their performance on matrices and figures is not necessarily exceptional. The alternative this literature proposes is to keep the nonverbal test within a battery that includes verbal and quantitative reasoning, and then to compare the student with those who have had similar opportunities, that is, against local norms.

## The General Ability Index and the Cognitive Proficiency Index

On the Wechsler Intelligence Scale for Children, Fifth Edition (**WISC-V**), the **Full Scale IQ** (**Full Scale IQ, FSIQ**) is influenced by the Working Memory and Processing Speed indexes (Chapter 3 defines these indexes). Many gifted students, particularly twice-exceptional learners, perform noticeably lower on these two indexes than on reasoning and comprehension.

Two composite scores are therefore derived from the scale to separate the two sides. The first is the **General Ability Index** (**General Ability Index, GAI**), built from five subtests in verbal comprehension, visual-spatial processing, and fluid reasoning: Similarities, Vocabulary, Block Design, Matrix Reasoning, and Figure Weights. The second is the **Cognitive Proficiency Index** (**Cognitive Proficiency Index, CPI**), built from four subtests in working memory and processing speed: Digit Span, Picture Span, Coding, and Symbol Search. The former describes what the student can reason through; the latter describes how efficiently he or she processes information quickly and accurately (Kaufman et al., 2016).

The difference between the two indexes is read like any difference between two scores: Is it statistically significant, that is, too large to be attributed to measurement error? And how rare is it in the norming sample? (See Chapter 2.) A student with a General Ability Index of 132 and a Cognitive Proficiency Index of 104 shows a 28-point gap. Such a difference merits interpretation if it proves to be both significant and rare, as established by statistical computation rather than by the visual impression of a jagged profile. The

National Association for Gifted Children recommends using the General Ability Index in identifying gifted students and not requiring the Full Scale IQ when a student's profile is uneven (NAGC, 2018).

When a student reaches the ceiling of several subtests, answering the last items or most of them, the Full Scale IQ and the General Ability Index can no longer capture the student's ability, because the score cannot go beyond the highest value in the norm tables. For this reason, the publishers of some scales, including the Wechsler, have issued **extended norms** (**Extended norms**) that extend scores beyond their usual limit for students who reach the ceiling (see Chapter 2). These help distinguish among the upper levels of giftedness, but they rest on very small numbers at the top of the distribution, so they should be read with caution, and the publisher's conditions for their use should be consulted.

## The ceiling effect and above-level testing

In assessing intellectual disability, a problem known as the **floor effect** (**Floor effect**) arises: the test lacks sufficiently easy items, so it cannot distinguish among very low levels of performance. Giftedness presents the opposite problem: the **ceiling effect** (**Ceiling effect**). An achievement test designed for a given grade lacks sufficiently difficult items, so a gifted student answers almost everything correctly and obtains a percentile rank of 99, just like a classmate who has merely mastered the grade-level curriculum. We then have no way of knowing where either student's ability actually lies.

The solution developed in talent search programs (**Talent Search**) is **above-level testing** (**Above-level testing**): an outstanding student takes a test designed for students several years older. Students who were tied at the ceiling spread out, revealing those who need genuine acceleration. A long-term investigation, the Study of Mathematically Precocious Youth (**Study of Mathematically Precocious Youth, SMPY**), followed thousands of students

identified in this way before the age of 13 and found that differences among them on above-level tests predicted large differences in their accomplishments decades later (Lubinski & Benbow, 2006).

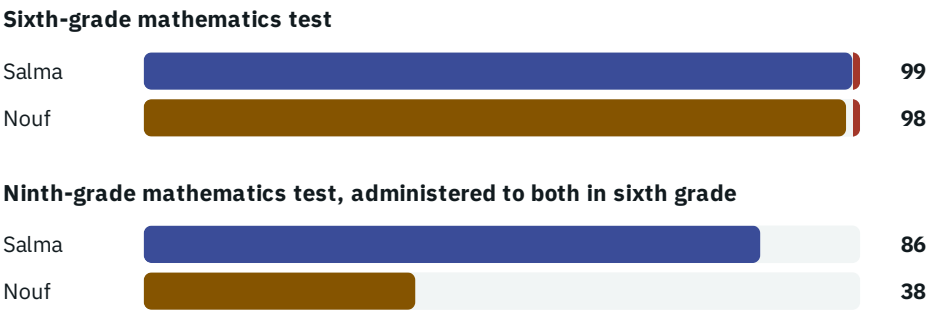
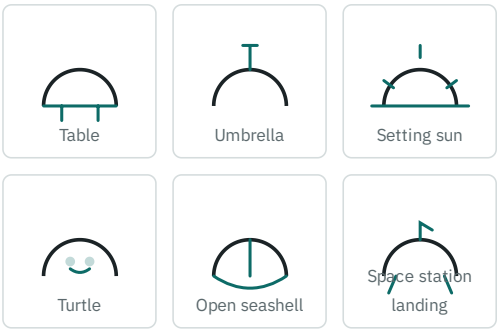



**Figure 12.3:** A hypothetical illustration of the ceiling effect. Values are percentile ranks. On the grade-level test, both students sit at the ceiling. On the above-level test, Salma outperforms most ninth graders, whereas Nouf falls below their average, an excellent performance for her age, but one that points to a different need.

## 12.5 Creativity, achievement, and performance

### Measuring creative thinking

Among the best-known instruments for measuring creative thinking are the Torrance Tests of Creative Thinking (**Torrance Tests of Creative Thinking, TTCT**), whose profile appears in Section 12.6. They come in a verbal form and a figural form. In the figural form, the student is given incomplete or repeated shapes and asked to complete them as meaningful drawings and to give each drawing a title. The responses are then scored on several dimensions, among them **fluency** (**Fluency**), the number of ideas; **flexibility** (**Flexibility**), the variety of their categories; **originality** (**Originality**), their rarity relative to other people's responses; and **elaboration** (**Elaboration**), how much the student adds to the basic idea (Kim, 2006).




**F** **Fluency:** six ideas.

**X** **Flexibility:** furniture, weather, nature, animals, and science fiction: five categories.

**O** **Originality:** "space station landing" and "open seashell" are rarer than "umbrella" and "table."

**E** **Elaboration:** sparse in most drawings, greater in the sun and the space station.

**Figure 12.4:** An illustrative activity designed by the author in the style of repeated-figures tasks; it is not taken from any published test. A hypothetical student completed a semicircle in six ways, and alongside the drawings is a reading of the responses on the four dimensions.

Creativity tests have important limitations. They measure **divergent thinking** (**Divergent thinking**), the production of many and varied ideas, which is only one facet of creativity. Performance on them is sensitive to the instructions and to the student's mood on the day of testing, and their reliability is lower than that of intelligence tests. Their predictive validity, that is, their capacity to predict creative achievement in adulthood, remains contested: some longitudinal studies have found a relationship between their scores and creative achievement decades later, while critics consider the relationship modest and inconsistent across studies (Kim, 2006). For this reason they serve as a supplementary source, alongside assessment of the student's actual creative products over time.

For evaluating the product itself, a well-known method is the **Consensual Assessment Technique** (**Consensual Assessment Technique**), proposed by Teresa Amabile (Amabile, 1982). Its premise is that the creativity of a product is defined by the judgment of those who know the field: the product, whether a story, a drawing, or a design, is presented to several judges with expertise in the field, each of whom rates its creativity independently of the others, and

their agreement is then examined. If they agree to an acceptable degree, the mean of their ratings is taken as a measure of the product's creativity. This method comes closest to what "performance before a panel" means in the arts, writing, and design, provided that the judges are experts in the field and that they rate independently before any discussion.

### Achievement and performance

Very high achievement in a domain signals a talent already realized in it. It is measured with achievement tests, preferably supplemented by above-level testing to overcome the ceiling effect, as discussed earlier. Performance, for its part, is measured through authentic performance tasks, that is, tasks resembling what specialists in the field actually do, and through portfolios that collect samples of the student's work over time. In some domains, such as the arts, sports, and leadership, performance before an expert panel is the primary source of evidence, because paper-and-pencil tests do not capture it.

## 12.6 Instrument profiles

The following profiles present four instruments that represent the main types of evidence used in identifying gifted learners: a group ability test for screening, a teacher rating scale, a test of creative thinking, and the above-level testing procedure. The profile of the Mawhiba test appears in Section 12.7, together with the regulatory context in which it is used. None of the profiles reproduces items from any published test.

### CogAT Cognitive Abilities Test, Form 7

Group ability test

### The idea

Students learn at school through three symbol systems: words, numbers, and shapes and pictures. This test was therefore designed to measure reasoning within each system separately rather than to yield a single IQ. The aim is to identify the student's ability profile so that suitable instruction or programming can be chosen (Lohman, 2011).

### What it looks like

Three batteries, each with three subtests. **Verbal battery:** verbal or picture analogies, sentence completion, and classification. **Quantitative battery:** number analogies, number puzzles, and number series. **Nonverbal battery:** figure matrices, paper folding, and figure classification. At the early levels, from kindergarten to second grade, items are presented as pictures and directions are read aloud, so the student does not need to read. The test is administered on paper or online to a whole class.

### A scene from administration

At a private international school that teaches in English, the administration decides to conduct universal screening of its third graders. Teachers administer the three batteries over two consecutive mornings. The profile of one student, whose first language is Arabic, reads as follows: Quantitative 131, Nonverbal 128, Verbal 112. The school summarizes the profile with a code such as "9B (V-)": the number is the median stanine, the letter describes the extent of variation across the batteries, and the parenthesis identifies the relatively low battery. The school psychologist asks: is the lower verbal score a weakness in verbal reasoning, or an effect of learning in a second language? She decides to compare him with peers who are also learning in a second language before reaching any judgment.

**Table 12.5: Profile of the Cognitive Abilities Test.**

| | |
|---|---|
| Purpose | Measurement of verbal, quantitative, and nonverbal reasoning for universal screening and instructional planning |

| | |
|---|---|
| **Ages** | Kindergarten through Grade 12, in grade-based levels |
| **Edition and year** | Form 7, 2011 |
| **Publisher** | **Riverside** |
| **Qualification level** | Administered in class by the teacher under standardized directions; results interpreted by a trained professional |
| **Time** | The manual specifies the time for each battery by level; the three batteries are administered in one or more sessions |
| **Score types** | Standard age score with a mean of 100 and a standard deviation of 16 (not 15); percentile ranks and stanines based on age and grade norms; a composite score; and an ability profile summarizing level and pattern |
| **Cutoff scores** | Set by each program according to its service |
| **Reliability** | High for the composite according to the technical manual, and lower for individual batteries |
| **Classification accuracy** | Not applicable in the diagnostic sense. What matters is its predictive validity for achievement, which is stronger for the verbal and quantitative batteries than for the nonverbal battery |
| **Year of norm collection** | U.S. norms; consult the publisher for the norm year used in scoring |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

Interpretation begins with pattern before level: are the three batteries close together or divergent? Scores are then chosen to fit the service. A mathematics acceleration program looks first at the quantitative and nonverbal scores, and

a language program looks at the verbal score. Note that the standard age score has a standard deviation of 16, so a score of 130 on it is not equivalent to 130 on a Wechsler scale.

### Its limits

Its items and norms are in American English, so it is unsuitable for students learning in Arabic unless adapted and normed. A group test is also less precise for the individual than an individual test; it is a screening tool that paves the way for more precise assessment. Relying on the nonverbal battery alone reintroduces the limitations of nonverbal tests discussed in Section 12.4. Among the best-known nonverbal alternatives is the Naglieri Nonverbal Ability Test, Third Edition (**NNAT3**), a 48-item matrices test completed in about 30 minutes that yields an index with a mean of 100 and a standard deviation of 15 (Naglieri, 2016). We likewise know of no published, normed Arabic version of it.

## GRS Gifted Rating Scales

Rating scale

### The idea

A teacher sees a student in many situations, yet one teacher's impression is not easily compared with another's. This scale was designed to convert teacher observation into a standard score that compares the student with same-age peers across six domains of giftedness that extend beyond intellectual ability (Pfeiffer & Jarosewich, 2003).

### What it looks like

The school form contains 72 items distributed across six scales of 12 items each: intellectual ability, academic ability, creativity, artistic talent, leadership, and motivation. The teacher rates each item on a nine-point scale

divided into three ranges: below average (1–3), average (4–6), and above average (7–9), relative to students of the same age. A second form covers preschool and kindergarten.

### A scene from administration

A fifth-grade teacher completes the school form for a student she has taught for two semesters. She pauses at an item on leadership and recalls that the girl has never led a group, because the teacher has always assigned the roles herself. Rather than choosing a number from the middle range, she notes in the margin that she has had no opportunity to observe this behavior.

**Table 12.6: Profile of the Gifted Rating Scales.**

| | |
|---|---|
| **Purpose** | Teacher rating of gifted behaviors in six domains, as one source within a multi-source assessment |
| **Ages** | School Form for Grades 1–8 (ages 6–13); Preschool Form from 4 years to 6 years 11 months |
| **Edition and year** | First edition, 2003 |
| **Publisher** | **Pearson** |
| **Qualification level** | Completed by a teacher who has known the student long enough; interpreted by a professional in accordance with the publisher's requirements |
| **Time** | A few minutes per student |
| **Score types** | T-score for each scale, with a mean of 50 and a standard deviation of 10 |
| **Cutoff scores** | The manual classifies T-scores into categories of likelihood of giftedness; consult it for the cut points |
| **Reliability** | The manual reports very high internal consistency for all six scales |

| | |
|---|---|
| **Classification accuracy** | The manual reports good discrimination between gifted and non-gifted samples; this requires independent studies in the local context |
| **Year of norm collection** | A U.S. sample collected shortly before publication, so the norms are more than twenty years old |
| **Arabic version** | At the time of writing we know of no published standardized Arabic version |

### How are the results interpreted?

The six scales are read as an integrated profile before any total. A high score in artistic talent alone points toward a performance assessment in art rather than a general academic program. The score is also compared with what another source says: if intellectual ability is high on the rating but low on the test, the question becomes what the teacher saw that the test did not, or vice versa.

### Its limits

It reflects the teacher's impression and what the classroom allowed the teacher to see, and so it carries the limitations of nomination discussed in Section 12.3. A student's scores on the six scales may also cluster closely, which suggests a global impression more than discrimination among domains, a phenomenon known as the halo effect. It does not cover the middle and secondary school years. The Scales for Rating the Behavioral Characteristics of Superior Students, by contrast, are flexible and cover more domains, but they lack national norms and therefore require local norms built by the program (Renzulli et al., 2010).

## TTCT Torrance Tests of Creative Thinking

Performance test

### The idea

If creativity is one component of giftedness, we need a task that shows how a student generates ideas when a question has no single answer. That is what this test was designed to do: open-ended tasks in which ideas, their variety, and their rarity are counted, regardless of whether they are correct (Kim, 2006).

### What it looks like

The test has two forms, each with two equivalent versions, A and B. **The figural form** consists of three activities of ten minutes each: picture construction from a given shape, completion of incomplete figures, and transformation of repeated shapes into different drawings, with a title for each drawing. **The verbal form** comprises written activities such as asking questions about a picture, guessing causes and consequences, improving a product, and proposing unusual uses for a familiar object. Under the current scoring method for the figural form, fluency, originality, and elaboration are scored, along with abstractness of titles and resistance to premature closure, that is, keeping a figure open long enough before closing it with a familiar solution. Flexibility is no longer scored on the figural form, although it remains a useful concept for reading responses, as in Figure 12.4.

### A scene from administration

The specialist administers the figural form to a small group of students. He reads the directions exactly as they appear in the manual, because a single extra word of encouragement can raise the number of ideas. He notices one student who finished the first activity in four minutes and then sat staring at the ceiling. He records this in his notes: a low fluency score may reflect that day's motivation more than the student's ability.

**Table 12.7: Profile of the Torrance Tests of Creative Thinking.**

| | |
|---|---|
| **Purpose** | Measurement of divergent thinking, as a supplementary source in identifying gifted learners |
| **Ages** | Kindergarten to adulthood |

| | |
|---|---|
| **Edition and year** | First published in the 1960s; the norms manual has been reissued several times, and the publisher should be consulted for the latest |
| **Publisher** | **Scholastic Testing Service** |
| **Qualification level** | Administered by a teacher or specialist after training; scored by a trained scorer or by the publisher's scoring service |
| **Time** | Figural form about 30 minutes; verbal form longer |
| **Score types** | Standard scores and percentile ranks for each dimension, and an overall creativity index |
| **Cutoff scores** | Set by the program, if it uses the test |
| **Reliability** | Lower than that of intelligence tests; varies across dimensions and is affected by scorer training |
| **Classification accuracy** | Not applicable in the diagnostic sense. Its predictive validity for creative achievement in adulthood is contested (Kim, 2006) |
| **Year of norm collection** | U.S. norms, with the norm year varying by edition of the manual |
| **Arabic version** | Figural Form B was normed in Saudi Arabia with support from King Abdulaziz City for Science and Technology in 1416 AH, through a pilot study of 598 male and female students followed by norms based on 1,227 male and female students aged 9 to 16, with scores scaled to a mean of 100 and a standard deviation of 15, according to a published account by one of its developers (Al-Nafi, 1428 AH [Arabic]). We have not seen the original norming report, and the norms are nearly thirty years old |

### How are the results interpreted?

The profile of dimensions is read before the overall index. High fluency with average originality means many familiar ideas, whereas high originality with average fluency may signal a student who thinks carefully before drawing. The result is then compared with the student's actual products, preferably evaluated with the Consensual Assessment Technique described in Section 12.5.

### Its limits

It measures divergent thinking, which is only one facet of creativity, and it is sensitive to the instructions and to the student's mood on the day of testing. Some items in the verbal form depend on cultural experiences and therefore require adaptation. The available Saudi norms are dated, and they were built on originality in its statistical sense, that is, the rarity of an idea within the sample. Such rarity shifts over time and with changes in what children are exposed to: an idea that was rare thirty years ago may be commonplace today.

## Above-level Above-level testing

Assessment procedure

### The idea

We saw in Section 12.4 that a grade-level test stops at its ceiling, where a student who has mastered the grade's curriculum becomes indistinguishable from one who is years ahead. The idea is to raise the ceiling by giving the student a test designed for older students. The procedure was developed by talent search programs, the best known of which is the Study of Mathematically Precocious Youth, which administered a college admission test to twelve-year-olds (Lubinski & Benbow, 2006).

### What it looks like

Two steps. **First,** a grade-level test; students who score at the 95th percentile or higher advance to the second step, the usual requirement in talent search programs. **Second,** a standardized test designed for a grade two to four years higher, in the subject in which acceleration is being considered. The result is then interpreted against the norms of the higher grade and by comparing the student with same-age peers who took the same test.

### A scene from administration

After Salma's result came in (Figure 12.3), the counselor did not settle for calling her "very advanced." He asked for an analysis of her answers on the ninth-grade test to find out what she did not yet know. A learning plan was then drawn up covering only those topics, followed by an end-of-course test in an advanced course to confirm her mastery before she moved on to a higher course. This is a well-known model in talent search programs called **Diagnostic Testing → Prescriptive Instruction** (**Diagnostic Testing → Prescriptive Instruction, DT→PI**), which rests on the principle that students should not be taught what they have already shown they know (Assouline et al., 2015).

**Table 12.8: Profile of above-level testing.**

| | |
|---|---|
| **Purpose** | Differentiating among students who sit at the ceiling of a grade-level test, and determining how much acceleration they need in a particular subject |
| **Ages** | Typically Grades 3 through 9 in talent search programs |
| **Edition and year** | A standardized test for a higher grade, chosen by the program |
| **Publisher** | Depends on the test used |
| **Qualification level** | Administered by whoever is qualified to administer the original test under its conditions; interpreted by a professional familiar with the norms of the higher grade |

| | |
|---|---|
| **Time** | The duration of the test used |
| **Score types** | Percentile ranks based on the norms of the higher grade and on the norms of same-age program participants |
| **Cutoff scores** | The 95th percentile or higher on the grade-level test is a common entry requirement; the program then sets the cut points for each service |
| **Reliability** | The reliability of the test used, with the advantage that its difficulty suits high performers and thus discriminates among them better |
| **Classification accuracy** | Differences on above-level tests predicted large differences in accomplishment decades later (Lubinski & Benbow, 2006) |
| **Year of norm collection** | The higher-grade norms of the test used |
| **Arabic version** | At the time of writing, we know of no published Arabic above-level testing program with publicly available norms. The idea can be applied using Arabic achievement tests for higher grades, provided this is stated explicitly and the result is interpreted against the norms of the grade for which the test was designed |

### How are the results interpreted?

The percentile rank here compares the student with older students, so a rank of 50 for a sixth grader on a ninth-grade test is an excellent performance for his or her age. The question the result answers is: "How far ahead of grade level is the student in this subject? And what does the student know of it, and what not yet?"

### Its limits

It measures a specific subject and so cannot serve as evidence of general giftedness. It may also be affected by content the student has not yet been taught, so low performance should be read as lack of exposure before it is taken as lack of ability. It must also lead to an actual educational decision; otherwise it becomes a number that changes nothing. The literature on acceleration summarizes extensive evidence of its academic benefit when it is grounded in assessment of this kind (Assouline et al., 2015).

## 12.7 Identification in the Saudi context and the decision pathway

The **National Program for Gifted Identification** began in 2011 as an annual project run in partnership by the King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba), the Ministry of Education, and the National Center for Assessment, initially targeting students in the third and sixth grades of elementary school and the third grade of intermediate school (Ministry of Education, 1437 AH [Arabic]). Today it targets students from the third grade of elementary school through the first grade of secondary school, that is, grades 3 to 10, with its main instrument, the **Mawhiba Multiple Cognitive Abilities Test** (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]). Alongside it are the Ministry of Education's school-based gifted programs, which have their own criteria. A recent review offers a comprehensive picture of gifted education and its programs in the Kingdom (Alfaiz et al., 2022). The box below sets out what the published documents stipulate on identification specifically.

**In the Saudi context: identification criteria in the published documents**

**Table 12.9: Criteria for identifying gifted learners in the published Saudi documents we were able to review.**

| Document | What it stipulates | Reading it with the chapter's tools |
|---|---|---|
| **Regulations for Special Education Institutes and Programs (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 8)** | A student is admitted if any one of the following conditions is met: high academic achievement over two semesters (97%, by our reading of the printed copy) together with an IQ of at least 115 on a standardized test; an IQ of at least 125 on the Wechsler scale or 127 on the Stanford–Binet; outstanding skills and abilities attested by the student's teachers and peers; or creative or innovative thinking | An "or" rule: one criterion suffices. The cutoff of 115 corresponds to roughly the top 16% and requires achievement alongside it; the cutoff of 125 corresponds to roughly the top 5%. The difference between 125 and 127 is explained by the standard deviation of 16 in older editions of the Stanford–Binet. The rule specifies neither an instrument nor a score for creativity |
| **Special Education Organizational Guide (Ministry of Education, 1436 AH [Arabic])** | Does not list the gifted among its categories | Regulation of giftedness moved to the gifted program guides, so it is no longer housed in a single document with disabilities |

| Document | What it stipulates | Reading it with the chapter's tools |
|---|---|---|
| **Guide to Gifted Classes (Ministry of Education, n.d. [Arabic])** | Students are nominated from among those classified in the national gifted identification project. If places remain unfilled, they are completed using the Ministry's four criteria: a cognitive abilities measure, personal traits, academic achievement, and creative output, provided the student meets at least two criteria, one of which must be the cognitive abilities measure with a score of no less than 120 | A compound rule: at least two criteria, with ability as a necessary condition. The cutoff of 120 corresponds to roughly the top 10%, approximately Gagné's threshold |
| **Organizational and Procedural Guide for the School Program for Gifted Girls (Ministry of Education, 1437 AH [Arabic])** | Admission for the top 5% of the national project's results, from fourth grade of elementary school to second grade of secondary school | The percentage is computed only among those who sat the test, since registration is optional |

| Document | What it stipulates | Reading it with the chapter's tools |
|---|---|---|
| **National Program for Gifted Identification (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic])** | The Mawhiba Multiple Cognitive Abilities Test, with optional registration; results are announced through the Mawhiba website | We found neither the score scale nor the cutoff score in any official source. A peer-reviewed study reported an operational cutoff of 1540 (Almulhim, Aljughaiman, & Alfaiz, 2026) |

Three observations arise from these documents. **First,** the older regulations treated the criteria as alternatives, any one of which suffices, whereas the gifted classes guide made them cumulative, requiring two criteria, one of which is ability. An "or" rule widens the door and increases errors of the "identifying a student who is not gifted" kind, while an "and" rule narrows it and increases errors of the "missing a gifted student" kind (Chapter 6 presents these combination rules). **Second,** the documents use three IQ cutoffs, 115, 120, and 125, without stating which level of giftedness each is meant to capture, and without any provision directing that the standard error of measurement or the confidence interval be taken into account. A student with a score of 118 and a confidence interval of 112–123 does not differ statistically from a student who scores 120. **Third,** in the documents we reviewed, we found no provision on adapting identification instruments for students with disabilities, nor any provision permitting or prohibiting simultaneous enrollment in a gifted program and special education services, even though a recent Saudi study estimated that about 5.9% of a sample of 6,875 students in four private schools in Riyadh may be twice-exceptional (Almulhim, Aljughaiman, Alnaim, et al., 2026).

**A question for reflection:** A student scored 126 on the Wechsler scale but did not sit the Mawhiba test because her family did not know when registration opened. Under which of the three documents would she be admitted, and under which would she not? And what does that say about the effect of the national instrument on access?

## MMCAT Mawhiba Multiple Cognitive Abilities Test

National ability test

### The idea

A single national instrument taken by applicants from every region, measuring reasoning in more than one domain and opening the way to gifted programs at Mawhiba and the Ministry of Education. It was built in Arabic and for the Saudi context from the outset, so it carries none of the problems of translation and adaptation.

### What it looks like

The test measures four domains: mental flexibility, linguistic reasoning and reading comprehension, mathematical and spatial reasoning, and scientific and mechanical reasoning (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]; Almulhim, Aljughaiman, & Alfaiz, 2026). It has three levels: the first for grades 3 to 5 of elementary school, the second for grade 6 of elementary school and the first and second grades of intermediate school, and the third for the third grade of intermediate school and the first grade of secondary school. A student may take it once at each level.

### A scene from administration

During the announced registration period, the parent opens an account on the National Center for Assessment website, confirms the child's details, selects the test, the venue, and the date, accepts the terms, and pays the parental contribution. After the test, the parent waits for the results to be announced, then looks them up on the Mawhiba website and reviews the programs for which the child qualifies.

**Table 12.10: Profile of the Mawhiba Multiple Cognitive Abilities Test, based on what we found published at the time of writing.**

| | |
|---|---|
| **Purpose** | Identifying students with high cognitive abilities on a national scale and nominating them for gifted programs |
| **Ages** | Grade 3 of elementary school through Grade 10 (first year of secondary school), at three levels |
| **Edition and year** | Used in the national program since its first cycle in 2011. We found no information on its current edition |
| **Administering body** | The Education and Training Evaluation Commission, through its National Center for Assessment, in partnership with Mawhiba and the Ministry of Education |
| **Registration** | Optional, through the National Center for Assessment website, with a "parental contribution" fee of SAR 200, waived for subsidized students. The most recently announced registration period ran from 1 October to 23 December 2025 |
| **Score reporting** | Through the national program results inquiry service on the Mawhiba website |
| **Score types and cut points** | We found neither the score scale nor the cutoff score in any official source. A peer-reviewed study of 5,146 students in Al-Ahsa reported an operational cutoff of 1540 (Almulhim, Aljughaiman, & Alfaiz, 2026) |

| | |
|---|---|
| **Consequences** | Mawhiba classes, the Mawhiba Advanced Program in Science and Mathematics, research and enrichment programs, and nomination for university scholarships; it is also the starting point for nomination to gifted classes and Ministry programs |
| **Reliability and validity** | We found no technical manual or published report presenting reliability coefficients or validity evidence. The study cited above analyzed the test's scores using latent profile analysis and found distinct cognitive profiles beneath the total score among high-ability students (Almulhim, Aljughaiman, & Alfaiz, 2026) |
| **Accommodations for students with disabilities** | We found no published statement of the accommodations available to students with disabilities |
| **Year of norm collection** | We did not find these published |
| **Arabic version** | Originally developed in Arabic, for the Saudi context |

### How are the results interpreted?

Because the official scale is unpublished, the practitioner reads the result from two angles: Did the student exceed the threshold that qualifies him or her for a particular program? And where does the student rank among applicants from the same grade? Note that the comparison here is only with those who sat the test. If the four domain scores are available, the pattern among them matters more for service decisions than the total score alone.

### Its limits

This is a reading of the test in light of the principles in Section 12.2, and fairness requires stating what it achieves before what it does not. It **achieves** a uniform instrument across regions, measurement of more than one domain of reasoning, an authentically Arabic construction, a repeated opportunity at each level, and fee waivers for subsidized students. In its published form,

however, it **does not achieve** four things. **First,** optional registration is not universal screening: students whose families do not know about the test, or do not consider it important, never reach it, which is precisely what studies of universal screening have warned against (Card & Giuliano, 2016). **Second,** admission to programs rests on a single cutoff score, whereas the principles call for multiple sources. **Third,** the norms are national, so a student from a small village is compared with a student from a large city. **Fourth,** its technical evidence and its accommodations for students with disabilities are unpublished, so practitioners cannot judge the precision of a score near the cutoff. The Ministry's programs fill part of these gaps: the gifted classes guide allows unfilled places to be completed using multiple school-based criteria, among them personal traits and creative output. But this completion depends on places remaining unfilled, and the national test remains the main gateway.

Table 12.11 summarizes what we know about normed Arabic versions of the instruments discussed in this chapter:

**Table 12.11: Arabic instruments available for identifying gifted learners, based on what we found published at the time of writing.**

| Instrument | What we know about its Arabic version | What is missing |
|---|---|---|
| **Mawhiba Multiple Cognitive Abilities Test** | Originally Arabic; in national use since 2011 | A published technical manual, the score scale, the norm year, and accommodations |
| **Wechsler Intelligence Scale for Children** | The revised edition for ages 6–16 was normed as part of the Gifted Identification and Care Program project (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024) | A published Saudi norming of the fourth and fifth editions, and hence of the General Ability Index and the Cognitive Proficiency Index |

| Instrument | What we know about its Arabic version | What is missing |
|---|---|---|
| — | Saudi norms for Figural Form B, based on 1,227 male and female students aged 9 to 16 (Al-Nafi, 1428 AH [Arabic]) | Updated norms, publication of the original norming report, and norming of the verbal form |
| **Gifted Rating Scales, Scales for Rating the Behavioral Characteristics of Superior Students, and observation tools** | We found no published Saudi norming study for any of them | Norming, or local norms built by the program |
| **Cognitive Abilities Test and Naglieri Nonverbal Ability Test** | We know of no published standardized Arabic version | Full adaptation and norming |

## The decision pathway

Figure 12.5 brings the preceding material together in a general pathway suited to a school or regional program, in which the Mawhiba test can serve as the screening instrument in the second step.

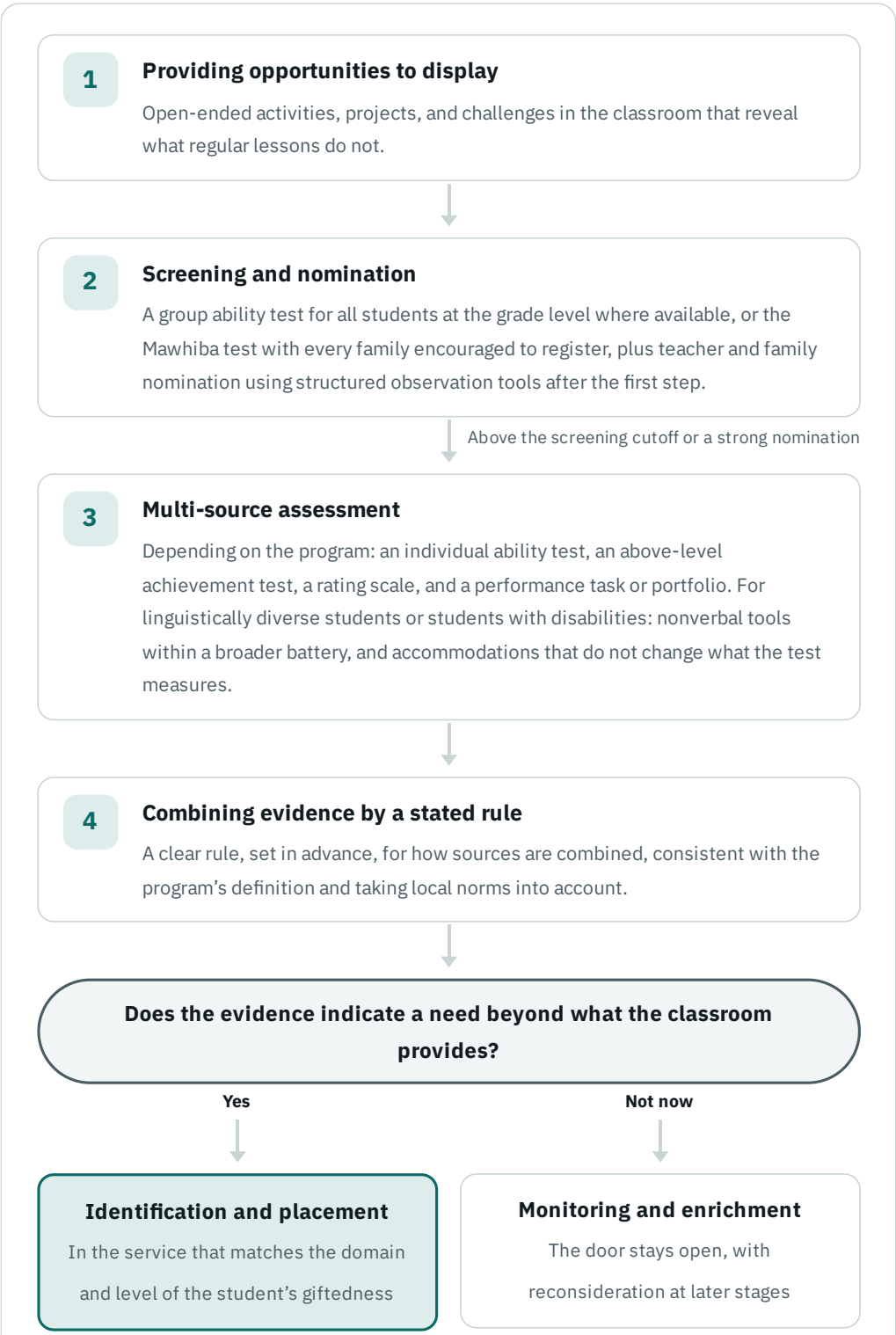

1
Providing opportunities to display
Open-ended activities, projects, and challenges in the classroom that reveal what regular lessons do not.
2
Screening and nomination
A group ability test for all students at the grade level where available, or the Mawhiba test with every family encouraged to register, plus teacher and family nomination using structured observation tools after the first step.
Above the screening cutoff or a strong nomination
3
Multi-source assessment
Depending on the program: an individual ability test, an above-level achievement test, a rating scale, and a performance task or portfolio. For linguistically diverse students or students with disabilities: nonverbal tools within a broader battery, and accommodations that do not change what the test measures.
4
Combining evidence by a stated rule
A clear rule, set in advance, for how sources are combined, consistent with the program's definition and taking local norms into account.
Does the evidence indicate a need beyond what the classroom provides?
Yes
Not now
Identification and placement
In the service that matches the domain and level of the student's giftedness
Monitoring and enrichment
The door stays open, with reconsideration at later stages

**Figure 12.5:** A general pathway for identifying gifted learners. Note that it begins by providing opportunities before screening, that the combination rule is announced before the evidence is gathered, and that it ends with a placement matched to the type of giftedness rather than a global verdict of "gifted" or "not gifted."

## 12.8 When is the judgment made? Three cases

### Case 1: Jana, age 10 General giftedness with a strength in science

**The rest of the opening scene:** Jana is a fourth grader whom her teachers did not nominate because she daydreams a lot and her spelling grades are average. After the science teacher mentioned her soil project, she proposed to the school that all the girls in the grade be offered open-ended activities for two weeks, after which the teachers would complete a structured observation checklist. Jana stood out in curiosity, reasoning, and inquiry. The school then encouraged all the students' families to register for the Mawhiba test, and her family registered her at the first level. The figures in the table are hypothetical, for illustration.

**Table 12.12: Evidence in Jana's case.**

| Source | Result | What it means |
|---|---|---|
| **Observation checklist after the open-ended activities** | High in curiosity, reasoning, and inquiry; average in self-discipline and organization | Gifted behaviors that emerged once the opportunity was provided |
| **Mawhiba Multiple Cognitive Abilities Test** | 97th percentile among applicants from her grade; her strongest domain was scientific and mechanical reasoning | High reasoning relative only to those who sat the test nationally |

| Source | Result | What it means |
|---|---|---|
| **Seventh-grade science test, administered to her in fourth grade** | 72nd percentile on seventh-grade norms | At least three years ahead of her grade in science |
| **Spelling** | Standard score of 98 (average) | Not inconsistent with giftedness, and does not call for further assessment at this point |

**Decision:** Jana is identified as gifted, with a particular strength in science, and is placed in a science enrichment program with acceleration in science. Her daydreaming is not settled for now. Eight to twelve weeks after she joins the program, she is observed systematically in a regular lesson and in one that offers suitable challenge, and the question is asked: does the daydreaming diminish when she meets a challenge? Yet this is a **supporting question, not a decisive one**. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (**DSM-5-TR**), states that signs of attention-deficit/hyperactivity disorder may be minimal or absent when the child is receiving frequent rewards for appropriate behavior, is under close supervision, is in a novel setting, or is engaged in especially interesting activities (American Psychiatric Association, 2022). If the daydreaming persists across more than one setting and visibly affects her organization and performance at home and at school, an assessment for the disorder or for anxiety is completed as a possible case of twice-exceptionality, not as an alternative to giftedness (Chapters 10 and 13 take up this distinction in detail).

> **What evidence could change the decision?** If her performance in the advanced science course falls short of what the above-level test result suggests, the team would revisit the amount of acceleration and the

> starting point. If the daydreaming persists at home and at school after she has found challenge, with effects on her organization and performance, an assessment for possible twice-exceptionality would be added to the giftedness decision.

## Case 2: Mishari, age 11 Local norms bring him into view

**The story:** Mishari is a fifth grader at a small school in a remote village. There are 18 students in his class, the school has a basic laboratory, and there is no regular enrichment activity. His score on the group ability test fell below the national admission cutoff for one of the programs. Yet he is the top student in his school by a wide margin, and from parts of old devices he has built a robotic arm that moves on simple commands, which he taught himself to program from online videos. The figures in the table are hypothetical, for illustration.

**Table 12.13: Evidence in Mishari's case.**

| Source | Result | What it means |
|---|---|---|
| **Group ability test, national norms** | 116, i.e., the 86th national percentile; the program's cutoff is the 95th percentile | Below the cutoff under the national-norms rule |
| **The same score among students in his grade at his school** | First of 18 students | An important indicator, but unstable: in a group of this size, each rank is separated from the next by about six percentiles |

| Source | Result | What it means |
|---|---|---|
| **The same score on local norms for rural schools in the education directorate (420 fifth graders, two years combined)** | 97th percentile | Within the top 3% of students who had comparable opportunities, a more stable comparison than with his class alone |
| **Portfolio and evaluation by a panel of three specialists, each rating independently** | The robotic arm project with documentation of its steps; close agreement among the judges on originality and craftsmanship | Advanced design and programming ability, and self-directed learning |
| **Individual nonverbal intelligence test** | 124 (confidence interval 117–129) | High reasoning that supports the other evidence but does not replace it |

**Decision:** Had a single rule been applied with national norms, Mishari would have been missed. Had the team relied only on his rank in his class, the judgment would have hinged on one or two students. The team therefore made the rural schools in the directorate its reference group, combined two years of data to enlarge it, and stated this in the report. It then combined the local norms, the portfolio, and the individual test, identified him as gifted in the technical domain, and recommended placement in a remote enrichment program with mentoring from a robotics specialist. The report cautioned that services must start from his actual level in programming and mathematics as revealed by the evidence, because the absolute level of students identified through local norms varies. This applies the idea of local norms, which compare students with others who have had comparable opportunities, and the phrase “their environment” in the definition of giftedness.

**What evidence could change the decision?** If his local rank dropped markedly once a third year was added to the reference group, or if his performance in the enrichment program fell short of what his portfolio suggests, the team would revisit both the placement and the level of service.

## Case 3: Dana, age 9

A high achiever who needs a different service

**The story:** Dana is in third grade, and all her teachers nominated her for the gifted program for girls. She has been top of her class since first grade, her marks are perfect, and she is highly organized and hardworking. The figures in the table are hypothetical, for illustration.

**Table 12.14: Evidence in Dana's case.**

| Source | Result | What it means |
|---|---|---|
| **Teachers' rating scales** | T-scores of 68 in motivation, 66 in academic ability, and 54 in creativity | A high-achieving, persistent student |
| **Individual ability test** | General Ability Index 114 (confidence interval 108–119) | Above average, with the entire interval below the usual threshold for general gifted programs |
| **Sixth-grade Arabic language test, administered in third grade** | 68th percentile on sixth-grade girls' norms | Language achievement three years ahead of her grade |

| Source | Result | What it means |
|---|---|---|
| **Fifth-grade mathematics test, administered in third grade** | 46th percentile on fifth-grade girls' norms | Advanced in mathematics, though to a lesser degree |

**Decision:** Dana did not meet the criteria for the general gifted program for girls, which requires very high general ability. Yet the team did not stop at saying she is "not gifted." Her language achievement, three years ahead, means that her grade's curriculum does not challenge her in language. The team therefore recommended **acceleration** (**Acceleration**) in Arabic, meaning that she study it with a higher grade or through an advanced curriculum, together with an enrichment activity in creative writing. Nor did the team rely on the above-level test result alone to set her starting point. It analyzed her answers to find out which parts of the higher grades' curriculum she had mastered and which she had not, then built a plan addressing only what she had not yet mastered, with the requirement that she pass the end-of-course test in the advanced course before moving beyond it, following the Diagnostic Testing → Prescriptive Instruction model (see the profile of above-level testing in 12.6). This is what aligning identification with services means: the question is not "Is she gifted?" but "What service does she need?"

> **What evidence could change the decision?** If a later reassessment of her ability yielded a General Ability Index whose entire confidence interval lay above the program threshold, or if independent judges rated her creative writing products highly, the team would reconsider placing her in the general gifted program or in a specialized language program.

## 12.9 What can be mistaken for giftedness, and what can hide it

**Table 12.15: Situations in which giftedness may be wrongly suspected or hidden, and how the team handles them.**

| Situation | What happens | How we handle it |
|---|---|---|
| **The high achiever** | High marks earned through effort and discipline, with above-average but not exceptional ability | Needs challenge in the subjects where he or she excels, which may suffice without a general gifted program, as in Dana's case |
| **The gifted underachiever** | Very high ability with average or low achievement, owing to boredom or to emotional or environmental factors | Detected through ability measures, since achievement conceals it; needs support for motivation alongside challenge |
| **Giftedness that resembles a disorder** | Boredom, daydreaming, repeated questions, and intense sensitivity may be misread as attention-deficit/hyperactivity disorder or an emotional disorder (Webb et al., 2005) | The student is observed systematically in an ordinary lesson and in one that offers suitable challenge, and a **supporting question** is asked: does the behavior diminish when the student is challenged? Its disappearance does not rule out the disorder, since symptoms may lessen during interesting or novel activities, under close supervision, or with frequent rewards (American Psychiatric Association, 2022). It is therefore combined with a developmental history, ratings from more than one informant across more than one setting, and evidence of functional impact at home and at school |

| Situation | What happens | How we handle it |
|---|---|---|
| **Giftedness with a disability (twice-exceptional learners)** | The disability masks the giftedness, the giftedness masks the disability, or each masks the other so that the student appears average | A comprehensive assessment that looks for strengths and weaknesses together and compares the General Ability Index with the Cognitive Proficiency Index rather than relying on the full-scale score alone. A Saudi study estimated that about 5.9% of a sample of 6,875 students in four private schools in Riyadh may belong to this group (Almulhim, Aljughaiman, Alnaim, et al., 2026) (Chapter 13 treats it in detail) |
| **The gifted student who is linguistically diverse or from a low-opportunity environment** | Lower scores on verbal tests and on national norms | Local norms, performance assessment, and a nonverbal test within a battery that includes verbal and quantitative reasoning (Lohman, 2005), as in Mishari's case |

# Conclusion

The central argument of this chapter is that the definition of giftedness determines the tools of identification, and that every definition and every admission rule determines, at the same time, who may be missed. It follows that teacher nomination should not be relied on alone, since it is limited by teachers' biases and by the opportunities to display that the classroom affords; that universal screening and local norms should widen access, with due attention to their conditions; that a nonverbal test should form part of a broader battery; and that above-level testing should be used when the ceiling limits what we can read of a student's ability. Creativity tests serve as a complement, alongside the evaluation of real products.

Some questions in the chapter nonetheless remain contested, and the evidence on others has yet to be published. The predictive validity of creativity tests is disputed. The Mawhiba test is a national instrument developed in Arabic, but registration is optional, and its score scale and technical evidence are unpublished. Saudi documents use different IQ thresholds (115, 120, and 125) with different combination rules. Distinguishing giftedness from a disorder remains a judgment that no single question can settle: the disappearance of a behavior under challenge is a supporting clue that needs others alongside it. And identification extends beyond the question "Is the student gifted?" to the service that suits the domain and level of the student's giftedness.

## 12.10 Summary, key terms and activities

### Chapter summary

- The definition of giftedness determines the tools of identification; contemporary definitions encompass both potential and achievement and take the environment into account. Each of the four models has those it may miss.
- Gagné distinguishes natural gifts from developed talents and sets a threshold at the top 10%, with five levels that help in reading admission rules.
- Teacher nomination is limited by teachers' biases and by the opportunities the classroom affords for giftedness to be displayed. Universal screening and local norms widen access; local norms carry conditions on the size of the reference group and limits on how far the label transfers.
- A nonverbal test is part of a broader battery. The General Ability Index is useful with uneven profiles, and its difference from the Cognitive Proficiency Index is read in terms of significance and

rarity. Above-level testing solves the ceiling problem and sets the starting point for acceleration.

- Creativity tests measure divergent thinking; their predictive validity is disputed, and they are used as a complement alongside the evaluation of real products.
- The Mawhiba test is a national instrument developed in Arabic, but registration is optional and its score scale and technical evidence are unpublished; Saudi documents use different IQ thresholds (115, 120, and 125) with different combination rules.
- In distinguishing giftedness from a disorder, the disappearance of a behavior under challenge is a supporting question that needs further evidence.
- The most important question after "Is the student gifted?" is "What service suits the domain and level of the student's giftedness?"

## Glossary of terms

| | |
|---|---|
| **Gift** | High natural ability not yet refined through learning and practice. |
| **Talent** | A mastered skill in a specific field, developed from gifts through learning and practice. |

| | |
|---|---|
| **Opportunity to display** | The situations a learning environment offers the student in which to display his or her abilities. |
| **Universal screening** | Assessing all students at a given grade level rather than relying on nomination. |

| | |
|---|---|
| **Ceiling effect** | A test's inability to discriminate among high-performing students because it contains too few difficult items. |
| **Above-level testing** | Administering a test designed for older students to an able student in order to gauge how far ahead he or she is. |
| **General Ability Index**<br>General Ability Index, GAI | A composite score measuring reasoning and comprehension without working memory or processing speed. |
| **Cognitive Proficiency Index, CPI** | A composite score drawn from working memory and processing speed subtests, compared with the General Ability Index to describe an uneven profile. |
| **Talent pool** | A broad group of students, about 15–20%, offered enrichment services in Renzulli's model, with flexible entry for those who display gifted behavior. |
| **Excellence gaps** | Differences between groups in the proportions of students who reach advanced levels of performance. |
| **Consensual Assessment Technique** | Evaluating a product's creativity through independent ratings by judges familiar with the field, then examining their agreement. |
| **DT→PI** | A model used in talent search programs that pinpoints what a student does not yet know of an advanced course, teaches only that material, and then tests the student at the end of the course. |

| | |
|---|---|
| **Divergent thinking** | The capacity to generate many varied and original ideas for a single problem. |
| **Three-Ring Conception** | Renzulli's conception of gifted behavior as the intersection of above-average ability, task commitment, and creativity. |
| **Local norms** | Comparing a student's performance with that of peers in the same school or area rather than with the national norming sample alone. |
| **Creative-productive giftedness** | The capacity to produce original ideas and products for a real audience; Renzulli contrasts it with schoolhouse giftedness, that is, skill at learning lessons and taking tests. |
| **Acceleration** | Presenting content to a student at a younger age or a faster pace than usual, such as studying a subject with a higher grade. |
| **Twice-exceptional** | Gifted students who also have a disability or disorder. |
| **Gifted underachiever** | A student whose ability is very high and whose achievement falls clearly below it. |

## Review questions

1. A program offers robotics enrichment to intermediate school students. Which of the four models in Table 12.2 best suits its definition of giftedness? What tools does that model call for, and whom might it miss?
2. An education directorate admits to its gifted program students who are nominated by their teachers and score above 130 on an individual intelligence test, and it does not reconsider those who are not admitted. Identify three principles in Table 12.3 that this procedure violates, and propose a modification for each.
3. What is meant by the opportunity to display and by inference load? Why might a teacher fall back on nominating high achievers when such opportunities are absent?
4. A student has a General Ability Index of 132, a Cognitive Proficiency Index of 104, and a full-scale IQ of 121. What two questions would you ask about the difference between the two indexes before interpreting it? Which score would you use to judge his eligibility for a program that requires 125, and why?
5. Two students scored at the 99th percentile on a grade-level mathematics test. How would you determine which of them needs acceleration? And what would be added afterward to set the starting point in the advanced course?
6. What are the limits of creative thinking tests? How does the Consensual Assessment Technique complement them? And why do the available Saudi norms for the Torrance test call for particular caution on the originality dimension?
7. A student scored 117 (confidence interval 111–122) on the cognitive abilities test, her achievement was 98% over the last two semesters, and her creative product was rated high. What decision does the rule of the Regulations (1422 AH) yield for her, and what decision does the rule of the gifted classes

guide yield? And what does the confidence interval say about the difference between her score and the threshold of 120?

8. Evaluate the Mawhiba test against the identification principles in 12.2: which two principles does it satisfy, and which two does it fail to satisfy in its published form? How could a school program you know fill one of these two gaps?

9. A fourth grader daydreams in class and asks questions far removed from the lesson; his mother says he spends hours building model airplanes without looking up. His teacher concluded that his focus during an open-ended science activity rules out attention-deficit/hyperactivity disorder. Evaluate her reasoning, and identify the evidence you would need before reaching any judgment.

10. **Case for application:** Thamer is fifteen and in tenth grade. For the past two years he has been making short animated films, one of which won second place in a regional competition for secondary school students. His school grades range from good to very good. On the Mawhiba test, at the third level, he scored at the 81st percentile among applicants from his grade, and so was not admitted to the advanced program. The art teacher who taught him in intermediate school completed the school form of a gifted rating scale: his T-score for artistic talent was 72, and his scores on the other scales ranged from 50 to 56. He was also given an individual intelligence test, with a full-scale score of 117 (confidence interval 111–122). The school principal wants a decision: should he be nominated for a gifted program? (a) Which of the four models would detect Thamer, and which would miss him? (b) What are the two problems with relying on the gifted rating scales here? (c) What missing evidence must be gathered before a judgment can be made in his domain of talent, and how should it be gathered? (d) What is your decision now, and what service do you propose? Justify your decision.

**Table 12.16: Alignment of chapter objectives with review questions and learning activities.**

| Objective | Assessed by |
|---|---|
| **1. Compare the four models of giftedness** | Question 1; part (a) of Question 10 |
| **2. Determine what is consistent with the principles of identification and what violates them** | Question 2; Activity 2 |
| **3. Design opportunities to display and determine the effect of their absence** | Question 3; Activity 1 |
| **4. Interpret the results of ability and creative thinking measures** | Questions 4, 5 and 6; Activity 3 |
| **5. Apply and critique the Saudi criteria** | Questions 7 and 8; Activity 2 |
| **6. Judge the sufficiency of evidence and distinguish the gifted student from the high achiever and from the student with a disorder** | Question 9; Question 10 (case for application) |

## Learning activities

Any activity in which an activity or instrument is administered to students, or in which a real student's data are used, is carried out with the consent of the family and the school administration, without any information that could identify the student, and under the supervision of a qualified specialist. Its results are not used in any decision concerning the student and are reported only through official channels.

1. **A week of opportunities to display:** Design five open-ended activities for a class you teach or are able to observe, and before administering them, write a list of the students you would have nominated for a gifted program. Then administer the activities under the conditions above, record, using codes

rather than names, the students who showed indicators you had not expected, and compare the two lists. If no class is available to you, apply the steps to Jana's scene and Thamer's case. *Product: five written activities, a comparison table, and a reflective paragraph. Time: one week for administration and about two hours for writing.*

2. **A program's definition:** Choose a gifted program in your area, such as gifted classes or one of Mawhiba's programs, and extract from its published documents its definition of giftedness, its tools, its thresholds, and its combination rule. Then compare these with Tables 12.3 and 12.9, identify whom the program may miss, and propose one feasible modification. *Product: a table and a paragraph. Time: about two hours.*
3. **A creative activity:** Design an activity in the style of Figure 12.4, using an incomplete geometric shape other than a semicircle. Administer it to ten of your classmates, or to students under the conditions above, and score the responses for fluency, flexibility, originality, and elaboration. Then ask two colleagues to rate the creativity of the same responses independently using the Consensual Assessment Technique, and compare the three sets of ratings. *Product: the activity, a table of ratings, and a paragraph on interrater agreement. Time: about two hours.*

## Chapter references

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Al-Nafi, A. (1428 AH). *Creative thinking* [Slide presentation] [In Arabic]. King Saud University. https://faculty.ksu.edu.sa/sites/default/files/_ltfkyr_lbdy-d.bdllh_lnf_0.pdf

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (1437 AH). *Organizational and procedural guide for the school program for gifted girls* [In Arabic]. General Directorate for Gifted Girls. https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20(1).pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alfaiz, F. S. (2026). The accuracy of parent and teacher nominations in gifted identification in Saudi Arabia. *Cogent Education, 13*(1), Article 2654922. https://doi.org/10.1080/2331186X.2026.2654922

Alfaiz, F. S., Alfaid, A. A., & Aljughaiman, A. M. (2022). Current status of gifted education in Saudi Arabia. *Cogent Education, 9*(1), Article 2064585. https://doi.org/10.1080/2331186X.2022.2064585

Almulhim, N., Aljughaiman, A., Alnaim, F., Alsayed, A., Alfawzan, S., & Alarfaj, A. (2026). Hidden potential: An exploratory study of the prevalence of twice-exceptional students in Saudi Arabia. *Journal of Intelligence, 14*(5), Article 83. https://doi.org/10.3390/jintelligence14050083

Almulhim, N., Aljughaiman, A. M., & Alfaiz, F. S. (2026). Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis. *Journal of Intelligence, 14*(8), Article 166. https://doi.org/10.3390/jintelligence14080166

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997–1013.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Assouline, S. G., Colangelo, N., VanTassel-Baska, J., & Lupkowski-Shoplik, A. (Eds.). (2015). *A nation empowered: Evidence trumps the excuses holding back America's brightest students*. Belin-Blank Center.

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678–13683.

Coleman, M. R., Shah-Coltrane, S., & Harrison, A. (2010). *U-STARS~PLUS: Teacher's observation of potential in students*. Council for Exceptional Children.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Frasier, M. M., Hunsaker, S. L., Lee, J., Mitchell, S., Cramond, B., Krisel, S., García, J. H., Martin, D., Frank, E., & Finley, V. S. (1995). *Core attributes of giftedness: A foundation for recognizing the gifted potential of minority and economically disadvantaged students* (RM95210). University of Connecticut, The National Research Center on the Gifted and Talented.

Gagné, F. (2009). Building gifts into talents: Detailed overview of the DMGT 2.0. In B. MacFarlane & T. Stambaugh (Eds.), *Leading change in gifted education: The festschrift of Dr. Joyce VanTassel-Baska* (pp. 61–80). Prufrock Press.

Gagné, F. (2013). The DMGT: Changes within, beneath, and beyond. *Talent Development & Excellence, 5*(1), 5–19.

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open, 2*(1).

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Wiley.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal, 18*(1), 3–14.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111–138.

Lohman, D. F. (2011). *Cognitive Abilities Test* (Form 7). Riverside.

Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted, 29*(4), 451–484.

Lubinski, D., & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1*(4), 316–345.

Naglieri, J. A. (2016). *Naglieri Nonverbal Ability Test* (3rd ed.). Pearson.

National Association for Gifted Children. (2018). *Use of the WISC-V for gifted and twice exceptional identification* [Position statement].

National Association for Gifted Children. (2019a). *A definition of giftedness that guides best practice* [Position statement].

National Association for Gifted Children. (2019b). *2019 Pre-K–Grade 12 gifted programming standards*.

Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open, 5*(2).

Pfeiffer, S. I. (2015). *Essentials of gifted assessment*. Wiley.

Pfeiffer, S. I. (Ed.). (2018). *APA handbook of giftedness and talent*. American Psychological Association.

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales*. Pearson.

Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180–184.

Renzulli, J. S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 53–92). Cambridge University Press.

Renzulli, J. S., & Reis, S. M. (2014). *The schoolwide enrichment model: A how-to guide for talent development* (3rd ed.). Prufrock Press.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., Gavin, M. K., Reis, S. M., Siegle, D., & Sytsma Reed, R. E. (2010). *Scales for rating the behavioral characteristics of superior students: Technical and administration manual* (3rd ed.). Prufrock Press.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*(1), 3–54.

Terman, L. M. (1925). *Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children*. Stanford University Press.

Webb, J. T., Amend, E. R., Webb, N. E., Goerss, J., Beljan, P., & Olenchak, F. R. (2005). *Misdiagnosis and dual diagnoses of gifted children and adults*. Great Potential Press.

Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixson, D. D. (2019). Gifted students. *Annual Review of Psychology, 70*, 551–576.

Chapter 13

# 13 Twice-exceptional learners

A student who is gifted and has a disability at the same time: the phrase may sound contradictory to anyone who sees giftedness and disability as opposite poles. Yet reality shows them coexisting in a single student, each concealing the other, so that the student falls between two systems, neither of which sees him fully as he is.

---

**Scene from a meeting that never took place**

Imagine two meetings at the same intermediate school in the same week. In the first, the gifted program team reviews the list of students whose performance has slipped and proposes withdrawing Yasser from the program: “He didn’t hand in his research project report, and his written test scores aren’t what you’d expect from a student in the program.” In the second, the learning disabilities team discusses the list of struggling students, and Yasser’s name never comes up: “He reads well, his Arabic grades are average, and he doesn’t need services.”

The two teams never met. Had they done so, they would have seen a student who solves math problems taught in secondary school and explains the idea behind his project

> orally with a fluency that astonishes his teachers, then spends a full hour writing a single paragraph about it, riddled with errors.

Yasser is fourteen, in the third year of intermediate school (ninth grade), and has been in the gifted program since fourth grade. His teacher in the program noticed that the gap between what he says and what he writes widens year after year: he avoids writing reports, asks to present his project orally, and his papers are full of spelling errors no one would expect from a student at his grade level. In this chapter we show why his difficulty was nearly overlooked and how to build an assessment that captures both sides at once, and then we return to his full case in Section 13.8.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Define twice-exceptional learners using the two reference definitions, and identify the difference between them.
2. Classify a student in a given description into one of the three groups, and explain how each of his two sides masks the other.
3. Identify what each of the usual identification models misses, and compare the arguments for normative comparison and intra-individual comparison.
4. Design an assessment protocol that opens both the giftedness and the disability tracks together for a given student, and justify each of its components.
5. Interpret an uneven cognitive profile using the difference between the General Ability Index and the Cognitive Proficiency Index, and explain why no judgment can rest on it alone.

6 Determine, in a given case, which of three possibilities is most likely—giftedness only, giftedness with a disorder, or a disorder only—and state the deciding evidence.

7 Judge whether the evidence in a new case suffices for a determination, and justify that judgment.

## Introduction

In twice-exceptional learners, giftedness and disability coexist, each concealing the other, so that their performance often appears average and catches the attention of neither the gifted team nor the special education team. Their cases therefore fall into the zone where the usual models break down: a student may be proposed for withdrawal from the gifted program because the disability has lowered his performance, and excluded from learning disabilities services because his giftedness has lifted his achievement to the average range—just as the opening scene portrays through two meetings that never converged. Judgment in such cases requires applying the criteria for each side separately, reading the uneven profile systematically, and knowing when a discrepancy counts as evidence and when it does not.

We open the chapter by defining twice-exceptional learners and locating them within the book's chapters (Section 13.1), then present masking and the three groups (Section 13.2), the reasons the usual models fail these students (Section 13.3), and the debate over identification criteria (Section 13.4). We then take up the assessment protocol, the reading of an uneven cognitive profile, and the adaptation of identification tools for learners with sensory and physical disabilities (Section 13.5), followed by the decision pathway that opens both tracks together (Section 13.6) and the three possibilities in differential diagnosis (Section 13.7). We apply all of this to four cases (Section 13.8) and close with the Saudi and Arab context (Section 13.9).

## 13.1 Who are twice-exceptional learners?

Twice-exceptional (**Twice-exceptional, 2e**) learners are students who show potential for high performance or creative productivity in one or more domains and who, at the same time, have one or more disabilities under recognized eligibility criteria (Reis et al., 2014). The literature stresses that the combination of the two exceptionalities produces a unique situation that is more than the sum of the two conditions: giftedness changes how the disability presents, the disability changes how the giftedness presents, and the student needs instruction that attends to both sides at once (Baldwin et al., 2015). Two definitions are widely cited in the literature, and they serve different functions:

**Two reference definitions**

**The operational definition (Reis et al., 2014)**

It specifies **who** counts as twice-exceptional: a student who shows potential for high achievement or creative productivity in one or more domains and who has one or more disabilities according to recognized eligibility criteria, such as learning disabilities, speech and language disorders, emotional and behavioral disorders, physical disabilities, autism spectrum disorder, and health impairments such as attention-deficit/hyperactivity disorder. By tying the category to existing criteria, the definition is easy to apply in schools.

**The community of practice consensus definition (2014; in Baldwin et al., 2015)**

It describes **how** the two exceptionalities interact and **what** the student needs: exceptional ability may dominate the disability, may conceal it, or may itself be concealed by it. These students need specialized identification methods that take the interaction between the two sides into account; dual differentiated instruction that develops their talents while meeting their learning needs; concurrent services that include the social and emotional domain; and teachers prepared to work with them.

The first, then, is a definition for **eligibility**, answering the question "Is this student one of them?"; the second is a definition for **practice**, answering the question "What does that mean for how we assess and teach him?" The assessment team needs both: the first in order to decide, and the second in order to know where to look and why the usual tools fall short.

Their number is not precisely known, because definitions and identification methods vary widely across studies and because many of these students are never identified at all (Foley Nicpon et al., 2011). Among recent studies, one examined the cognitive and achievement characteristics of students from a U.S. national sample who were identified as potentially twice-exceptional (Maddocks, 2020). In the Kingdom, an exploratory study of 6,875 students in

grades 2 through 12 at four private schools in Riyadh estimated that about 5.9% of them may be twice-exceptional: 23.6% showed a discrepancy of more than 1.5 standard deviations between their performance in two subjects, and 24.5% of these then appeared possibly gifted (Almulhim et al., 2026). This is a **preliminary estimate** based on indicators of discrepancy in achievement, and it would need individual diagnostic assessment to confirm. It does not estimate the prevalence of the group in Saudi schools, but it does signal that the group may be less rare than is commonly assumed.

Recent literature favors planning for these students that begins with their talents and interests rather than their deficits, an approach known as the strength-based, talent-focused approach (Baum et al., 2014). The examples in this chapter include Yasser, gifted in mathematics with a writing difficulty; Joud, linguistically gifted with autism spectrum disorder; and Faris, gifted in mathematics with attention-deficit/hyperactivity disorder. Set against them is Muhannad, a gifted student some of whose characteristics resemble a disorder.

## Twice-exceptional learners across the chapters of this book

This chapter stands on its own, but twice-exceptional learners also appear in the chapters on specific categories, each viewed from the angle of that category. Readers who wish to go further can turn to the following places:

**Table 13.1: Forms of twice-exceptionality across the book's chapters, and what each place adds.**

| Location | Form of twice-exceptionality | What it adds |
|---|---|---|
| **Chapter 8, Section 8.8 (Case 4)** | An eleven-year-old gifted in mathematics, with a specific learning disorder in reading fluency and spelling accuracy | Limited intervention and response data, and the tension between the logic of ability comparison and the response to intervention model |

| Location | Form of twice-exceptionality | What it adds |
|---|---|---|
| **Chapter 9, Section 9.8 (Case 3)** | A twelve-year-old girl who excels in mathematics and science, with autism spectrum disorder at support level 1 | Choosing the high-functioning form of the autism rating scale, and the effect of girls' camouflaging of their traits |
| **Chapter 10, Section 10.10** | Attention-deficit/hyperactivity disorder with giftedness | What to ask when a disorder is suspected in a gifted student |
| **Chapter 12, Sections 12.4 and 12.9** | Giftedness concealed by a disorder, and giftedness that resembles one | The General Ability Index, the ceiling effect, and above-level testing |
| **Chapter 15, Section 15.9 (Case 3)** | A twice-exceptional intermediate school student undergoing reevaluation | Involving the student in decisions about accommodations and his plan |
| **This chapter, Section 13.8** | Giftedness with a writing difficulty, with autism, and with attention deficit, and giftedness that resembles a disorder | Judging whether the evidence suffices on both sides together |

## 13.2 Masking and the three groups

### How does each side mask the other?

One way to picture it is as two forces pulling the student's performance in opposite directions: giftedness lifts it, and the disability drags it down. What the teacher sees in the grades is the net result of this tug-of-war, and it may well be an average performance that draws no one's attention.

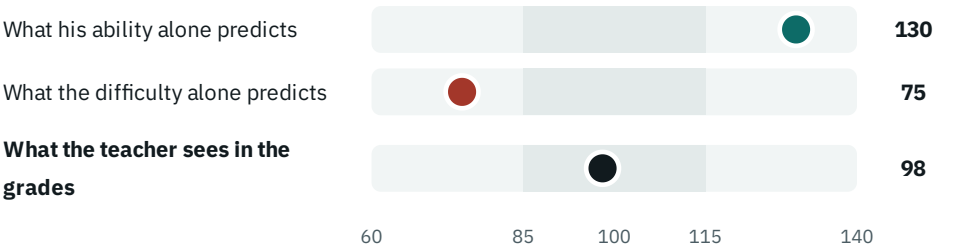



**Figure 13.1:** A hypothetical example of masking. The gray band is the average range, 85 to 115. High ability compensates for the difficulty, and the difficulty pulls performance down to a level that looks ordinary. The gifted team does not notice him because his performance is not high, and the learning disabilities team does not notice him because it is not low. The figure is a **conceptual representation**: the three numbers illustrate the direction of the pull between the two sides and are not meant to be added to or subtracted from one another.

## The three groups

Because of this masking, twice-exceptional learners appear in schools in three groups (Brody & Mills, 1997; Baum & Owen, 2004):

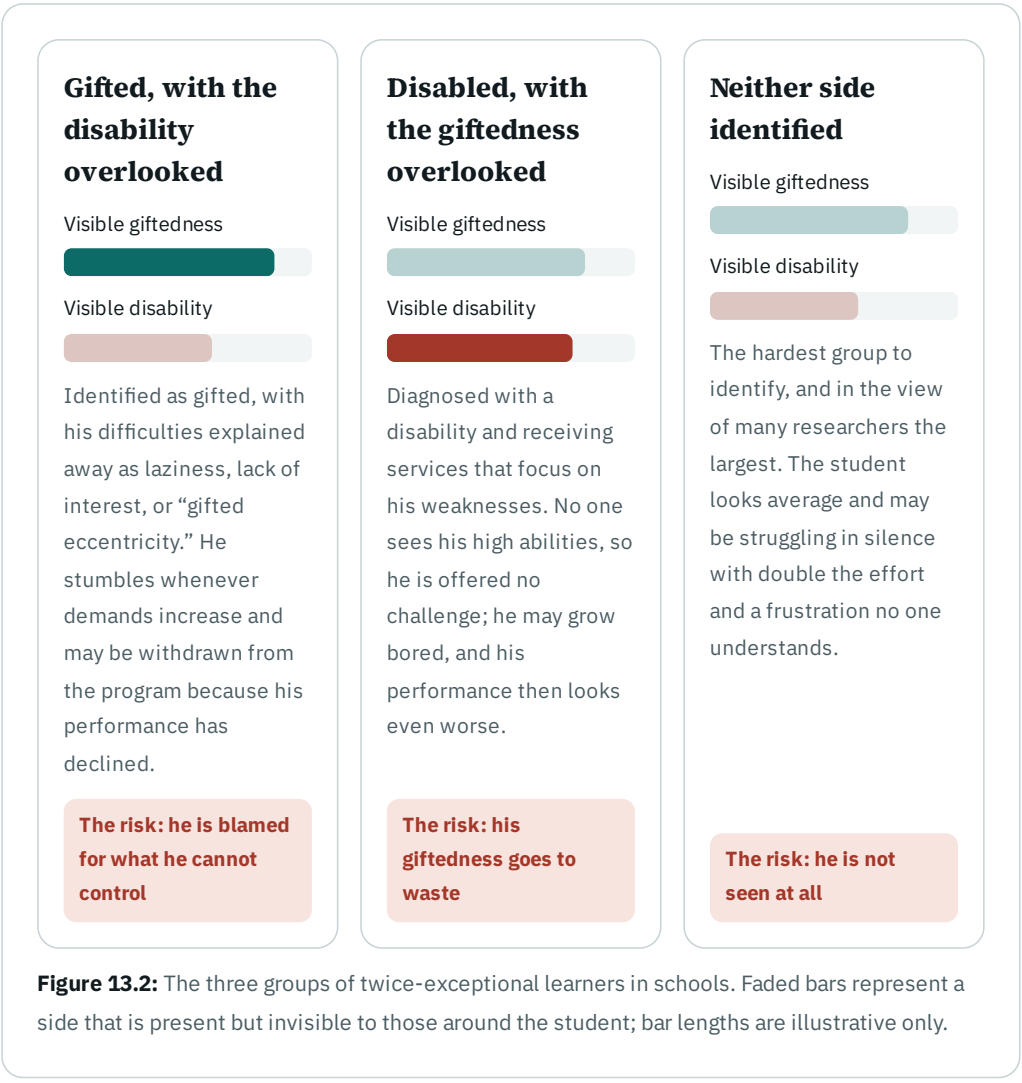



**Figure 13.2:** The three groups of twice-exceptional learners in schools. Faded bars represent a side that is present but invisible to those around the student; bar lengths are illustrative only.

The claim that the third group is the largest is a **common hypothesis** in the literature (Baum & Owen, 2004) that still awaits proof, since by their very nature we have no prevalence data on students who go unidentified. Yasser, in the scene that opens this chapter, belongs to the first group: he is in the gifted program, and his writing difficulty is invisible to anyone who hears him speak—so much so that he was nearly withdrawn from the program over the reports he never handed in. Had he never been nominated for the program in the first place, he would have belonged to the third group.

## 13.3 Why do the usual models fail them?

To identify learning disabilities and giftedness, schools rely on familiar models, among them response to intervention, which treats a student's failure to improve with instruction of graduated intensity as an indicator of difficulty; the criteria for specific learning disorder in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition, text revision (**DSM-5-TR**), which require the skill to be clearly below what is expected for the student's age (APA, 2022); and the admission criteria of gifted programs. When we examine these models, we find that each of them may miss the twice-exceptional student (Chapter 6 discusses identification models in detail). Researchers warned early on that relying on the response to intervention model alone may hide these students, because its screening looks for those who fall below their peers (Crepeau-Hobson & Bianco, 2011):

**Table 13.2: The usual identification models, what each requires, and why each may miss the twice-exceptional student.**

| Model | What it requires | Why it may miss him |
|---|---|---|
| — | Performance below peers that does not improve with intervention | His performance is at the level of his peers or slightly above, so he never shows up in screening |
| **Diagnostic manual criteria for specific learning disorder** | A skill clearly below what is expected for age | His skill is in the low-average range, above the threshold for clearly low performance for his age |
| **Gifted program criteria** | A high score on the Full Scale IQ or in achievement | The Full Scale IQ is pulled down by working memory and processing speed, and achievement is pulled down by the difficulty |

| Model | What it requires | Why it may miss him |
|---|---|---|
| **The "and" rule across criteria** | Exceeding the threshold on every criterion | He exceeds the threshold in ability but falls short of it in achievement, and so is excluded |
| **Teacher nomination** | Behavior that matches the image of a gifted student | The teacher sees frustration, avoidance, and slowness, and the giftedness is hidden from view |

## 13.4 The debate over identification criteria

The most important scholarly debate in this area turns on a single question: **what do we compare the student's performance against?**

### Normative comparison

**Is the student's performance below that of his same-age peers?**

This is what the diagnostic manual's criteria require. Proponents of this comparison argue that a student who reads in the average range does not suffer a functional impairment that would justify a diagnosis, and that diagnosing everyone with a gap between their abilities would expand the category without justification and turn "relative weakness" into a disability (Lovett & Lewandowski, 2006).

### Intra-individual comparison

**Is the student's performance clearly below what his abilities predict?**

Proponents of this comparison argue that a student whose ability lies in the top 1% of his peers, and who reads at an average level only with double the effort, does indeed have a real difficulty with real consequences, even if he is not below his peers. In their view, requiring normatively low performance denies these students understanding and support (Assouline et al., 2010).

**Figure 13.3:** The two lenses in the debate over identifying twice-exceptional learners.

One study compared different methods of identifying gifted students with learning disabilities within a single sample and found that the methods differed considerably in how many students they identified and in which

students, with limited agreement among them (Maddocks, 2018). In other words, who counts as twice-exceptional depends in part on the criterion we choose.

**A balanced practical position**

The gap between ability and achievement is not sufficient on its own, since everyone has uneven abilities. Nor is normatively low performance sufficient on its own, because it misses these students. The position toward which good practice leans is to base judgment on the gap together with converging evidence: a **history** of early difficulty in the same skill; **processing indicators** that explain it, such as weak phonological awareness in reading or weak orthographic knowledge in writing; a genuine **functional impact**, such as double effort, avoidance, or exhaustion; the **persistence** of the difficulty despite appropriate instruction targeting it; and a skill that actually lies in the low-average range or near the threshold, not merely below the student's other abilities.

## 13.5 Assessment: the protocol and the uneven profile

### Reading an uneven cognitive profile

Many twice-exceptional learners show a recurring pattern on intelligence scales such as the Wechsler Intelligence Scale for Children, Fifth Edition (**WISC-V**): very high reasoning and verbal comprehension, alongside clearly lower working memory and processing speed. These two indexes enter into the calculation of the Full Scale IQ and pull it down. For this reason, identifying giftedness in these students relies on the **General Ability Index** (**General**

**Ability Index, GAI**), which is calculated without these two indexes (NAGC, 2018). Chapter 12 (Section 12.4) explained this index, the ceiling effect, and above-level testing in detail.

What this chapter adds is that the index has a counterpart that completes it: the **Cognitive Proficiency Index** (**Cognitive Proficiency Index, CPI**), calculated from the working memory and processing speed subtests—that is, from the efficiency with which the student processes information, independent of his reasoning ability. Comparing the two indexes is the systematic way to describe an uneven profile, rather than looking at the graph and judging by eye (Kaufman et al., 2016). Two questions are asked, in order:

1. **Is the difference statistically significant?** That is, is it larger than measurement error alone would produce? This is settled by consulting the tables in the manual.
2. **Is it rare?** Many statistically significant differences are common in the norm sample, and their presence signals nothing unusual. We therefore look at the proportion of the norm sample who showed a difference of this size, that is, the base rate.

When a student reaches the ceiling on several subtests, ordinary scores do not reveal the full extent of his ability. Extended norms have been published for the WISC-V Full Scale IQ and General Ability Index that allow scores above the ceiling of the standard norms; the publisher should be consulted on the conditions for their use. But the uneven profile as a whole, however wide its gap, **is not diagnostic**: it appears in many gifted students with no disability whatsoever, and it is not sufficient on its own for any judgment.

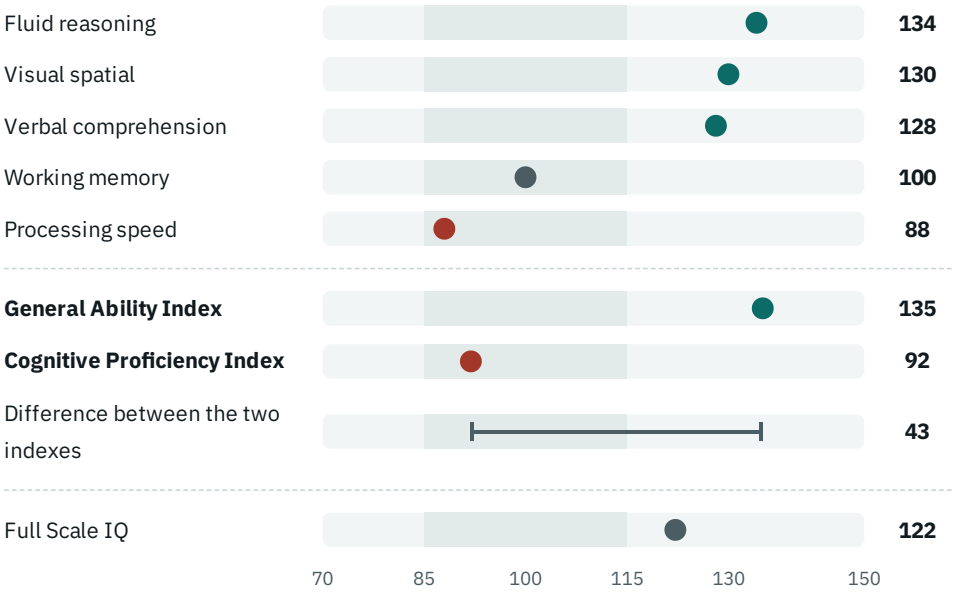



**Figure 13.4:** A hypothetical profile for Yasser on the WISC-V indexes. The gray band is the average range, 85 to 115, that is, one standard deviation on either side of the mean, as in the book's other graphs. (The test manual applies the label "average" only to the range 90 to 109 in its descriptive classification.) The difference between the General Ability Index and the Cognitive Proficiency Index is 43 points, a difference whose significance and rarity are checked in the manual's tables before it is interpreted. The Full Scale IQ of 122 falls short of the cut-off score in many gifted programs, whereas the General Ability Index of 135 exceeds it.

## Principles of assessment

1. **Two lenses in one assessment:** The assessment involves both someone who knows giftedness and someone who knows disability, or else an assessor who brings both kinds of expertise. Two questions are asked about the student together: What are his abilities? And what is holding him back?
2. **Start with strength:** What the student excels at is measured with high-ceiling tools, such as above-level testing, performance tasks, and authentic products, rather than grade-level tools on which he reaches the highest possible score and which therefore cannot reveal how far his ability extends (see Chapter 12, Section 12.4).

3. **Measure skill by time and effort:** Many twice-exceptional learners arrive at the correct answer eventually. Performance is therefore measured once under time limits and once untimed, and the assessor observes the effort the student expends and the compensatory strategies he relies on.
4. **Developmental and educational history:** When did the difficulty first appear? How did the student compensate for it? Did his performance drop when demands rose at a particular stage, such as the move to intermediate school, when writing and self-regulation demands multiply? What instruction has he received in the area of his difficulty, and how did he respond to it?
5. **Guard against overinterpretation:** No judgment rests on a gap between indexes alone, nor on analyzing each subtest in isolation. The Full Scale IQ and the General Ability Index are more reliable than any single index, and the difference between two indexes is less reliable than either of them.
6. **Emotional aspects:** Twice-exceptional learners are vulnerable to frustration and a poor self-concept, because they know they are capable yet cannot understand why they keep stumbling. This side is therefore assessed and taken into account in the plan.

## Assessment protocol profile

No single tool reveals twice-exceptionality. What reveals it is a **protocol**: a set of tools planned in advance and selected according to the referral question, with each component answering a specific question. The profile below presents this protocol in the same format used for instruments in the chapters on specific categories.

### 2e Two-lens assessment protocol for a student suspected of being twice-exceptional

Assessment protocol

### The idea

Both tracks are opened together from the outset: one track searches for strength with tools unconstrained by a ceiling, and the other searches for difficulty with tools that expose what compensation hides. The team does not wait for one to be settled before opening the other, because each side changes how the other should be read.

### Its components

**Table 13.3: Components of the two-lens assessment protocol, and the question each component answers.**

| Component | What is administered | Question it answers |
|---|---|---|
| **Senses and learning opportunities** | A recent hearing and vision screening, and a review of attendance and of the instruction the student has received | Is the difficulty explained by a sensory impairment or by a lack of learning opportunities? |
| **General ability** | An individual intelligence scale with all its indexes, calculating the General Ability Index, the Cognitive Proficiency Index, and the difference between them | What is the student's ability on an index that the difficulty does not depress? Is the discrepancy significant and rare? |
| **High-ceiling strength** | An above-level test in the area of strength, and a performance task or authentic product rated against clear criteria | Where does the student truly stand in his area of excellence? |
| **Basic skills, timed and untimed** | A standardized achievement test in reading, writing, and mathematics, part of which is readministered without time limits, along with copying, dictation, and free-writing samples | What does open time conceal? What is the difference between accuracy and fluency? |

| Component | What is administered | Question it answers |
|---|---|---|
| **Processing indicators** | A phonological processing tool, such as the Comprehensive Test of Phonological Processing, Second Edition (**CTOPP-2**; Wagner et al., 2013), or Arabic tasks measuring phonological awareness and rapid naming, along with tasks of orthographic and morphological knowledge when the concern lies in writing | Is there a deficit in a process that explains the weak skill? |
| **Attention and behavior** | Standardized rating scales from the teacher and the family, and from the student himself in adolescence, interpreted using T-scores, as well as observation checklists | Do the symptoms appear in more than one setting, and how severe are they? |
| **Structured observation** | Interval-recording observation in an ordinary lesson and in a lesson that offers real challenge suited to his level | What remains of the behavior when challenge is present? |
| **History** | A family interview and a review of school records going back to kindergarten | When did the difficulty begin? Did symptoms appear before school and before age twelve? |
| **Emotional domain** | A student interview, and a self-concept or anxiety measure when needed | How does the student explain his struggles to himself? And how have they affected him? |

**Table 13.4: Protocol information profile.**

| | |
|---|---|
| **Purpose** | To verify the coexistence of giftedness and disability, or to determine which of the three possibilities in Section 13.7 is most likely |

| | |
|---|---|
| **Ages** | School age; instruments are selected according to the student's age |
| **Who administers it** | A team that includes expertise in both giftedness and disability; individual measures are administered by a qualified psychologist in accordance with each publisher's requirements |
| **Time** | Typically three to five sessions, with at least two classroom observations |
| **Score types** | Standard scores, T-scores, and percentile ranks from each tool, along with observation and interview data. A percentile rank is the percentage of the student's peers in the norm sample who scored lower than he did; a percentile rank of 84 means that his performance exceeds that of 84% of them |
| **Cutoff scores** | There is no cut-off score for twice-exceptionality. The criteria for giftedness and the criteria for the disorder are applied separately, and judgment rests on converging evidence |
| **Reliability** | The reliability of each tool is as reported in its manual. Differences between indexes are less reliable than the indexes themselves |
| **Arabic version** | We know of no standardized Arabic protocol for twice-exceptional learners. The Wechsler scale has a Saudi standardization of the revised edition (**WISC-R**) for ages 6 to 16 (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024), but we know of no published Saudi standardization of the fifth edition. Nor did we know of any published standardized Arabic version of the **CTOPP-2** at the time this book was prepared |

### How are the results interpreted?

The results are read in three layers. First, are the criteria for giftedness met on an index that the difficulty does not depress? Second, are the criteria for the disorder met, with evidence from history, processing, functional impact, and

persistence? Third, does the evidence from different sources converge, or does some of it contradict the rest, so that the team needs additional evidence before reaching a judgment?

### Its limits

The protocol lengthens the assessment and requires a team that is not available in every school. The large number of tools raises the likelihood that a low score will appear by chance, so no judgment should rest on a single score. Most of its tools are standardized on foreign samples, so their scores are read with caution where Arabic norms are unavailable.

## Adapting gifted identification tools for learners with sensory and physical disabilities

A deaf or blind student, or a student with a physical disability, may be gifted, yet the usual identification tools measure the disability before they measure the ability. A written verbal reasoning test measures a deaf student's command of written Arabic; a test of printed matrices is inaccessible to a blind student; and a timed test measures the hand speed of a student with a physical disability before it measures his thinking. The following should therefore be observed:

- **For deaf and hard-of-hearing students:** Instructions are given in sign language or in whatever mode the student uses to communicate, and nonverbal ability measures are used, such as the Universal Nonverbal Intelligence Test, Second Edition (**UNIT2**; Bracken & McCallum, 2016). Nonverbal tests, however, are weaker predictors of academic achievement and may miss students who excel in verbal and quantitative reasoning (Lohman, 2005), so they are supplemented with the student's products and with ratings from teachers who know him.
- **For blind and low-vision students:** Oral verbal measures are relied on, and materials are presented in braille or large print. Any adaptation that changes how an item is presented may change what it measures, so the

score is not compared with the original norms unless the adaptation is approved in the tool's manual.

- **For students with physical disabilities:** Alternative response modes are permitted, such as pointing, dictating to a scribe, or using a computer; tests that depend on motor speed are excluded from the judgment, and this is stated explicitly in the report.

The general rule is this: an adaptation that removes a barrier unrelated to the ability being measured improves the validity of the result, whereas an adaptation that changes the measured ability itself invalidates comparison with the norms. In either case, the adaptation is documented in the report. As for the Mawhiba Multiple Cognitive Abilities Test, administered within the National Program for Gifted Identification (Ministry of Education, 1437 AH [Arabic]), we found no description, in the published sources we consulted, of the accommodations available in it for students with disabilities. It is therefore advisable for the family or the school to ask the administering body about them before registration, and for the performance of a student with a disability on the test to be read in light of what he was given or denied.

**For the teacher: signs worth noticing**

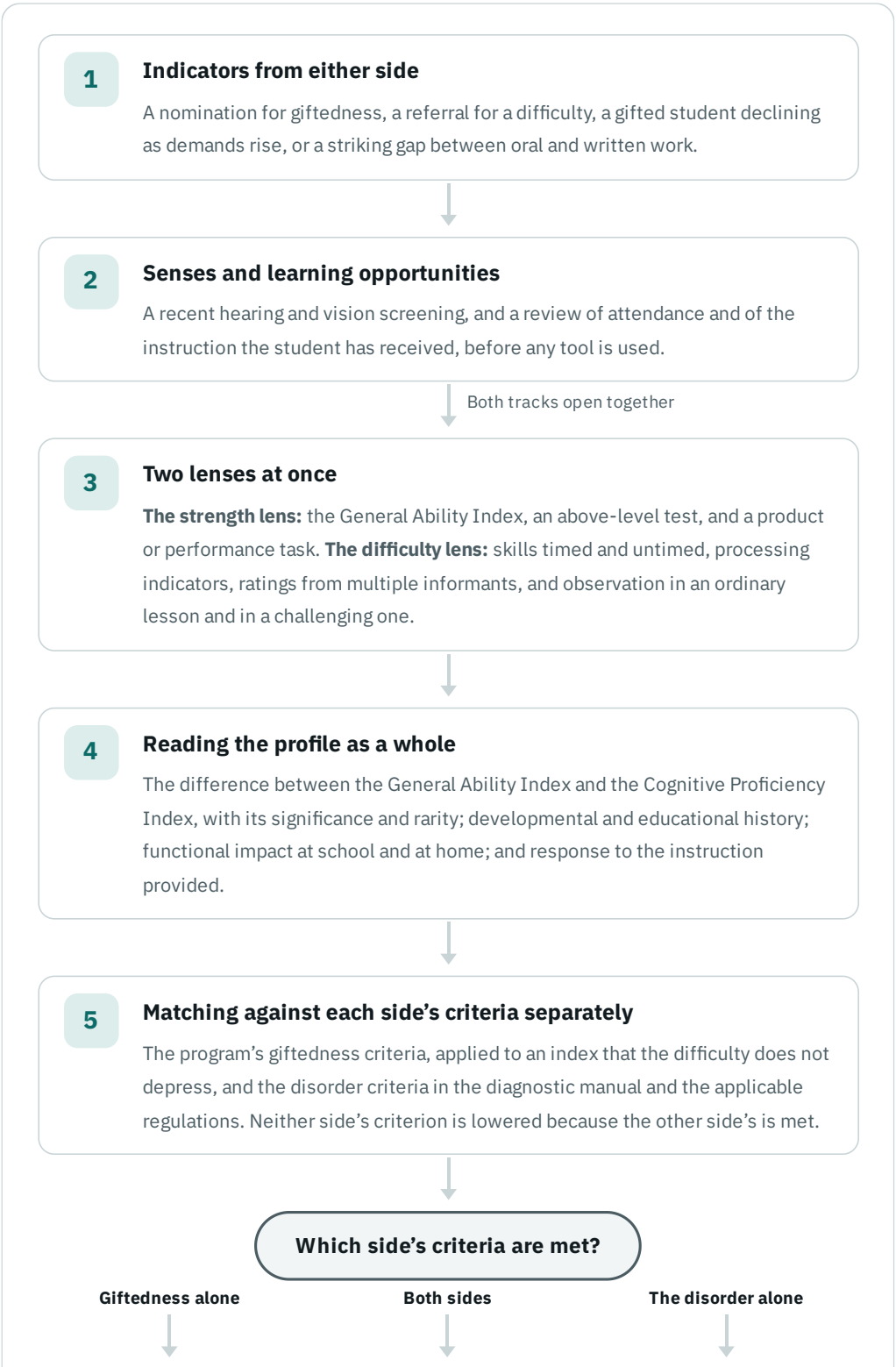
- A wide gap between what the student says orally and what he writes.
- A gifted student who declines on moving to a stage with heavier writing and self-regulation demands, such as intermediate school.
- A deep grasp of complex ideas alongside stumbling over simple skills such as spelling or math facts.
- Erratic performance: excellent on one day or task and weak on others, for no apparent reason.
- Severe fatigue or strong avoidance of particular tasks, alongside enthusiasm for others.
- Frustration and remarks like "I'm stupid" from a student who strikes you as very bright.

Your role is to document what you see with examples, situations, and dates; to refer the student to both teams together; and to leave the judgment of "gifted only" or "has a difficulty only" to the integrated assessment.

## 13.6 The decision pathway: opening both tracks together

In the chapters on specific categories, the decision pathway runs along a single line, from initial indicators to matching against the criteria of one category. With a student suspected of being twice-exceptional, however, the most common error is to open one track and close the other prematurely: the student is admitted to the gifted program and no one asks about his difficulty, or he is diagnosed with a disability and no one asks about his giftedness. For

this reason, after the second step the pathway branches into two lenses that work together and then converge in a single reading of the profile before the decision.


1
Indicators from either side
A nomination for giftedness, a referral for a difficulty, a gifted student declining as demands rise, or a striking gap between oral and written work.
2
Senses and learning opportunities
A recent hearing and vision screening, and a review of attendance and of the instruction the student has received, before any tool is used.
Both tracks open together
3
Two lenses at once
The strength lens: the General Ability Index, an above-level test, and a product or performance task. The difficulty lens: skills timed and untimed, processing indicators, ratings from multiple informants, and observation in an ordinary lesson and in a challenging one.
4
Reading the profile as a whole
The difference between the General Ability Index and the Cognitive Proficiency Index, with its significance and rarity; developmental and educational history; functional impact at school and at home; and response to the instruction provided.
5
Matching against each side's criteria separately
The program's giftedness criteria, applied to an index that the difficulty does not depress, and the disorder criteria in the diagnostic manual and the applicable regulations. Neither side's criterion is lowered because the other side's is met.
Which side's criteria are met?
Giftedness alone
Both sides
The disorder alone

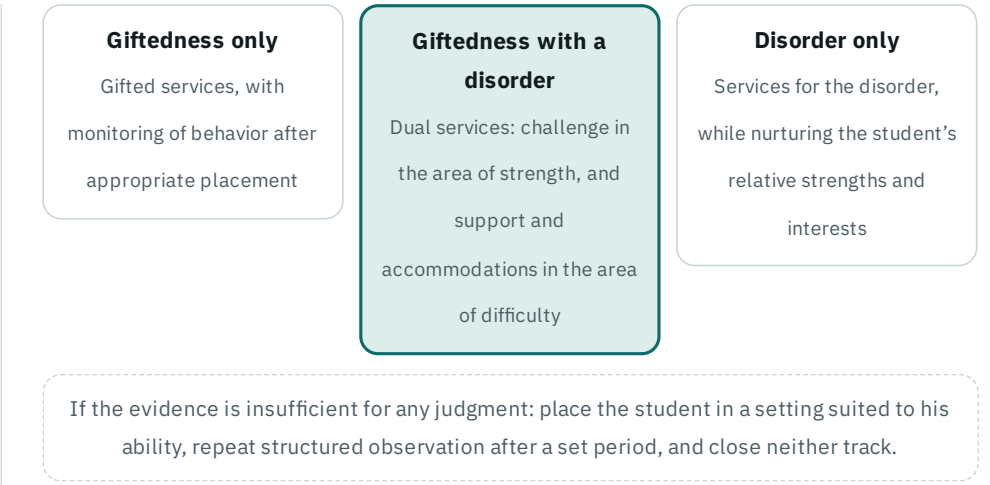



**Figure 13.5:** The decision pathway for a student suspected of being twice-exceptional. After sensory screening, the assessment branches into two lenses that work together; the criteria for each side are applied separately; and the pathway ends in one of three decisions, or in appropriate placement followed by renewed evidence gathering.

The crucial point in the fifth step is that the team must not "go easy" on one side's criteria because the other side's are met. Giftedness does not prove a disorder, and a disorder does not rule out giftedness. What changes for the twice-exceptional student is the **evidence** accepted for each criterion, while the criterion itself stays fixed: the General Ability Index is used in place of the Full Scale IQ for the giftedness criterion, and for the difficulty criterion the skill is read under time limits and in terms of the effort expended.

## 13.7 Differential diagnosis: three possibilities

When a student shows signs suggestive of both giftedness and a disorder, the team faces three possibilities, all of them plausible. It may be giftedness only, some of whose characteristics resemble a disorder when the student finds nothing suited to him (Webb et al., 2005). It may be giftedness with a disorder, which is twice-exceptionality. Or it may be a disorder only, with a relative strength in one part of the profile that does not rise to the level of giftedness. Error is possible in all three directions, and every error has its cost.

**Table 13.5: The three possibilities in differential diagnosis, what favors each, and what is not sufficient on its own.**

| Possibility | What favors it | Not sufficient on its own | The cost of error |
|---|---|---|---|
| **Giftedness only, resembling a disorder** | The concern centers on low-challenge situations and diminishes clearly after weeks of appropriate placement. There is no functional impact outside these situations, no developmental history of difficulty, and basic skills are intact on standardized, timed measurement | The behavior disappearing during a favored or novel activity; this also happens with the disorder | If the disorder is in fact present, the student is denied understanding and support and is blamed for what he cannot control |
| **Giftedness with a disorder (twice-exceptionality)** | Evidence of high ability from more than one source, such as the General Ability Index, above-level testing, and products, together with the disorder criteria being met: early history, a processing indicator or a behavioral pattern in two settings, functional impact, and persistence despite appropriate instruction or appropriate placement | The gap between indexes; it is common among gifted students with no disability at all | If the disorder is not present, the student is labeled with a diagnosis he does not need and may be given treatment he has no need for |

| Possibility | What favors it | Not sufficient on its own | The cost of error |
|---|---|---|---|
| **Disorder only, with a relative strength** | The disorder criteria are met, and the strength is a relative strength within the student's own profile that does not clearly exceed his peers, or broad knowledge confined to a narrow interest without high-level reasoning in that domain | An extensive memory for facts on a favorite topic, early mechanical reading without matching comprehension, or a single high index amid average ones | If giftedness is present, it goes to waste, and the student receives services that focus on his weaknesses alone |

⚠ **Caution: "What happens when he finds challenge?" is a helpful question, not a decisive one**

Comparing a student's behavior in an ordinary lesson with his behavior in a lesson that offers genuine challenge is a useful comparison, but it does not settle the matter by itself. The diagnostic manual states that signs of attention-deficit/hyperactivity disorder may be minimal or absent when the child is receiving frequent rewards for appropriate behavior, is under close supervision, is in a novel setting, or is engaged in especially interesting activities (APA, 2022). A student's ability to concentrate for hours on an activity he loves therefore **does not rule out** the disorder. What tips the balance toward one possibility or the other is the convergence of evidence: ratings from two settings, developmental history, functional impact at school and at home, and what remains of the behavior after weeks of appropriate placement. These carry more weight than an hour in a club he loves.

## 13.8 When is a judgment warranted? Four cases

The four cases that follow cover the three possibilities set out in the previous section. In three of them, the criteria for both sides are met through different kinds of evidence: a writing difficulty, autism, and attention-deficit/hyperactivity disorder. In one, Muhannad's case (Case 2), the evidence is not yet sufficient for any judgment about a disorder.

### Case 1: Yasser, age 14, ninth grade (third year of intermediate school)

**Giftedness with a writing difficulty**

**The story:** Yasser, whose scene opened this chapter, has been in the gifted program since fourth grade and trains with his school's team for mathematics competitions. At the start of ninth grade, his program teacher referred him after he failed to submit the report for his research project, even though his oral presentation of the idea had astonished the judging panel. His family reported that he had been slow to learn letter formation and spelling in first and second grade, and that he "used to write words the way he said them"; they were told he would improve with time. The gap then widened in intermediate school, as reports and essay answers multiplied, and his older brother began writing down what Yasser dictated to him.

**What the two-lens assessment added:**

**Table 13.6: Evidence in Yasser's case through the two lenses.**

| Lens | Evidence | What it means |
|---|---|---|
| **Senses** | Hearing and vision screening this year: normal | No sensory explanation for the difficulty |

| Lens | Evidence | What it means |
|---|---|---|
| **Giftedness** | General Ability Index 135, Cognitive Proficiency Index 92 and Full Scale IQ 122, as shown in Figure 13.4. The difference between the two indexes is statistically significant and rare according to the manual's tables | Very high ability, which the full-scale score underestimates |
| **Giftedness** | A mathematics test designed for twelfth-grade students: percentile rank 70 among students in that grade | Realized mathematical talent three years ahead of his grade |
| **Reading** | Reading a fully vowelized text and an unvowelized text at grade level: accuracy 98% and 96%, fluency in the average range, and reading comprehension 118 | Reading is intact in both conditions, so the difficulty must be sought beyond decoding |
| **Difficulty** | Standardized achievement test in writing: spelling 77, writing fluency 80, and written expression 84 | At or below one standard deviation under the mean, and far below his ability |
| **Difficulty** | Analysis of spelling errors across three samples: most involve the hamza; tāʾ marbūṭa, open tāʾ and hāʾ; and letters of similar shape distinguished only by their dots, such as writing "ج" for "ح" or dropping dots; as well as spelling words as they are pronounced in his dialect | A consistent weakness in orthographic and morphological knowledge that recurs in a clear pattern |

| Lens | Evidence | What it means |
|---|---|---|
| **Difficulty** | Copying from a ready, legible text at a speed close to his peers'. Free handwritten composition, however, yielded forty-five words in ten minutes, in short, loosely connected sentences. When he wrote on the same topic using a keyboard, his text grew slightly longer but the spelling errors remained; when he spoke about it orally, his presentation was organized and rich | The difficulty is linguistic, orthographic and organizational, not motor |
| **Fine motor** | Screening by an occupational therapist: average | Developmental coordination disorder unlikely |
| **Intervention** | Twelve weeks in eighth grade of explicit instruction in spelling rules and the hamza in a small group: errors decreased on the trained words, but the gains did not transfer to his free writing | Persistence of the difficulty despite targeted instruction |
| **Emotion** | He says, "My thoughts are faster than my hand," and "I'm stupid in Arabic," and he grows tense before essay tests | An effect on self-concept and anxiety tied to writing |

| A ✓ | B ✓ | C ✓ | D ✓ |
|---|---|---|---|
| Persistent despite 12 weeks of intervention | Spelling 77 and written expression 84, with supporting evidence | Began in the early grades and intensified as demands increased | Intact senses, high ability, regular instruction, and intact motor skills |

**How did the team read the evidence?** Yasser reads both vowelized and unvowelized text accurately, and this was the first thing that set his case apart from dyslexia. His writing, however, breaks down at two points where difficulty is common in Arabic: orthographic knowledge, such as

the rules of the hamza and tāʾ marbūṭa and distinguishing letters that differ only in their dots; and the distance between his spoken dialect and written Standard Arabic. Comparing copying with dictation, free composition and oral presentation further showed that his hand is not where the problem lies (Chapter 8, in Sections 8.3 and 8.4, set out the characteristics of Arabic reading and writing and the logic of this comparison).

**Decision:** The evidence that the balanced position calls for had come together: a large discrepancy, an early history, an orthographic processing indicator, functional impact, persistence despite intervention, and a skill below the threshold. The team therefore concluded that Yasser is twice-exceptional: gifted in mathematics, with a specific learning disorder with impairment in written expression, affecting spelling accuracy and the clarity and organization of written expression, at a mild level. He remains in the gifted program and receives explicit instruction in spelling grounded in root and pattern, and in planning a text before writing it, with accommodations: word processing for long assignments, speech-to-text for project drafts, extra time, and assessment of content in subjects other than Arabic independently of spelling. His profile is also discussed with him openly, so that he understands that his struggles with writing do not cancel out his ability.

**On the term "dysgraphia"**

The term dysgraphia is used in the literature in two senses: a difficulty with handwriting and the physical act of writing, and a difficulty with spelling and written expression. The diagnostic manual speaks instead of a specific learning disorder "with impairment in written expression," and does not count poor handwriting alone as part of it (APA, 2022). Yasser's difficulty is of the second kind, which is why the solution was not to improve his handwriting.

**What evidence could change the decision?** If, after a semester of explicit instruction, his free writing samples showed his spelling errors falling to the level of his peers without accommodations, the description of the disorder and its severity would be revisited. And if a full evaluation by an occupational therapist revealed a motor impairment that the screening had missed, the possibility of developmental coordination disorder would come back under consideration.

## Case 2: Muhannad, age 9, fourth grade

**Giftedness that resembles an attention disorder: the evidence is not yet sufficient**

**The story:** Muhannad was referred for evaluation on suspicion of attention-deficit/hyperactivity disorder. In class he does not finish his exercises, interrupts the teacher with off-topic questions and doodles in the margins of his notebook. His teacher's ratings on an attention rating scale are elevated. Before reaching any judgment, and after confirming that his hearing and vision were normal, the team gathered the following evidence over six weeks.

**Table 13.7: Evidence in Muhannad's case.**

| Source | Result | What it means |
|---|---|---|
| **Standardized rating scale: teacher** | Inattention: T-score 68 (six symptoms); hyperactivity-impulsivity: 57 | Elevated in the regular classroom |
| **Standardized rating scale: family** | Inattention: T-score 58 (three symptoms: his mind wanders when given tasks that do not interest him, he avoids repetitive homework, and he sometimes forgets what he is asked to do); hyperactivity: 52 | Few symptoms at home, with a score below the elevated range |
| **Chess club supervisor** | One symptom; describes him as "the most focused student in the club" | The behavior decreases in an activity he enjoys, which by itself does not rule out the disorder |
| **Systematic observation with interval recording** | Off task in 38% of intervals in an ordinary mathematics lesson, and in 9% in a lesson in which he was given sixth-grade-level problems. The rates for two classmates in the ordinary lesson were 12% and 15% | The behavior rises markedly when challenge is absent in class |
| **Developmental history** | No concerns about attention or activity level in kindergarten or in first and second grade. The concerns began in third grade, when he was getting far ahead of the mathematics curriculum | No symptoms before school entry, and the onset of concerns is tied to placement |

| Source | Result | What it means |
|---|---|---|
| **Functional impact at home** | He completes his homework on his own in a reasonable time, keeps his materials organized, keeps to his schedule, and spends hours building complex models | No clear functional impact outside the classroom |
| **Ability test** | General Ability Index 136, working memory 124 | Very high ability, with no weakness in working memory |
| **Above-level mathematics test** | At the level of the average sixth-grade student | About two years ahead of his grade's curriculum |

**How did the team read the evidence?** It would be inaccurate to say that the symptoms "appear in only one setting." Criterion C in the diagnostic manual requires only that **several symptoms** be present in two or more settings, without requiring the full threshold in each, and three symptoms at home may satisfy it (APA, 2022). What prevents a judgment at this point is two other things. The first is the absence of clear functional impact outside the classroom, whereas Criterion D requires that the symptoms interfere with, or reduce the quality of, functioning. The second is a strong alternative explanation that the manual itself mentions under differential diagnosis: individuals of high ability may appear inattentive in an environment that does not challenge them. This explanation is strengthened by the onset's link to placement and by the absence of any symptoms before school. Yet the decline in the behavior in the challenging lesson and at the chess club is **not enough** to rule out the disorder, since this also happens when the disorder is present.

**Decision:** The evidence is not yet sufficient either to diagnose attention-deficit/hyperactivity disorder or to rule it out definitively. Muhannad is therefore identified as gifted and accelerated in mathematics

(**Acceleration**), that is, moved through its curriculum at a pace that suits his level. The systematic observation and the teacher and family rating scales are repeated after 8 to 12 weeks of appropriate placement. If the symptoms decline in class and there is still no functional impact at home, the "giftedness only" possibility gains weight. If the symptoms persist in two settings with functional impact, the evaluation of the disorder is completed as a possible case of twice-exceptionality.

**The lesson**

Not every gifted student who struggles is twice-exceptional. Some characteristics of giftedness, such as boredom with repetition, constant questioning and absorption in ideas, may resemble a disorder when the student finds nothing suited to him (Webb et al., 2005). The question "What happens when he finds the right level of challenge?" is a **supporting question**, answered through actual placement and follow-up over several weeks, and it remains one piece of corroborating evidence to be weighed alongside others.

**What evidence could change the decision?** If his off-task behavior in class remained high after 8 to 12 weeks of acceleration, or if functional impact emerged at home where none had existed before, such as late homework or lost belongings, the balance would tip toward completing the evaluation of the disorder. And if kindergarten records turned up notes on attention or activity level that had not reached the team, the argument for an onset tied to placement would be weakened.

## Case 3: Joud, age 15, tenth grade (first year of secondary school)

**Verbal giftedness with autism, hidden for years**

**The story:** Joud was identified as gifted in elementary school for her verbal talent; today she writes metrical poetry and has won a literary competition at the regional level. Her teachers describe her as "polite, quiet, outstanding." Her name did not reach the assessment team through the gifted program but through the school counselor: she had been repeatedly absent on Sunday mornings with headaches, had begun spending recess alone in the learning resource center, and then broke down in tears after her only friend moved to another group. The counselor assumed she was dealing with social anxiety or depression. Her mother said that Joud "shuts down" after school, closing her bedroom door for hours without speaking to anyone. Joud herself told the counselor: "I study my classmates the way I study texts: I memorize what they say, pick out what fits the situation, and practice my facial expressions in the mirror."

**Table 13.8: Evidence in Joud's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing and vision** | Both normal | No sensory explanation |
| **Social Responsiveness Scale, Second Edition (SRS-2)** | Mother's rating: T-score 72 (moderate range). Teacher's rating: 57 (within normal limits) | A wide discrepancy between home and school, consistent with camouflaging of traits at school |

| Source | Result | What it means |
|---|---|---|
| **Autism Diagnostic Interview with the mother (ADI-R)** | Scores reached the algorithm cut-offs in all three domains: reciprocal social interaction, communication, and restricted and repetitive patterns. From her history before age three: she used words to label things and repeated lines from television in an adult tone, did not point to objects to show them to others, arranged books by color, and became upset if the route to her grandmother's house changed | Clear early indicators before 36 months; the difficulty long predates adolescence |
| **Autism Diagnostic Observation Schedule (ADOS-2), Module 4** | Classification: "autism spectrum." Fluent speech with a rich vocabulary that turns into a lecture when the topic is poetry and prosody; difficulty explaining why her friend was upset or describing the feelings of characters in a story; and eye contact that the specialist describes as "studied" | The observation is consistent with the interview, although camouflaging may lower the score in a structured session |
| **Self-report anxiety scale** | Social anxiety in the elevated range; depression within normal limits | Co-occurring anxiety that needs attention but falls short of explaining her entire history |
| **—** | Verbal Comprehension 141, General Ability Index 131, Processing Speed 94 | Very high verbal ability |

| Source | Result | What it means |
|---|---|---|
| **Her products** | A small collection of her poems, the literary competition award, and her Arabic teacher's evaluation of her writing | Realized talent in a specific domain |
| **Observation at recess and in a group activity** | She withdraws from noise, avoids the cafeteria, sticks in the activity to a role she has set for herself, and becomes upset if it is changed | Social difficulty and sensory sensitivity in unstructured situations |

| A ✓ | B ✓ | C ✓ | D ✓ | E ✓ |
|---|---|---|---|---|
| In the interview and the observation | Intense interest, routines and sensory sensitivity | Indicators before age three | Absences, exhaustion and breakdown | High ability and no better explanation |

**How did the team read the evidence?** The differential question was whether this was social anxiety alone. A person with social anxiety has social skills that fear holds back, and interacts naturally with familiar people. Joud's difficulty with social reciprocity, by contrast, has been evident since before age three, including with her family, and it comes with an intense interest, sensory sensitivity and an aversion to change. The anxiety here co-occurs with autism; it is not a better explanation for it. Her identification was delayed until adolescence because her verbal talent and her powers of observation allowed her to camouflage her traits, and because her intense interest, in poetry and prosody, is one that school praises and that does not seem unusual.

**Decision:** Joud was diagnosed with autism spectrum disorder, without intellectual impairment and without language impairment, at support level 1, the lightest of the three support levels in the diagnostic manual, with co-occurring anxiety for which she was referred for evaluation and

treatment. She is twice-exceptional. The team designed services that bring both sides together, an approach known as dual differentiation: advanced enrichment in literary writing with a specialist mentor who follows her work; a quiet space available at recess; explicit teaching of what is expected in social situations, rather than leaving her to work it out alone through exhausting effort; and a conversation with the family about the after-school "shutting down" as the price she pays for camouflaging throughout her school day.

**Camouflaging among gifted autistic girls**

Some autistic individuals, and girls in particular, learn to camouflage their traits (**Camouflaging**): they watch others, imitate how they speak and their facial expressions, and memorize ready-made phrases for different situations (Hull et al., 2017; see also Chapter 9, Section 9.4). Verbal giftedness makes this camouflaging more polished and longer-lasting, so that a student can reach secondary school without anyone asking about autism, while her anxiety and exhaustion are seen as the whole problem. For this reason, in assessing this group the team gives particular weight to early developmental history, to discrepancies between home and school ratings, and to what the student herself says about the effort she spends to appear "normal." A low score in a structured observation session is not reassuring: the session is short and structured, precisely the kind of situation in which the student camouflages most skillfully.

**What evidence could change the decision?** If early records or an interview with another relative showed that the indicators before age three were not as her mother described them, or if her social

> reciprocity improved markedly after treatment for anxiety, social anxiety would return as the leading explanation, and the autism diagnosis would need to be revisited.

## Case 4: Faris, age 13, seventh grade (first year of intermediate school)

**Giftedness with attention-deficit/hyperactivity disorder**

**The story:** Faris is gifted in mathematics and programming and was admitted to the gifted program in fifth grade. His program teacher referred him because a pattern kept recurring in every project: he starts with striking enthusiasm, then leaves the work unfinished, loses parts of the team's robot and misses the deadline. In an advanced mathematics competition, he got problems wrong that he knew how to solve because he skipped steps and read the given information in a hurry. His mother said this was "nothing new" and that she sits beside him every evening so that he finishes his homework.

**Table 13.9: Evidence in Faris's case.**

| Source | Result | What it means |
|---|---|---|
| **Hearing, vision and sleep** | Hearing and vision normal; sleeps about nine hours, according to the family | No sensory explanation and no sleep deprivation |
| **Standardized rating scale: inattention** | T-scores: mathematics teacher 72, gifted program teacher 69, mother 71, and Faris himself 65. Seven symptoms in both the teacher's and the mother's ratings | Elevated symptoms in three settings, including a challenging one |

| Source | Result | What it means |
|---|---|---|
| **Standardized rating scale: hyperactivity-impulsivity** | T-scores: teacher 62, mother 60 | Mildly elevated; the predominant picture is inattention |
| **Systematic observation with interval recording** | Off task in 34% of intervals in an ordinary lesson, and 27% in a robotics session he enjoys and that challenges him. Two classmates in the same session: 10% and 12% | The behavior decreases slightly with challenge but remains far higher than his peers' |
| **History** | Kindergarten report: "Doesn't settle in circle time and leaves activities before finishing them." A note in second grade: "Very bright but forgets his books and needs constant reminders" | Symptoms before age twelve and before school entry |
| **Functional impact at home** | Homework takes three hours under his mother's supervision; he forgets test dates, loses his belongings every week, and this leads to frequent conflict at home | Clear functional impact outside school |
| — | General Ability Index 131, Cognitive Proficiency Index 91 (working memory 97, processing speed 86) | Very high ability. The difference between the two indexes is consistent with the picture but does not diagnose it on its own |

| Source | Result | What it means |
|---|---|---|
| **A mathematics test designed for tenth-grade students** | Percentile rank 80 among students in that grade | Realized mathematical talent |
| **Self-report measure of anxiety and depression, and achievement in reading and writing** | Anxiety and depression within normal limits; achievement in the high average range | No better explanation from another disorder, and no co-occurring learning disability |

| A ✓ | B ✓ | C ✓ | D ✓ | E ✓ |
|---|---|---|---|---|
| Seven symptoms of inattention | Symptoms before age 12 | Home, classroom and program | Impact on achievement and family | No anxiety or depression accounts for them |

**Decision:** The team concluded that the evidence favors attention-deficit/hyperactivity disorder, predominantly inattentive presentation, and that Faris is twice-exceptional. He was referred to a physician to complete the medical evaluation and to discuss treatment options with the family, since the Special Education Organizational Guide requires, for eligibility in this category, that the student's need for intervention be established "according to an approved report" (Ministry of Education, 1436 AH [Arabic]). Faris remains in the gifted program, with support for executive functions: breaking each project into short stages, each with its own deadline and review; a checklist against which he verifies the steps of his solution before submitting; external organizational tools such as a calendar and reminders; and teaching him to monitor his own attention.

**The lesson: What distinguished Faris from Muhannad?**

Both are gifted in mathematics, and both had teachers who rated their inattention symptoms as elevated. But Faris's symptoms remained far above his peers' even in a session he enjoys and that challenges him; they appeared in three settings, predated school, and affected his life at home. Muhannad's behavior, by contrast, dropped markedly with challenge, with no functional impact at home and no symptoms before school. What made the difference was the convergence of evidence from different sources.

**What evidence could change the decision?** If the medical evaluation revealed another cause for the symptoms, such as a sleep disorder the family had not noticed or a health condition affecting attention, the conclusion would be revisited. And if ratings from other teachers who know him well in different subjects came back clearly low, the requirement that symptoms appear in more than one setting would be called into question and would need to be examined anew.

## 13.9 In the Saudi and Arab context

**In the Saudi context: Can a student receive both gifted program services and learning disabilities services?**

**The short answer:** We found no published regulatory text that explicitly permits a student to combine a gifted program with a special education service, nor any text that prohibits it. The Special Education Organizational Guide lists eleven disability categories and does not address the gifted (Ministry of Education, 1436 AH [Arabic]). The gifted program guides we reviewed regulate nomination and admission and make no mention of students with disabilities (Ministry of Education, n.d., 1437 AH [Arabic]). In practice, then, the matter is left to the judgment of teams in schools and education directorates.

Even so, the texts contain provisions that open the door rather than close it. The Regulations for Special Education Institutes and Programs defined both "superior ability and giftedness" and "learning disabilities" in a single document, set separate admission requirements for each, and did not make giftedness a reason to exclude a learning disability. Moreover, the admission requirement for learning disabilities in the Regulations is "a clear discrepancy between the student's level of ability and level of academic achievement" in areas that include writing. This is the logic of intra-individual comparison, which suits twice-exceptional learners better than comparison with peers, although the Regulations do not specify a figure for this discrepancy (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). Gifted classes, for their part, require meeting two or more criteria, one of which is a cognitive abilities test with a score of no less than 120 (Ministry of Education, n.d. [Arabic]), and the guide does not specify whether this refers to the full-scale score or to an index that the difficulty does not depress.

Nor may the group be rare, as the preliminary estimate in Section 13.1 suggests (Almulhim et al., 2026). Saudi researchers have studied the experiences of teachers in the Hail region with these students (Alsamiri & Alsamani, 2023), teachers' professional development needs for identifying them in inclusive schools (Alsamiri & Alsamani, 2026), and the barriers that elementary school teachers see to identifying and supporting them (Alsamani et al., 2026).

**Table 13.10: Twice-exceptionality in the diagnostic manuals, US law and Saudi regulations.**

| Element | DSM-5-TR | US Individuals with Disabilities Education Act (IDEA) | Saudi regulations |
|---|---|---|---|
| **Is giftedness a category?** | No; the manual is designed for classifying disorders | No; giftedness falls outside the law's categories and is regulated by the states | Yes; it was defined in the 1422 AH Regulations and is governed by the gifted program guides |
| **Does giftedness preclude a diagnosis of a disorder?** | No. The learning disorder criterion compares the student with age expectations, with leeway up to one standard deviation if other evidence supports it | No. Eligibility requires both a disability and a need for special education | No text prohibits it. The learning disabilities requirement is a discrepancy between ability and achievement, with no specified figure |
| **Is combining the two services explicitly provided for?** | Not applicable | No explicit federal provision, and the law does not prohibit it | We found no text that permits or prohibits it |

**A question for reflection:** If Yasser's case came before your school tomorrow, who would request his evaluation: the gifted team or the learning disabilities team? And who has the authority to decide that he receives both services at once?

## Conclusion

The central argument of this chapter is that twice-exceptional learners are identified through a multi-instrument protocol that opens the giftedness and disability tracks together and applies the criteria for each separately. It follows that response to intervention, normative criteria or the "and" rule cannot be expected to identify these students; that the difference between the General Ability Index and the Cognitive Proficiency Index should be read as a systematic description of an uneven profile, one that precedes any diagnosis and requires further evidence; and that early history and home–school discrepancies deserve particular weight, especially for gifted autistic girls, who may camouflage their traits for years. In every case, the team faces three possibilities: giftedness only, giftedness with a disorder, or a disorder only.

Debate continues over the basis for judgment: normative comparison or intra-individual comparison. The balanced position presented in this chapter combines the discrepancy with history, processing indicators, functional impact and persistence despite appropriate instruction, but it remains a matter of professional judgment that has not yet settled into a rule. Two further questions also remain open: the claim that students in whom neither side has been detected form the largest group is a common hypothesis that awaits proof, and no published Saudi regulatory text has been found that either permits or prohibits combining a gifted program with learning disabilities services.

# 13.10 Summary, key terms and activities

## Chapter summary

- Twice-exceptional learners are gifted and also have a disability. The operational definition specifies who they are, while the consensus definition describes the interaction between the two sides and what these students need.
- Mutual masking makes their performance look average, so they appear in three groups. The claim that students in whom neither side has been detected form the largest group is a common hypothesis that has yet to be proven.
- The usual models, such as response to intervention, normative criteria and the "and" rule, may miss them.
- The debate centers on normative versus intra-individual comparison; the balanced position combines the discrepancy with history, processing indicators, functional impact and persistence despite appropriate instruction.
- Twice-exceptional learners are identified through a multi-instrument protocol that opens both tracks together and applies the criteria for each side separately.
- The difference between the General Ability Index and the Cognitive Proficiency Index describes an uneven profile systematically, but it does not diagnose on its own.
- The team faces three possibilities: giftedness only, giftedness with a disorder, or a disorder only. The question "What happens when the student finds the right challenge?" is a supporting one, to be weighed with other evidence, because symptoms of a disorder may be absent during a preferred activity.

- Gifted autistic girls may camouflage their traits for years, so early history and home–school discrepancies are given particular weight.
- We found no published Saudi regulatory text that either permits or prohibits combining a gifted program with learning disabilities services.

## Glossary of terms

| | |
|---|---|
| **Twice-exceptional** Twice-exceptional (2e) | A student who has high performance potential in one or more domains and, at the same time, one or more disabilities. |
| **Masking** | Giftedness compensating for the disability and the disability depressing the giftedness, so that performance appears average. |

| | |
|---|---|
| **Normative comparison** | Comparing a student's performance with that of same-age peers. |
| **Intra-individual comparison** | Comparing a student's performance in one domain with the student's abilities in other domains. |
| **Compensatory strategies** | Methods a student uses to overcome a difficulty, such as memorization or reliance on context, which may conceal it. |
| **Dual differentiation** | Instruction that provides challenge in areas of strength and support in areas of difficulty at the same time. |

| | |
|---|---|
| **Cognitive Proficiency Index** Cognitive Proficiency Index (CPI) | A Wechsler index calculated from the working memory and processing speed subtests and compared with the General Ability Index to describe an uneven profile. |
| **Assessment protocol** | A set of instruments and procedures planned in advance, each answering a specific question in the assessment. |
| **Camouflaging** | An autistic person's concealment of their traits by observing and imitating others and memorizing ready-made phrases, at an exhausting cost in effort. |
| **Dysgraphia** | A term used for a difficulty with handwriting or a difficulty with spelling and written expression; its counterpart in the diagnostic manual is specific learning disorder with impairment in written expression. |

## Review questions

1. A specialist wrote in a report: "The student is twice-exceptional because he exceeds the gifted program criteria and has a diagnosis of autism." Which of the two reference definitions does this judgment rest on? And what does the other definition add to what the report should recommend? **(Objective 1)**
2. A fifth-grade student in the gifted program has seen her grades decline in reading-heavy subjects, and her teacher attributes this to her being "busy with competitions." In which of the three groups in Figure 13.2 would you tentatively place her? And what might each side be concealing in the other? **(Objective 2)**

3. A student has a General Ability Index of 128 and reading fluency of 90, and has received no intervention because he never showed up in screening in the first place. Which of the models in Table 13.2 would miss him, and why? And what would each side of the debate over normative versus intra-individual comparison say about his case? **(Objective 3)**
4. Design an assessment protocol for an eighth-grade student referred because his answers on essay tests are much weaker than his oral participation. Choose six components from Table 13.3 and explain the reason for each choice, as well as one component you do not need and why. **(Objective 4)**
5. A student's General Ability Index came out at 126 and his Cognitive Proficiency Index at 99. What two questions would you ask before interpreting this difference? And why is it not enough, even if significant and rare, to diagnose him as twice-exceptional? **(Objective 5)**
6. The father of a gifted student says, "My son can focus for three hours on coding games, so there is no way he has an attention disorder." How would you respond, drawing on what the diagnostic manual states? And what evidence would you need to favor one of the three possibilities? **(Objective 6)**
7. Compare the cases of Muhannad and Faris in Tables 13.7 and 13.9: what three pieces of evidence separated the two decisions? And why is it unsound, in Muhannad's case, to rely on the family's rating of only three symptoms to rule out Criterion C? **(Objectives 6 and 7)**
8. Why was Joud's autism not identified until secondary school? Which evidence in Table 13.8 carried the most weight in the judgment, and why was the observation session alone not considered reassuring? **(Objectives 6 and 7)**

## Case for application

### Arwa, age 13, seventh grade (first year of intermediate school)

**The story:** Arwa has been hard of hearing since birth, uses two hearing aids and studies in an inclusion program. Her art teacher considers her designs to be "on another level"; she once built a model of her school to a precise scale without anyone asking her to. Her Arabic teacher, however, sees her as "weak in comprehension," no one has nominated her for the gifted program, and she has not been registered for the Mawhiba test. With her parent's consent, the special education teacher gathered the following data over the course of a month.

**Table 13.11: Data in Arwa's case.**

| Source | Result |
|---|---|
| **Audiological assessment (last year)** | Moderate bilateral sensorineural hearing loss; 55 decibels in the better ear with hearing aids |
| **Nonverbal ability test, administered with written instructions** | Standard score 134 |
| **Classroom reading comprehension test** | Below her grade's average |
| **Mathematics** | Excellent in geometry, average on word problems |
| **Art teacher** | A portfolio of twelve designs, each rated "highly outstanding" |

| Source | Result |
|---|---|
| **Mother** | "She lip-reads well at home, but struggles in class when there is a lot of talk and more than one person speaking" |

**Tasks:** (1) Which of the three groups does Arwa currently fall into? And why has no one noticed her? (2) What is the value of the nonverbal ability test here, and what are its limits in judging her giftedness? (3) What evidence is missing before any judgment about giftedness can be made, and what accommodations would you request in the identification instruments? (4) Does her hearing loss alone explain her weak reading comprehension, or is further assessment needed? And what decision would you propose now, and why? **(Objectives 2, 4, 6 and 7)**

## Learning activities

Any activity carried out with students or families is conducted with the consent of the family and the school administration, without any data that could identify the student, and under the supervision of a specialist; its results are not used in any decision concerning the student.

1. **A two-lens profile:** Choose the case of a student with a disability or a gifted student from the cases in this book or from a published case, and rewrite the assessment through two lenses: what does the question of giftedness add to a disability case, or the question of disability to a giftedness case? Product: a three-column evidence table and a decision paragraph, in about two hours. **(Objectives 4 and 6)**
2. **Reviewing criteria:** Set out the admission requirements for gifted classes and the eligibility requirements for learning disabilities from the Saudi documents cited in Section 13.9 in two side-by-side tables, and identify where Yasser and Arwa fall between them and what the texts leave

unresolved. Product: a table and a paragraph, in about an hour and a half. **(Objective 3)**

3. **Interview:** Interview the family or teacher of a twice-exceptional student and ask: When did they notice each of the two sides? Who saw what? And how long did it take to arrive at the full picture? *Desk-based alternative:* Write the same interview in the voice of Joud's mother, drawing on the details of her case, then identify in her answers three pieces of information that would never have surfaced at school. Product: one page, in about an hour. **(Objectives 2 and 6)**

## Chapter references

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (1437 AH). *Organizational and procedural guide for the school program for gifted girls* [In Arabic]. General Directorate for Gifted Girls. https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20(1).pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Almulhim, N., Aljughaiman, A., Alnaim, F., Alsayed, A., Alfawzan, S., & Alarfaj, A. (2026). Hidden potential: An exploratory study of the prevalence of twice-exceptional students in Saudi Arabia. *Journal of Intelligence, 14*(5), Article 83. https://doi.org/10.3390/jintelligence14050083

Alsamani, O. A., Alsamiri, Y. A., & Alabdulwahab, R. A. (2026). Twice-exceptionality in primary schools: Teachers' perceived barriers to identifying and supporting 2e students. *Acta Psychologica, 262*, Article 106031. https://doi.org/10.1016/j.actpsy.2025.106031

Alsamiri, Y. A., & Alsamani, O. A. (2023). From self-doubt to self-efficacy: Saudi elementary teachers' reflections on their experiences and challenges of teaching 2E students. *Islamic University Journal for Educational and Social Sciences, 16*(2), 367–388. https://doi.org/10.36046/2162-000-016-019

Alsamiri, Y. A., & Alsamani, O. A. (2026). Teachers' perceptions of professional development needs for identifying and supporting twice-exceptional students in inclusive schools in Saudi Arabia. *Education Sciences, 16*(3), Article 366. https://doi.org/10.3390/educsci16030366

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Assouline, S. G., Foley Nicpon, M., & Whiteman, C. (2010). Cognitive and psychosocial characteristics of gifted students with written language disability. *Gifted Child Quarterly, 54*(2), 102–115.

Baldwin, L., Baum, S., Pereles, D., & Hughes, C. (2015). Twice-exceptional learners: The journey toward a shared vision. *Gifted Child Today, 38*(4), 206–214.

Baum, S. M., & Owen, S. V. (2004). *To be gifted and learning disabled: Strategies for helping bright students with LD, ADHD, and more*. Creative Learning Press.

Baum, S. M., Schader, R. M., & Hébert, T. P. (2014). Through a different lens: Reflecting on a strengths-based, talent-focused approach for twice-exceptional learners. *Gifted Child Quarterly, 58*(4), 311–327.

Bracken, B. A., & McCallum, R. S. (2016). *Universal Nonverbal Intelligence Test* (2nd ed.). PRO-ED.

Brody, L. E., & Mills, C. J. (1997). Gifted children with learning disabilities: A review of the issues. *Journal of Learning Disabilities, 30*(3), 282–296.

Crepeau-Hobson, F., & Bianco, M. (2011). Identification of gifted students with learning disabilities in a Response-to-Intervention era. *Psychology in the Schools, 48*(2), 102–109.

Ejbara, M. A. (2024). Psychometric characteristics of a Saudi version of Wechsler Intelligence Scale for Preschool and Young Children. *Dirasat: Educational Sciences, 51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Foley Nicpon, M., Allmon, A., Sieck, B., & Stinson, R. D. (2011). Empirical investigation of twice-exceptionality: Where have we been and where are we going? *Gifted Child Quarterly, 55*(1), 3–17.

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M.-C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 47*(8), 2519–2534. https://doi.org/10.1007/s10803-017-3166-5

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Wiley.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111–138.

Lovett, B. J., & Lewandowski, L. J. (2006). Gifted students with learning disabilities: Who are they? *Journal of Learning Disabilities, 39*(6), 515–527.

Maddocks, D. L. S. (2018). The identification of students who are gifted and have a learning disability: A comparison of different diagnostic criteria. *Gifted Child Quarterly, 62*(2), 175–192.

Maddocks, D. L. S. (2020). Cognitive and achievement characteristics of students from a national sample identified as potentially twice exceptional (gifted with a learning disability). *Gifted Child Quarterly, 64*(1), 3–18.

National Association for Gifted Children. (2018). *Use of the WISC-V for gifted and twice exceptional identification* [Position statement]. NAGC.

Reis, S. M., Baum, S. M., & Burke, E. (2014). An operational definition of twice-exceptional learners: Implications and applications. *Gifted Child Quarterly, 58*(3), 217–230.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing* (2nd ed.). PRO-ED.

Webb, J. T., Amend, E. R., Webb, N. E., Goerss, J., Beljan, P., & Olenchak, F. R. (2005). *Misdiagnosis and dual diagnoses of gifted children and adults*. Great Potential Press.

Part Four

# Applied issues

From diagnosis to decision

Chapter 14: Fairness in diagnosis

Chapter 15: From diagnosis to decision

# 14 Chapter 14 Fairness in diagnosis

A student can be wronged in diagnosis without anyone intending it: a hard-of-hearing child comes close to an autism spectrum disorder diagnosis because no one checked his hearing; another student is nearly diagnosed with intellectual disability because he was tested in a language he has not mastered; and the youngest child in the class has activity that is perfectly expected for his age read as a disorder. In this chapter we move from the individual case to a broader question: how can we tell whether the diagnostic system as a whole is fair, and whether its fairness reaches beyond the single student?

**Scene from an end-of-year report**

A special education supervisor in an education directorate reviewed the referral records of the schools in her charge. Three things caught her eye. Boys were referred for behavior roughly four times as often as girls. Students who were not native speakers of Arabic made up twice as large a share of learning disabilities referrals as of the school population. And the schools in remote villages had not referred a single student to the gifted program in three years.

> She asked herself: do these numbers reflect real differences among students, or do they reflect the way we ourselves work?

The fair answer is that both may be true at once. Our aim in this chapter is to show how to tell the two apart.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Calculate the risk index and risk ratio from a school's data, and explain the limits of the risk ratio when numbers are small.
2. Identify, in a given scenario, the point at which bias entered the diagnostic pathway, and propose a step to limit it.
3. Calculate the representation index and the minimum fair representation in a gifted program, and propose two steps to raise the representation of an underrepresented group.
4. Distinguish, in the description of a student learning in a second language, between conversational fluency and academic proficiency, and specify the evidence needed before referral.
5. Critique the interpretation of a score pattern with the Culture-Language Interpretive Matrix in light of studies that have examined its validity.
6. Name the cognitive bias at work in a given assessment scenario, and propose a strategy to reduce it.
7. Judge whether the evidence is sufficient for a fair decision in a new case, and justify that judgment.

# Introduction

A student can suffer injustice in diagnosis without anyone meaning it to happen, when his hearing goes unchecked, when he is tested in a language he has not mastered, or when activity expected for his age is read as a disorder. Yet fairness is not tested in the individual case alone. Patterns of referral and diagnosis across the schools of an entire education directorate, like those the supervisor noticed in the opening scene, may reveal that certain groups are referred more often, or less often, than others. Judging the fairness of a diagnostic system requires distinguishing disparities in rates from actual bias, knowing the points through which bias enters, and knowing the tools that expose it.

We open the chapter by defining what fairness in diagnosis means (Section 14.1), then turn to disproportionality, the ways it is measured and what happens when numbers are small (Section 14.2), and the points through which bias enters, along with its sources within the instrument (Section 14.3). We then address fairness in gifted identification, the representation index and the fair minimum (Section 14.4), followed by language and culture, the Culture-Language Interpretive Matrix and what studies say about its validity (Section 14.5), and assessor biases (Section 14.6). We offer a fairness checklist (Section 14.7) and close with three cases and a school audit (Section 14.8).

## 14.1 What is fairness in diagnosis?

Fairness in diagnosis does not mean that every group should appear in special education or gifted programs in equal proportions. Some differences are real: attention-deficit/hyperactivity disorder, for example, genuinely is more prevalent among boys. Fairness means, rather, that every student has the same chance of being identified if he needs services, and that no student is misdiagnosed because of something unrelated to need: his language, his sex, his family's poverty, the location of his school, or the expectations of the person assessing him. The "Standards for Educational and Psychological Testing"

devote their third chapter to fairness in testing, defining it as the extent to which score interpretations are equally valid for all individuals and groups to whom the instrument is administered. They approach it from four angles: fair treatment of examinees during administration, freedom of measurement from bias, access to the construct the test measures without obstacles unrelated to it, such as the language of the instructions or the format of presentation, and the validity of the interpretation of an individual's own score (AERA et al., 2014).

Diagnostic error runs in two directions, and each carries a cost:

**Overdiagnosis**

A student who does not need services is diagnosed: he may be separated from his peers, expectations of him drop, he carries a label that does not fit him, and resources are diverted from those who need them.

**Underdiagnosis**

A student who needs services goes unidentified: he struggles for years without understanding or support, is blamed for what lies beyond his control, and loses time when intervention would have done the most good.

## 14.2 Disproportionality and how it is measured

Disproportionality is present when the share of a group in a diagnostic category or program differs markedly from its expected share. It is measured with two simple indicators:

- **Risk index:** the proportion of a group's members who have been diagnosed or referred, out of the group's total number.
- **Risk ratio:** the risk index of one group divided by the risk index of a comparison group. A value of 1 indicates proportionate representation; a value above 1 indicates overrepresentation, and a value below 1 underrepresentation. U.S. regulations use the risk ratio to determine

significant disproportionality at the state and district levels (U.S. Department of Education, 2024).

**Worked example: behavior referrals in one school**

1. The school has 400 boys and 380 girls. Thirty-six boys and 8 girls were referred for behavior.
2. Risk index for boys: 36 ÷ 400 = 9%; for girls: 8 ÷ 380 ≈ 2.1%.
3. Risk ratio for boys compared with girls: 9 ÷ 2.1 ≈ 4.3.
4. In other words, a boy is more than four times as likely as a girl to be referred for behavior.
5. **Caution:** the figure of 4.3 rests on only 8 girls. Had two more girls been referred (10 girls), the risk ratio would be about 3.4; had two fewer been referred (6 girls), it would be about 5.7. A change of two referrals moves the ratio anywhere between “three times” and “nearly six times.”

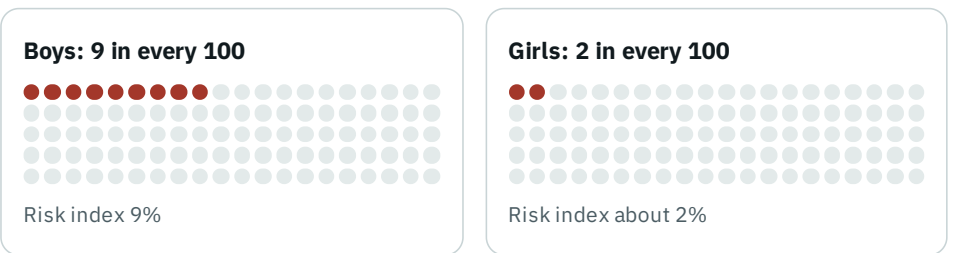



**Figure 14.1:** The risk index made visible. Each grid has one hundred dots, each dot representing a student; the shaded dots are those referred for behavior. The risk ratio here is about 4.3.

⚠ **The number raises a question; it does not answer it**

A high risk ratio does not, on its own, prove bias. Behavior problems may genuinely be more common among boys. The right questions are these: is the difference larger than studies of real differences would lead us to expect? Are boys referred for behaviors that are overlooked in girls? And are internalizing problems in girls, such as anxiety, withdrawal and sadness, missed because they do not disrupt the classroom the way overt behavior does? The literature holds a genuine debate on this point: one large study found that some groups were less likely to be diagnosed than expected once other factors were controlled (Morgan et al., 2015), while other researchers criticized it on the grounds that the method of control itself may conceal bias (Skiba et al., 2016). The lesson is that the numbers mark the beginning of an inquiry and an invitation to complete it.

## When numbers are small

The risk ratio is a fraction whose denominator is the risk index of the comparison group. When only a few members of that group have been referred, the denominator becomes fragile, and a change of one or two cases is enough to make the result jump, as the caution in the previous example showed. When the number is zero, the division cannot be performed at all. For this reason, a risk ratio in a small school, or in a rare category, should be read only alongside the raw counts on which it rests. More useful than a single year's ratio is to pool data across several years and look at the trend over time.

**For the specialist: how do U.S. regulations deal with the instability of the ratio?**

Section §300.647 of the implementing regulations of the Individuals with Disabilities Education Act (IDEA) governs how "significant disproportionality" is calculated, and it addresses the problem of small numbers with three tools. **The first** is a minimum cell size, meaning the number of group members diagnosed or referred, together with a minimum group size (**Minimum n-size**), meaning the group's total membership. The ratio is not calculated unless both thresholds are met, and the regulation regards a threshold of no more than 10 for the cell and 30 for the group as presumptively reasonable. **The second** is the alternate risk ratio: when the comparison group in the district falls below these thresholds, the group's risk index is compared with the risk index of all other students across the state. **The third** is that a state may decline to identify significant disproportionality unless the ratio exceeds its threshold for several consecutive years, up to three (U.S. Department of Education, 2024). We know of no published Saudi text that sets out a method or a threshold for monitoring disproportionality, but the same logic holds for any education directorate reviewing its own data.

## 14.3 Where does bias enter?

Many people assume that bias in diagnosis comes from the test. But the test is only one of four points along the way, and bias can enter at any of them:

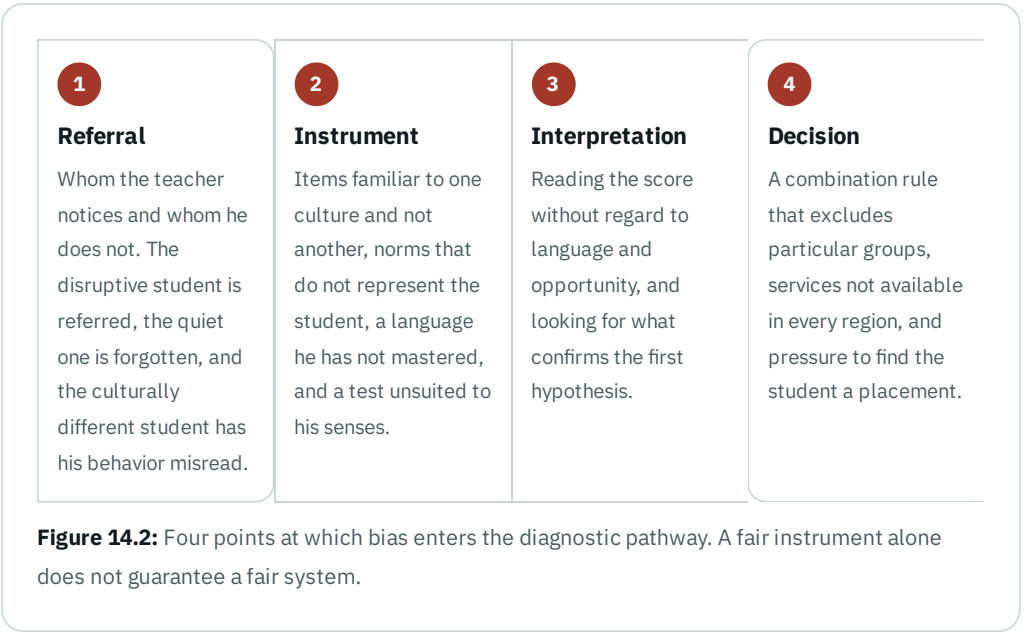



**Figure 14.2:** Four points at which bias enters the diagnostic pathway. A fair instrument alone does not guarantee a fair system.

## Sources of bias in the instrument

**Table 14.1: Sources of bias in the instrument and ways of detecting it.**

| Type | What it means | Example |
|---|---|---|
| **Content bias** | Items that assume knowledge or experience not equally available to all students | A question about snow or a train put to a child who has never seen either, or about vocabulary specific to one region |
| **Norm bias** | A norming sample that does not represent the student | Comparing a Saudi student with American norms, or a rural student with urban-only norms |
| **Language bias** | A test that measures ability through a language the student has not mastered | A student whose first language is not Arabic is tested on vocabulary and verbal comprehension in Arabic, and his verbal ability score drops because of his weak Arabic |

| Type | What it means | Example |
|---|---|---|
| **Predictive bias** | The same score predicts different performance in two groups. It is examined statistically by comparing the regression lines in the two groups: the slope of each and its intercept with the axis | A test that predicts the success of urban students in a program better than it predicts the success of village students |
| **Situational bias** | Testing conditions affect one group more than others | An unfamiliar examiner in a formal setting with a child unused to dealing with strangers |
| **Differential item functioning and measurement noninvariance** | An item on which two groups perform differently even though they are equal in the trait being measured, or an instrument that does not measure the same construct in the same way in the two groups | An item on a verbal reasoning scale that is harder for students from one region because it contains a word common in another dialect, even though students from the two regions with equal total scores should perform equally on it |

The last row of the table differs from the others: rather than describing a type of bias, it describes **two standard methods for detecting it**. The first is **differential item functioning** (**Differential item functioning, DIF**). The idea is to compare students from two groups who are *equal in ability*, rather than comparing the two groups as wholes. If an item is harder for one group even though its members are equal in ability, then for that group the item is measuring something else as well, such as cultural knowledge or a local word. A difference between the two groups' mean scores on an item does not by itself indicate bias, since it may reflect a real difference in the trait. The second is **measurement invariance** (**Measurement invariance**), which asks at the level of the whole instrument: is the factor structure the same in both groups? Are the items related to the factors with the same strength? Does the same score

signify the same level of the trait? If invariance does not hold, comparing the two groups' means is not warranted, nor is applying a single cutoff score to both (AERA et al., 2014). The practical question for the practitioner is whether the manual of the instrument he uses reports the results of these analyses for the groups with whom he uses it, such as boys and girls, different regions, and speakers of languages other than Arabic. If it does not, fairness remains an untested assumption.

## 14.4 Fairness in gifted identification

The examples in the preceding sections all concern overrepresentation, where a group appears in special education more often than it should. In gifted programs, the best-known problem runs the other way: **underrepresentation**. Students from low-income families, from remote schools, and those learning in a second language are usually fewer in these programs than their share of the student body. Nor does the matter stop at the test. Groups differ in the proportions who reach advanced levels of performance in the first place, differences known as **excellence gaps** (**Excellence gaps**), and these gaps reflect, in large part, differences in opportunity more than in potential (Plucker & Peters, 2016). A program that admits only those who have already reached advanced performance inherits these gaps, then widens them by serving those who were already ahead.

### The representation index and the fair minimum

The risk index and risk ratio apply here as well, but the gifted education literature often relies on two indicators that are easier to read:

- **Representation index** (**Representation index**): the group's share among program students divided by its share among students at the same grade level. A result of 1 means proportionate representation; anything below it means underrepresentation.

- **Equity formula** (**Equity formula**), proposed by Donna Ford: it accepts that the two percentages may differ somewhat, since perfect correspondence is unrealistic, but it places a ceiling on the difference. The minimum acceptable representation of a group in the program is its share among students minus 20% of that share, that is, **its share among students × 0.8**. Anything below that counts as underrepresentation that calls for review (Ford, 2013).

The 20% margin is a practical convention. It spares the program from chasing small differences while obliging it to account for large ones.

**Worked example: a gifted program in an education directorate**

1. Grades three through six in an education directorate enroll 4,000 students, 1,200 of them (30%) in village schools. The gifted program has 200 students, 20 of them (10%) from village schools.
2. Representation index for village students: 10 ÷ 30 ≈ 0.33. Their representation in the program is thus about one third of their share of the student body.
3. Fair minimum under Ford's formula: 30% × 0.8 = 24%, that is, at least 48 of the 200 students. The shortfall is 28 students.
4. In risk ratio terms: the risk index for village students is 20 ÷ 1,200 ≈ 1.7%, for urban students 180 ÷ 2,800 ≈ 6.4%, and the risk ratio ≈ 0.26. A village student, in other words, is about one quarter as likely as an urban student to be identified as gifted.
5. **The next question** is "Where are they falling out?" before "How many should we add?" Did nomination or registration reach them? Were they given the test? Did they pass it and then fail to enroll because the program was too far away? Each point calls for a different remedy.

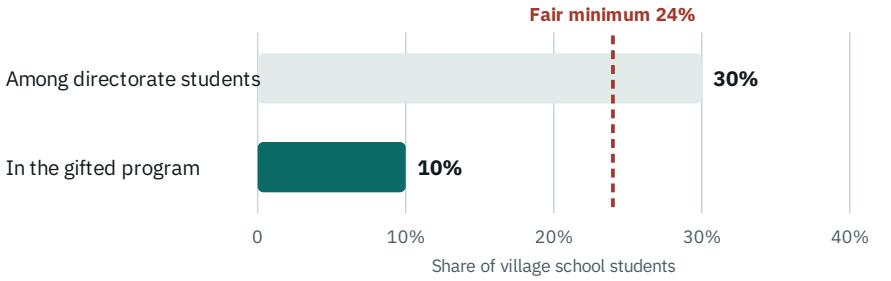



**Figure 14.3:** Share of village school students among the directorate's students and in the gifted program, in the worked example. The dashed line marks the fair minimum under Ford's formula (30% × 0.8 = 24%). Anything below it, like the second bar, is underrepresentation that calls for review.

## Where does the shortfall come from?

Studies indicate that the shortfall accumulates at points close to those in Figure 14.2, with one more point added after the decision: the service itself.

- **Nomination:** when the door to a program is teacher nomination or a family request, who walks through it reflects the teacher's judgment and the family's awareness as much as the student's ability. A large U.S. study found that the likelihood of a student being identified as gifted varied with the background of his teacher, even when students' test achievement was equal (Grissom & Redding, 2016). In Saudi Arabia, Alfaiz examined the accuracy of parent and teacher nominations in identifying gifted students (Alfaiz, 2026). When a large U.S. school district moved from nomination to **universal screening** (**Universal screening**), administering an initial instrument to every student in the grade, the representation of students from low-income families and from minority groups rose markedly, because many of them had never been nominated in the first place (Card & Giuliano, 2016).
- **Instrument and norms:** comparing a student in a less advantaged school with a national sample measures his opportunities as much as his ability.

Hence the proposal of **local norms** (**Local norms**), which compare the student with peers in his own school or in schools with similar circumstances; they have been found to increase the representation of underrepresented groups (Peters et al., 2019).

- **Decision rule:** a single high cutoff score on a single instrument, or a requirement to meet every criterion at once, narrows the door for students with a clear strength in one area and a weakness in another.
- **Service:** a program offered only in urban centers, or one that requires transportation the family does not have, loses students who have already been identified.

No single measure solves the problem. Peters, reviewing what these measures have achieved, argues that the effect of each on representation is limited, and that much of the gap forms years before identification, in opportunities to learn themselves, so that better identification alone cannot close it (Peters, 2022). Fair identification is necessary, then, but not sufficient.

**In the Saudi context: fairness issues**

Four issues in the Saudi system bear on fairness in diagnosis and identification. For each, we present what has been published, and we state plainly what we could not find in print.

1. **Voluntary registration and fees.** The Mawhiba Multiple Cognitive Abilities Test, the instrument of the National Program for Gifted Identification, is administered to students whose families register them through the National Center for Assessment website, with a parental contribution of 200 riyals from which supported students are exempt; a student may sit the test once at each level (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]). In the absence of universal screening, the first gate here is the family's decision: a student whose family does not know about the program, or does not see it as a priority, never reaches the test, however able he may be. The exemption eases the barrier of cost, but the barrier of awareness remains. We found no published data on registration rates by region, school or family income.
2. **Different admission rules in boys' and girls' programs.** The guide to the School Program for the Care of Gifted Girls admits the "top 5% of the outcomes" of the national project (Ministry of Education, 1437 AH [Arabic]). The gifted classes guide, by contrast, nominates students classified in the project and fills the remaining places using criteria of which at least two must be met, one of them a cognitive ability test with a score of no less than 120 (Ministry of Education, n.d. [Arabic]). The "top 5%" is calculated from test-takers alone, so it depends on which girls registered and which did not. And because boys and girls are educated separately, those who nominate and assess in girls' schools are mostly women, and those who nominate in boys' schools mostly

men, so any difference in nomination standards between the two groups turns into a sex difference in representation. We found no published data comparing referral rates to special education, or identification rates in gifted programs, across the two systems. On the other hand, survey studies have been published that offer a basis for comparison, among them a study of 5,495 boys and girls in Al-Zulfi Governorate that estimated the prevalence of dyslexia at about 6.5% among boys and 3.8% among girls (Aldakhil et al., 2023). Because figures of this kind do not pass through referral, they can serve as a benchmark against which actual referral rates in schools are compared.

3. **Regional variation.** The national test standardizes the instrument across regions, but we do not know how everything around it is distributed. The Special Education Organizational Guide requires that assessment instruments be administered and interpreted by "qualified specialists" (Ministry of Education, 1436 AH [Arabic]), yet we found no published data on their availability in cities and villages, nor on the distribution of gifted classes and programs. The peer-reviewed studies published on national test scores come from particular regions, such as a study of students in Al-Ahsa (Almulhim et al., 2026), and so do not allow regions to be compared.
4. **Learners of Arabic as a second language and non-Saudi students.** The Regulations exclude from learning disabilities difficulties attributable to "learning conditions or family care" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). This exclusion is what protects a student learning in a second language from misdiagnosis, but the published documents do not explain how it is verified. Nor did we find a published text specifying the eligibility of non-Saudi students for the National Program for Gifted Identification, or any data on their participation in it.

**A question for reflection:** if your education directorate wanted to know whether the Mawhiba test reaches students in all its schools equally, which three pieces of data would you request first, and from which body?

## 14.5 Language and culture

### The language of conversation and the language of learning

One of the most common sources of misdiagnosis is confusing second language learning with disability. Jim Cummins distinguished between two levels of language proficiency: **basic interpersonal communicative skills** (**Basic Interpersonal Communicative Skills, BICS**), the language of everyday talk and the playground, in which a second language learner reaches fluency close to that of peers in **about two years**; and **cognitive academic language proficiency** (**Cognitive Academic Language Proficiency, CALP**), the language of textbooks, abstract concepts and lengthy instructions, which takes roughly **five to seven years** to reach the level of native speakers (Cummins, 1981).




**Figure 14.4:** Approximate time, in years, that a second language learner needs to approach native-speaking peers, after Cummins (Cummins, 1981). Conversational fluency comes in about two years; academic proficiency between the fifth and seventh year. A student who speaks Arabic fluently at recess after two years may need several more years before he can learn in it as effectively as his classmates.

The danger is that a teacher sees a student chatting fluently in Arabic with classmates, then stumbling in reading and science, and concludes that he has a learning disability, when in fact he has not yet attained academic language proficiency. Saudi schools enroll many students from expatriate families whose first language is not Arabic, and these students most closely resemble those on whom the framework was built.

These figures have limits that readers should know before using them. **First,** they are drawn from data on immigrant children learning English in Canada, and they do not carry over unchanged to another language or another education system. **Second,** it is often suggested that an Arab student who speaks a colloquial dialect encounters in Modern Standard Arabic something akin to a second academic language, since it departs from his dialect in vocabulary and structure. The analogy is reasonable, but it is an inference awaiting research, and it does not justify applying Cummins's timeframes to that student. **Third,** the distinction between the two levels has itself drawn theoretical criticism: among other things, that it reduces language proficiency to two poles, and that it can be used to explain a student's difficulties as a deficit in him rather than in his instruction. Cummins has addressed these criticisms, presenting the distinction as a tool for understanding the difference between fluency and learning, with approximate timeframes (Cummins, 2008). The practical conclusion nonetheless stands: **fluency in conversation is not evidence of command of the language of learning**.

—

How can we tell whether the low scores of a linguistically diverse student reflect his ability or his language? Flanagan and colleagues (Flanagan et al., 2013) proposed a method called the **Culture-Language Interpretive Matrix** (C-LIM). The idea is to classify subtests along two dimensions: how much **cultural content** they carry, and how much **language** they demand. The student's scores are then placed in the cells of the matrix, and the scores in each cell are compared with an **expected range** for a student whose linguistic

and cultural difference from the norming sample is similar to his. A highly different student, such as a recent arrival, is expected to show steeper drops in the high-loading cells than a student who has spent years in schools taught in that language. If scores decline steadily as cultural and linguistic loading increases, *and fall within the range expected for such a student*, language and culture likely explain the decline. If they fall below the expected range, however, language alone is not a sufficient explanation, and another cause should be sought. The criterion, then, is the size of the graded decline relative to students with similar exposure to the language, not merely its presence (Flanagan et al., 2013).

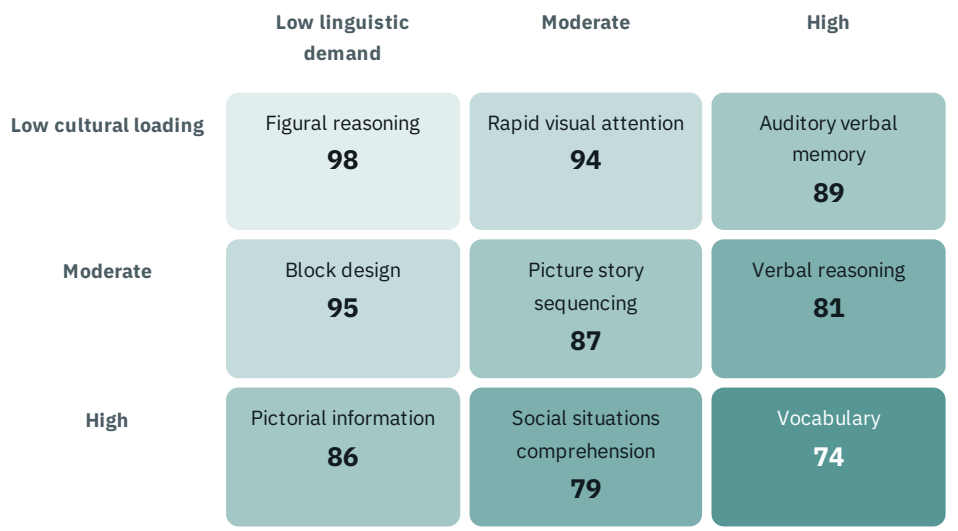



**Figure 14.5:** A simplified hypothetical example of the Culture-Language Interpretive Matrix. The numbers are standard scores with a mean of 100. Scores decline steadily from the top cell to the bottom, that is, as cultural and linguistic loading increases, from 98 to 74, and cell shading deepens in the same direction. This pattern is *consistent with* an effect of language and culture, but it does not prove one on its own (see the next paragraph). The task names here are generic and their classification is illustrative, independent of how particular subtests are classified in published tables.

## What studies say about the validity of the matrix

The matrix spread through practice faster than it was examined. And when it was examined, the results were troubling. Kranzler and colleagues administered a cognitive battery to students learning English as a second language and did not find the graded decline the matrix predicts with the regularity it assumes (Kranzler et al., 2010). Styck and Watkins studied the files of students referred for evaluation, some learning English as a second language and some native speakers, and found that the matrix patterns distinguished poorly between the two groups (Styck & Watkins, 2013). This is to be expected once we recall that a native speaker with a language difficulty will also show declining scores as linguistic demands increase. The graded pattern may mean "second language," or it may mean "genuine language weakness," and the matrix alone cannot tell them apart.

Add to this that its classifications are built on English-language tests, and we know of no published classification of Arabic tests by cultural and linguistic loading. The matrix, then, is **a tool for organizing one's thinking before judgment**. Its core idea remains valid in any language: **look at the pattern of scores as well as their level**. But the pattern is a hypothesis that needs independent evidence to confirm or refute it, of which three kinds matter most: the student's proficiency in his first language, measured directly rather than through the family's report; his level in the second academic language, compared with peers who have had the same length of exposure; and his response to direct, focused instruction monitored over several weeks. Among the tools for the last kind of evidence is dynamic assessment, which measures how much the student benefits from brief instruction provided during the assessment itself (see Chapter 5), complemented by asking the family about the student's language and educational history (Rhodes et al., 2005).

**Common misconceptions**

- **"A nonverbal test is fair to all students."** A nonverbal test reduces the influence of language, but it does not remove the influence of experience and schooling, and it does not measure the verbal and quantitative abilities a student needs in most academic programs. For this reason it is not sufficient on its own for identifying gifted students among second language learners; it is more useful to compare the student with others who share similar opportunities, using verbal, quantitative and nonverbal measures together (Lohman, 2005).
- **"Translating a test makes it valid for the student."** Translation can change the difficulty and meaning of items, and it does not carry over the norms. A translated instrument needs evidence of measurement invariance before its scores can be compared with the original.
- **"Equal proportions mean fairness, and unequal ones mean injustice."** The standard is equal chances of identification and equally valid interpretation, not identical proportions. A difference raises a question to be investigated.

## 14.6 Assessor biases

Assessors are human, and the mind sometimes falls back on mental shortcuts in judgment that lead to predictable errors. These errors have long been studied in medicine, where many diagnostic errors were found to stem from the way clinicians think more than from gaps in their knowledge (Croskerry, 2003). The most prominent in our field are the following:

**Table 14.2: Common assessor biases and how to reduce them.**

| Bias | How it shows | Example | What reduces it |
|---|---|---|---|
| **Anchoring** | The first idea governs everything that follows | A teacher notices that a student does not turn when his name is called, assumes autism from the outset, and never asks for a hearing check | Writing at least three alternative hypotheses before the assessment |
| **Confirmation bias** | Seeking what supports the hypothesis and overlooking what contradicts it | Dismissing a parent's rating because it differs from what the team expected, instead of asking why the two differ | Deliberately searching for evidence that contradicts the hypothesis |
| **Diagnostic overshadowing** | Attributing every symptom to a known diagnosis | Anxiety in a student with intellectual disability is explained away as "part of his disability" | Asking: does the known diagnosis actually explain this symptom? |
| **Halo effect, and its negative counterpart, the horn effect** | One salient trait colors judgment of everything else: a positive one raises it, a negative one lowers it | Nominating well-behaved high achievers for giftedness while overlooking a gifted girl who asks endless questions and keeps untidy notebooks | Observation tools with specific items, more than one rater, and judging work without knowing whose it is |
| **Availability bias** | Judging by whatever comes most readily to mind | A teacher who attended a course on attention-deficit/hyperactivity disorder begins to see it in every highly active student | Returning to the written criteria instead of relying on impressions |

**For the teacher: five minutes before you refer or nominate**

You are the first point in Figure 14.2, which is why bias either begins with you or ends with you. Before you sign a referral or nomination form, ask yourself:

- What specific behavior did I see? How often? In which lesson? Write it down as it happened, before any interpretation.
- If another student in the class had done the same thing, would I refer him? And if this student were a girl rather than a boy, or the reverse, or came from a different family?
- Who in my class has never once crossed my mind? The quiet student, the newcomer, the one who has not yet mastered Arabic may need referral or nomination more than those who caught my attention.
- Did I give the student a chance to show what he can do, through an open-ended task or in a form other than writing, before judging him?

## 14.7 A fairness checklist

These are questions for the team to ask before any diagnostic decision, arranged according to the points in Figure 14.2:

- [ ] **Referral**
  Did the student receive sound instruction and adequate support before referral? And was this documented?

- [ ] **Referral**
  Was the student's behavior compared with that of peers of the same age and background?

- [ ] **Instrument**
  Were hearing and vision checked?

- [ ] **Instrument**
  Was the instrument normed on a sample resembling the student in language and environment? And when was it normed?

- [ ] **Instrument**
  Was it administered in the student's stronger language or through his means of communication? Did he need a nonverbal instrument?

- [ ] **Instrument**
  Does the instrument's manual present evidence that its items are free of differential functioning, and of measurement invariance, for the group to which the student belongs?

- [ ] **Interpretation**
  Was the pattern of scores considered alongside their level? Were confidence intervals taken into account?

- [ ] **Interpretation**
  Did the team write alternative hypotheses and look for evidence against its first hypothesis?

- [ ] **Interpretation**
  Was information gathered from home and school, and was any conflict between them understood?

- [ ] **Decision**
  Is the decision rule made public, and does it avoid excluding any particular group without justification?

- [ ] **Decision**
  At the program level: are representation rates reviewed every year using raw counts and compared with the fair minimum?

- [ ] **Decision**
  Did the family take part in the decision and understand it?

## 14.8 Three cases and a school audit

### Case one: Zain, age 10

Most likely a second language effect

**The story:** Zain arrived with his family two and a half years ago; his first language is Urdu. He speaks Arabic fluently with his classmates but is struggling in reading, science and social studies. His teacher referred him on suspicion of a learning disability.

**Table 14.3: Evidence in Zain's case.**

| Source | Result | What it means |
|---|---|---|
| **Length of time learning Arabic** | Two and a half years | Conversational language mastered; academic language most likely not yet fully developed |
| **Direct measure of academic Arabic** | He read a passage from his grade's science textbook, answered comprehension questions and defined subject vocabulary. His performance fell clearly below that of his Arabic-speaking classmates and close to that of two classmates who had been learning Arabic for the same length of time | His difficulties are within what is expected for someone with his exposure to Arabic |
| **Ability test in Arabic** | Scores declining steadily with linguistic and cultural loading, as in the pattern of Figure 14.5 | A pattern consistent with a language effect, but one that cannot by itself distinguish a second language from a language weakness |

| Source | Result | What it means |
|---|---|---|
| **Family interview with an interpreter** | He reads Urdu well and was a high achiever at his previous school | Important information, but it rests on the family's report alone |
| **Reading sample in Urdu, with help from a teacher at the school who speaks it fluently** | He read a grade-level passage accurately and fluently and summarized what he read | Direct, though unstandardized, evidence of sound reading in his first language |
| **Dynamic assessment in Arabic reading** | Rapid improvement after two sessions of training in decoding strategies | An initial indicator of learning potential that needs further corroboration |

**Decision:** Zain is not diagnosed with a learning disability at this point. A genuine learning disability usually shows itself in both languages, and the direct evidence from the reading sample, beyond what the family reports, indicates that he reads soundly in Urdu. He will receive eight weeks of focused support in academic Arabic, including explicit teaching of subject vocabulary, with weekly measurement of his reading progress compared against that of peers who have been learning Arabic for the same length of time. If he grows as they do, the language explanation is confirmed. If his growth stalls despite the support, his case will be revisited through a more comprehensive evaluation, and the possibility of a learning disability remains open.

> **What evidence could change the decision?** If, after the eight weeks of focused support, his growth in Arabic reading remained clearly slower than that of peers with the same length of exposure, or if a closer assessment of his Urdu reading revealed difficulties with

> accuracy or fluency, the possibility of a learning disability would return to the fore, and the comprehensive evaluation would be completed.

## Case two: Waad, age 9 Horn effect

**The story:** On a gifted nomination scale, Waad's teacher rated her academic ability low, writing: "Her handwriting is very poor, and her notebook is a mess." The program coordinator, however, asked two other teachers to rate Waad's answers on a problem-solving task after they had been retyped on a computer, without telling them whose work it was. Both placed her among the top 10% of girls in the class.

**Decision:** Waad was referred for gifted assessment. Her case illustrates the negative face of the halo effect, known as the **horn effect** (**Horn effect**): one conspicuous negative trait, here her handwriting, colored the judgment of her ability as a whole. Rating work without knowing whose it is offers a simple way to curb it.

An assessment of her fine motor skills and handwriting was also requested, and for a reason that goes beyond the handwriting itself. The wide gap between her problem-solving ability and her written performance may be a sign of **twice-exceptionality**: giftedness combined with a writing difficulty (dysgraphia). A student who carries both is exposed to two errors at once: her difficulty may mask her giftedness, so that she is never nominated, as nearly happened here, and her giftedness may mask her difficulty, so that she is never supported. Both sides are therefore assessed together, rather than stopping at her admission to the program (see Chapter 13).

**What evidence could change the decision?** If the individual gifted assessment returned average results despite the two teachers' ratings, her performance on a single task would be treated as a relative strength that needs further evidence before nomination. And if the assessment of her fine motor skills and handwriting revealed a clear deficit, the question of twice-exceptionality would become part of the decision itself.

## Case three: Ohoud, age 14, ninth grade

**First, open the gate for her**

**The story:** Ohoud attends a small intermediate school in a remote village, with eleven girls in her class. The nearest gifted classes for girls are in a city about an hour and a half away by car. She has never been registered for the national test at any level, and no teacher has ever nominated her. As part of the review the supervisor launched in the opening scene, her education directorate ran open-ended problem-solving activities with girls in the village schools for four weeks. During that time, the mathematics teacher noticed that Ohoud was solving problems that had not yet been taught, and proposing more than one method for some of them.

**Table 14.4: Evidence in Ohoud's case.**

| Source | Result | What it means |
|---|---|---|
| **National test registration record** | Not registered at the first or second level. Her mother said she had never heard of the test | She never reached the first gate at all, and the reason lies in the family's awareness of the test, quite apart from the student's ability |

| Source | Result | What it means |
|---|---|---|
| **School grades** | Top of her class in mathematics and science since seventh grade | A limited indicator: ranking first among eleven girls says little about where she would stand in a larger group |
| **Teacher's observation after the open-ended activities** | Original solutions to unfamiliar problems, and questions that go beyond the lesson | Behavior that would not have surfaced in a classroom that offers no such tasks |
| **Brief group ability test administered by the directorate to all ninth-grade girls in the sector's village schools** | In the top 3% of 240 girls in these schools | A clear strength relative to those who share her opportunities; this is a screening indicator that comes before any eligibility decision |

**Decision:** No judgment is made yet as to whether Ohoud is gifted or not; the first task is to open the gate she never reached. With the school's help, she will be registered for the Mawhiba test this year, and her eligibility for a fee exemption will be checked. She is at the third level, where she is entitled to only one attempt, so that attempt should not be lost to missing information or to a lack of transport to the test center. In the meantime, she will receive mathematics enrichment within her own school, because her need exists whatever the test result. If she qualifies, the team will work out with her family how she can reach the service, since identification without access to services changes nothing.

**What the case teaches:** Ohoud was wronged because no one had measured her in the first place; neither a test nor an assessor ever reached her. This kind of injustice appears in no individual assessment report. Only a system-level review, like the one the supervisor began, can bring it to light.

**What evidence could change the decision?** If her Mawhiba test result fell below the nomination threshold after a registration carried out under suitable conditions, the mathematics enrichment in her school would continue, and her need for a gifted program would be reconsidered. And if it turned out that her family could not reach the service location, the decision would shift to finding an alternative that reaches her in her village.

**School audit: back to the supervisor**

The supervisor in the opening scene has three numbers in front of her. Here are the steps we suggest:

1. **Behavior referrals:** Calculate the risk ratio for each school, record the raw counts next to it, and pool three years of data for small schools before drawing conclusions. Then review a sample of referrals: which behaviors are cited? Are girls referred for the same behaviors? And look for girls with internalizing problems who were never referred.
2. **Learning disabilities referrals of non-native speakers of Arabic:** Review how long these students have been learning Arabic; whether their proficiency in their first language was measured directly, alongside asking the family; whether their academic Arabic was measured against peers with the same length of exposure; and whether the referral was preceded by language support whose effect was monitored.
3. **Village schools missing from gifted programs:** Calculate the representation index and the fair minimum, then trace where students drop out: Were they registered for the national test? Did screening or nomination reach them? Were local norms used? Were they given opportunities to display their abilities? Ohoud's case is an example of what such tracing can uncover.

## Conclusion

The central argument of this chapter is that fairness means an equal chance of being identified for those who need it, freedom from misdiagnosis for those who do not, and valid score interpretation for every group, which runs deeper than equal percentages. It follows that a risk ratio should be treated as a

question that calls for verification, read alongside raw counts and across several years, and that bias should be sought at its four points of entry: referral, instrument, interpretation and decision. It follows, too, that conversational language should not be confused with academic language, that the pattern of scores across cultural and linguistic loading should be read as a hypothesis requiring independent evidence, and that assessor biases should be curbed through alternative hypotheses, multiple raters and a return to the criteria.

Some questions in this chapter nonetheless remain open. Cummins's time frames are estimates drawn from a different context; the Culture-Language Interpretive Matrix discriminates poorly between second language learning and language impairment; and part of the representation gap in gifted programs arises before identification itself, so identification tools alone cannot close it. In the Saudi context, optional registration for the national test makes the family's decision the first gate, and no data are published that would allow comparison across regions or between boys' and girls' programs, so system-level monitoring of fairness remains limited by whatever data are available.

## 14.9 Summary, key terms and activities

### Chapter summary

- Fairness means an equal chance of being identified for those who need it, freedom from misdiagnosis for those who do not, and valid score interpretation for every group; it runs deeper than equal percentages.
- The risk ratio is a simple tool for detecting disproportionality, but it raises a question rather than proving bias, and it becomes unstable when numbers are small, so it is read alongside raw counts and across several years.

- Bias enters at four points: referral, instrument, interpretation and decision. Instrument bias is detected statistically through analyses of differential item functioning and measurement invariance.
- In gifted programs, underrepresentation is the dominant problem. The representation index and the fair minimum detect it, and universal screening and local norms reduce it, but part of the gap arises before identification.
- In the Saudi context, optional registration for the national test makes the family's decision the first gate, and no data are published that would allow comparison across regions or between boys' and girls' programs.
- Conversational language is acquired much faster than academic language, and confusing the two is a common source of misdiagnosis. Cummins's time frames are estimates drawn from a different context.
- The pattern of scores across cultural and linguistic loading is more telling than their level, but the Culture-Language Interpretive Matrix discriminates poorly between second language learning and language impairment, so the pattern it shows is a hypothesis that needs independent evidence.
- Assessor biases are to be expected; they are reduced by alternative hypotheses, rating work without knowing whose it is, multiple raters and a return to the criteria.

## Glossary of terms

| | |
|---|---|
| **Disproportionality** | A clear difference between a group's share of a category or program and its expected share. |
| **Risk index** | The number of students in a group who have been diagnosed, divided by the group's total number. |
| **Risk ratio** | The risk index of one group divided by the risk index of a comparison group. |
| **Representation index** | A group's share among program students divided by its share among all students; a value of 1 indicates proportionate representation. |
| **Equity formula** | A rule that sets the minimum acceptable representation of a group in a gifted program at its share among students multiplied by 0.8. |
| **Excellence gaps** | Differences between groups in the proportions of students who reach advanced levels of performance. |
| **Universal screening** | Administering an initial instrument to all students in a grade or school level, rather than waiting for nomination or a family request. |
| **Differential item functioning** Differential item functioning (DIF) | A difference in two groups' performance on an item among individuals of equal standing on the ability being measured, indicating that the item measures something else besides that ability. |

| | |
|---|---|
| **Measurement invariance** | The property of a scale measuring the same construct in the same way in two groups, so that their scores can be compared and a single cut-off score applied to both. |
| **Basic Interpersonal Communicative Skills (BICS)** | The language of everyday conversation, in which a second language learner reaches fluency close to that of peers in about two years. |
| **Cognitive Academic Language Proficiency (CALP)** | The language of learning and abstract concepts, which takes roughly five to seven years to develop. |
| **Culture-Language Interpretive Matrix (C-LIM)** | A method for interpreting the pattern of scores according to their cultural loading and linguistic demands, by comparing them with the range expected for students whose exposure to the language resembles the student's own. |
| **Confirmation bias** | Seeking evidence that supports a hypothesis while overlooking evidence that contradicts it. |
| **Anchoring** | Allowing the first idea that occurs to the assessor to govern the interpretation of everything that follows. |
| **Diagnostic overshadowing** | Attributing a new symptom to a diagnosis the student already has, so that the possibility of another problem is overlooked. |
| **Halo effect** | The influence of a single salient trait, positive or negative, on judgments across many unrelated areas. |

| | |
|---|---|
| **Horn effect** | The negative face of the halo effect: a salient negative trait that lowers judgments of areas unrelated to it. |
| **Internalizing problems** | Emotional problems the student directs inward, such as anxiety, withdrawal and sadness; they are easily overlooked because they do not disrupt the classroom. |

## Review questions

1. A school has 300 students who are non-native speakers of Arabic and 900 who are native speakers. Twenty-one students in the first group and 27 in the second have been diagnosed with a learning disability. Calculate the risk ratio for the non-native speakers and interpret it. Then suppose another school has 30 non-native speakers of Arabic, two of whom have been diagnosed: what should you be wary of before comparing its ratio with that of the first school? **(Objective 1)**
2. Why does a high risk ratio not, on its own, prove bias? What further questions does it raise? **(Objective 1)**
3. A gifted program has 150 students, nine of whom are learning Arabic as a second language; these learners make up 15% of students at that school level. Calculate the representation index and the fair minimum using Ford's formula, determine the shortfall if any, and propose two measures to close it, each addressing a different point of entry. **(Objective 3)**
4. For each of the following situations, identify the point at which bias enters and propose a measure to reduce it: (a) A teacher refers only students who disrupt the class. (b) A program requires a score of 130 on a single verbal test for admission. (c) A team interpreted the low scores of a recently arrived student as low ability without asking about his language. **(Objective 2)**

5. A study found that an item on a reasoning scale was harder, on average, for girls than for boys. Is this enough to conclude that the item is biased? What comparison do you need before reaching a judgment? **(Objective 2)**
6. After two years, a student speaks Arabic fluently but struggles with reading. What three pieces of evidence do you need before referral? And why is his fluency not sufficient evidence that he has mastered the language of instruction? **(Objective 4)**
7. The Culture-Language Interpretive Matrix showed a pattern of gradual decline for a referred student. One team member said: "The pattern proves the problem is linguistic, so there is no need for further assessment." Evaluate this view in light of studies that have examined the validity of the matrix, and identify two independent pieces of evidence the team needs. **(Objective 5)**
8. Choose one of the three cases in this chapter, or a case you have encountered in your own work, and name a cognitive bias that could have led to a wrong decision in it; then propose a strategy to reduce it. **(Objective 6)**
9. Design one procedure for your school that reduces the halo or horn effect in gifted nomination, and explain how you would know it had worked. **(Objectives 3 and 6)**
10. **Case for application:** Adam is eleven years old and in fifth grade. His first language is Somali, and he enrolled in a Saudi school four years ago. He was referred on suspicion of intellectual disability after obtaining a full-scale score of 68 (confidence interval 63–74) on an Arabic intelligence scale, with a verbal score of 62 and a nonverbal score of 84. His teacher says he speaks Arabic fluently but does not understand long instructions and struggles in every subject. His family says he helps out in the family shop, counts money and gives change quickly, and interprets for his parents at official appointments. His adaptive behavior has not been assessed, and his hearing has not been checked in three years. (a) At which of the four points might bias enter his case? (b) For eligibility under intellectual disability, the Special Education Organizational Guide requires an IQ of no more than 75 on a Wechsler scale or its equivalent, along with deficits in at least two

areas of adaptive behavior (Ministry of Education, 1436 AH [Arabic]). Is the available evidence sufficient to judge that both conditions are met? Explain. (c) What evidence is missing? Rank it by priority. (d) What is your decision now: diagnosis, no diagnosis, or deferring judgment? Justify your decision, and specify what is needed to finalize it. **(Objectives 2, 4 and 7)**

**Brief answers to the computational questions:** Question 1: The risk index for non-native speakers of Arabic is 21 ÷ 300 = 7%, and for native speakers 27 ÷ 900 = 3%, so the risk ratio is 7 ÷ 3 ≈ 2.33; that is, a non-native speaker is more than twice as likely as a native speaker to be diagnosed with a learning disability. In the second school the risk index is 2 ÷ 30 ≈ 6.7%, close to the first, but it rests on two cases: a change of a single case moves it between roughly 3.3% and 10%, so it can be compared only alongside the raw counts and several years of data. Question 3: Their share in the program is 9 ÷ 150 = 6%, the representation index is 6 ÷ 15 = 0.4, and the fair minimum is 15% × 0.8 = 12%, or 18 students, a shortfall of nine students.

**Table 14.5: Alignment of chapter objectives with review questions and learning activities.**

| Objective | Assessed by |
|---|---|
| **1. Calculate the risk index and risk ratio and interpret their limits** | Questions 1 and 2; Activity 1 |
| **2. Identify the point at which bias enters and propose ways to reduce it** | Questions 4 and 5; part (a) of Question 10; Activity 2 |
| **3. Calculate the representation index and the fair minimum and propose measures** | Questions 3 and 9; Activity 1 |
| **4. Distinguish conversational fluency from academic proficiency** | Question 6; part (c) of Question 10 |
| **5. Critique interpretations of the Culture-Language Interpretive Matrix** | Question 7 |

| Objective | Assessed by |
|---|---|
| **6. Name the cognitive bias and propose ways to reduce it** | Questions 8 and 9; Activity 3 |
| **7. Judge whether the evidence suffices for a fair decision** | Question 10 (case for application) |

## Learning activities

Any activity that uses students' data or work is carried out with the consent of the family and the school administration, without any information that could identify the student, and under the supervision of a qualified specialist. Its results are not used in any decision concerning the student and are reported only through official channels.

1. **Calculating from school data:** With the administration's permission, and using aggregate counts only, without names, obtain the numbers of referrals to special education, or of students enrolled in a gifted program, in one school or a group of schools, broken down by gender, grade or school type. Calculate the risk indices and ratios, or the representation index and the fair minimum; flag every cell with fewer than ten students; and write three questions the numbers raise. If no data are available to you, apply the steps to the figures in the opening scene and the two worked examples. *Product: a table of calculations and three questions on a single page. Time: about two hours.*
2. **Item analysis:** Choose ten items from an Arabized scale whose test booklet you have access to, without reproducing them in your report, and identify those that may carry cultural or dialectal content unfamiliar to a student in your region. Then search the scale's manual: does it report an analysis of differential item functioning or of measurement invariance across groups? *Product: a table of the items, numbered and classified, and a paragraph on the fairness evidence in the manual. Time: about an hour and a half.*

3. **Rating without knowing whose work it is:** Collect five pieces of student work, rewrite them in a uniform hand, and ask a colleague to rate them before and after their authors are revealed; then compare. If no student work is available to you, write two answers of equal content to a single task yourself, one in neat handwriting and the other in poor handwriting, and ask colleagues to rate them. *Product: a table showing both ratings for each piece of work, and a paragraph interpreting the differences. Time: about an hour.*

## Chapter references

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (1437 AH). *Organizational and procedural guide for the school program for gifted girls* [In Arabic]. General Directorate for Gifted Girls. https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20(1).pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Aldakhil, A. F., Ebrahim, M. T., & Gadelrab, H. F. (2023). Diagnostic survey of dyslexia and accompanying behavioral indicators in primary school students in Saudi Arabia. *Research in Developmental Disabilities, 134*, Article 104424. https://doi.org/10.1016/j.ridd.2023.104424

Alfaiz, F. S. (2026). The accuracy of parent and teacher nominations in gifted identification in Saudi Arabia. *Cogent Education, 13*(1), Article 2654922. https://doi.org/10.1080/2331186X.2026.2654922

Almulhim, N., Aljughaiman, A. M., & Alfaiz, F. S. (2026). Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis. *Journal of Intelligence, 14*(8), Article 166. https://doi.org/10.3390/jintelligence14080166

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678–13683.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775–780.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 2*(2), 132–149.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 2. Literacy* (2nd ed., pp. 71–83). Springer.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Wiley.

Ford, D. Y. (2013). *Recruiting and retaining culturally different students in gifted education*. Prufrock Press.

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open, 2*(1).

Kranzler, J. H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review, 39*(3), 431–446.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111–138.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher, 44*(5), 278–292.

Peters, S. J. (2022). The challenges of achieving equity within public school gifted and talented programs. *Gifted Child Quarterly, 66*(2), 82–94.

Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open, 5*(2).

Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. Guilford.

Skiba, R. J., Artiles, A. J., Kozleski, E. B., Losen, D. J., & Harry, B. (2016). Risks and consequences of oversimplifying educational inequities: A response to Morgan et al. (2015). *Educational Researcher, 45*(3), 221–225.

Styck, K. M., & Watkins, M. W. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition among referred students. *School Psychology Review, 42*(4), 367–382.

U.S. Department of Education. (2024). *Significant disproportionality*, 34 C.F.R. §§ 300.646–300.647. Retrieved September 29, 2026, from https://www.ecfr.gov/current/title-34/subtitle-B/chapter-III/part-300

Chapter 15

# 15 From diagnosis to decision

A diagnosis is worth what follows from it: a report the family can take in, a plan built on its findings, goals against which progress is measured, and periodic review that asks whether the judgment still holds and whether the services still fit. In this chapter we examine the links that connect a diagnostic judgment to an educational decision.

**Scene from a meeting room**

Mazen's father and mother sat facing a thirty-page report. His mother leafed through it quickly: tables, standard scores, percentile ranks, and terms such as “phonological processing” and “rapid automatized naming.” On the last page was a single line: “He meets the criteria for specific learning disorder in reading.”

His father looked up and asked two questions: “Does this mean something is wrong with our son?” And then: “So what do we do now?”

The report was accurate, yet it did not answer the two questions the family had come with. In this chapter we look closely at the distance between a correct finding and a useful decision, and at ways of bridging it.

## Chapter objectives

After studying this chapter, the reader should be able to:

1. Classify a set of decisions as a diagnostic judgment, an eligibility decision or a service decision, and justify each classification with an example.
2. Reorganize a test-by-test report around the referral question, and phrase its findings in language parents can understand.
3. Turn a general recommendation into a specific one that states what will be done, who will do it, how much and how often, and how success will be judged.
4. Apply the steps of a feedback meeting in a role-play with a family.
5. Build a chain linking an assessment finding to a need, a measurable goal, a service and a method of progress monitoring in the individualized education program (IEP).
6. Calculate a goal from the baseline and the rate of growth, and decide from a given progress graph whether the intervention should be changed, taking the number of data points into account.
7. Choose the appropriate type of decision for a gifted student (enrichment, single-subject acceleration, whole-grade acceleration or a domain-specific program), specify the evidence it requires, and write a measurable goal for the student.
8. Compare what Saudi regulations stipulate on eligibility, the plan, reevaluation and family rights with what US law stipulates, and identify what the published documents leave unaddressed.
9. Judge whether the evidence in a new case is sufficient for reevaluation or exit from services, and justify that judgment.

# Introduction

The team's task extends beyond the diagnostic judgment, because the value of a diagnosis lies in what follows from it. A report crowded with tables and terminology that ends with a single line about criteria, like Mazen's report in the opening scene, may be technically sound and still fail to answer the family's question: What does this mean for our son, and what will we do? Moving from diagnosis to decision requires distinguishing among the diagnostic judgment, the eligibility decision and the service decision; writing a report organized around the referral question; building a plan whose goals can be measured; and conducting periodic review that asks whether the judgment is still correct and whether the services are still appropriate.

We open the chapter with the eligibility decision and what is actually decided in it (Section 15.1), then turn to writing the report and its recommendations, with a sample report (Section 15.2); the meeting with the family (Section 15.3); and the move from the report to the individualized education program, including how to write measurable goals (Section 15.4). We then address progress monitoring, reevaluation and exit from services (Section 15.5); decisions for gifted learners, from identification to placement (Section 15.6); the regulatory framework in Saudi Arabia compared with US law (Section 15.7); and ethics and the student's voice (Section 15.8). We apply all of this to three cases (Section 15.9) and close with principles drawn from the book as a whole (Section 15.10).

## 15.1 The eligibility decision: what is actually being decided?

At the end of an assessment, three distinct decisions are made, and they are often confused:

- **The diagnostic judgment:** Does the student meet the criteria for a category, such as specific learning disorder, intellectual disability, autism spectrum

disorder or giftedness? (The chapters of Part Three discuss the criteria for these categories in detail.)

- **The eligibility decision (Eligibility):** Is the student entitled to special education services or to a gifted program under the regulations of the agency that provides them? It usually requires two conditions: the presence of the category, and the student's resulting need for specially designed instruction. In Saudi Arabia, this decision is made by the school's assessment and diagnosis team, which, according to the Special Education Organizational Guide, includes the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the student counselor or social worker, specialists relevant to the category, and the student's parent or guardian where possible (Ministry of Education, 1436 AH [Arabic]). Saudi regulations define a student with a disability as one whose impairment reaches "the extent that requires the provision of special education services" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), so need is built into the definition itself. The same logic applies in US law, namely the Individuals with Disabilities Education Act (IDEA), reauthorized in 2004 as the Individuals with Disabilities Education Improvement Act: the presence of a disability must be accompanied by the student's need, because of it, for special education (Code of Federal Regulations, 34 C.F.R. §300.8; Yell, 2019).
- **The service decision:** What exactly will the student receive, where, how much, and how will we know it is helping? This decision rests on the details of the student's *profile* more than on the name of the category.

A student may be diagnosed with autism spectrum disorder while succeeding in class and needing only minor accommodations. Another may meet the criteria for no category at all yet need Tier 2 support within response to intervention (RTI), a model that provides students with graduated support across tiers that grow more intensive as their response diminishes. The label does not determine the service: two students diagnosed with the same specific learning disorder may need entirely different plans.

**Key idea**

The question with which an assessment begins, "Does the student have X?", must end with another: "What does the student need, and how will we know that what we provide is helping?" An assessment that answers only the first has done only half its job.

## 15.2 Writing the report

The report is what remains of the assessment. It will be read by the parents, the teacher, the planning team, and perhaps another specialist years later. Most weak reports get the numbers right; they fall short in three ways: they are organized by test rather than by question, they are written in professional jargon, and they end with general recommendations that no one knows how to carry out (Harvey, 2006; Schneider et al., 2018).

### Organize the report around the question

A test-by-test report presents each instrument in its own section: the Wechsler results, then the achievement test results, then the behavior rating scale, leaving readers to pull the threads together themselves. A question-driven report, by contrast, begins with the referral question and gathers evidence from all sources under each sub-question, as Figure 15.1 illustrates.

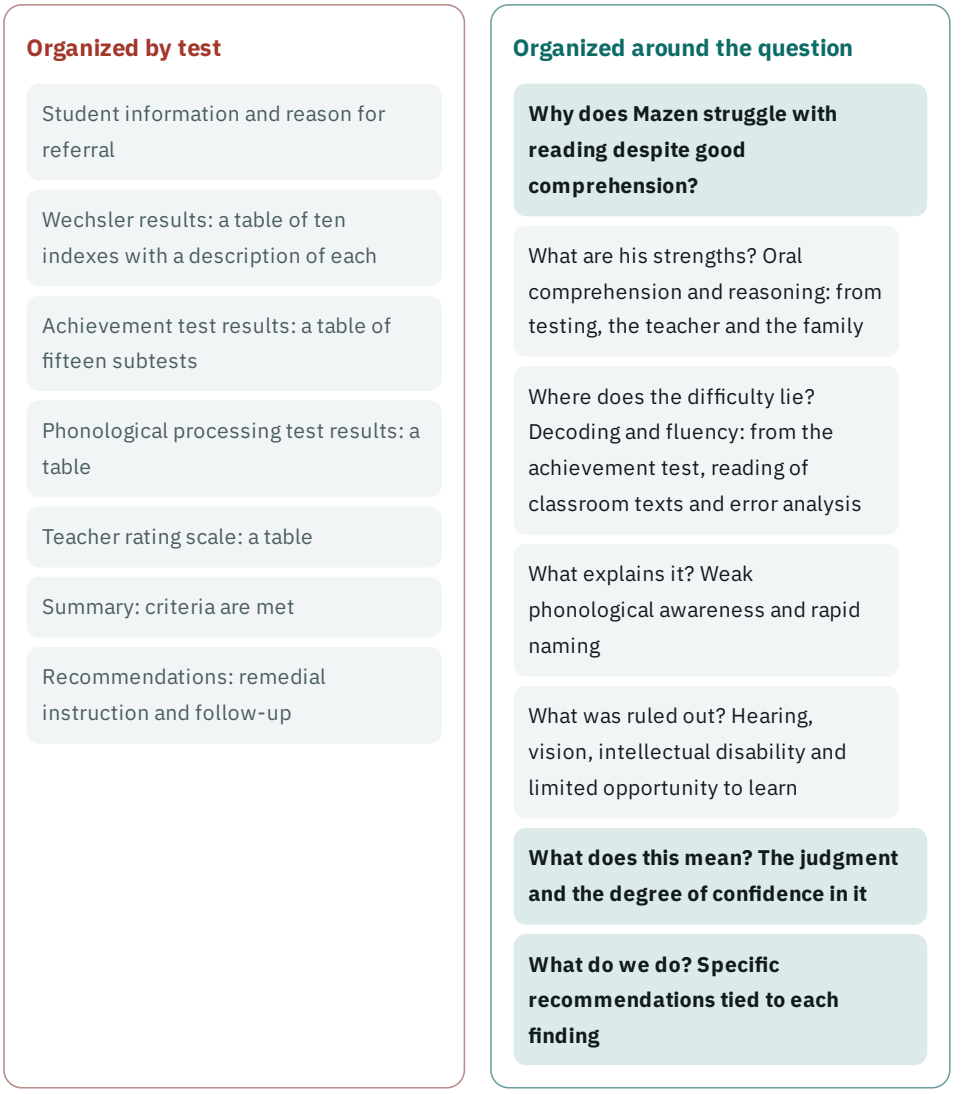



**Figure 15.1:** Two structures for the same report. On the left, the instruments are presented one after another; on the right, evidence from all sources is gathered under the questions the reader cares about. Full score tables move to an appendix at the end of the report.

## Principles of clear writing

- **Write for the least specialized reader:** If the parents understand the report, the specialist will too; the reverse does not hold.
- **Describe the behavior before naming the term:** "He has difficulty linking letters to their sounds, confusing letters that sound alike, such as sīn and ṣād, and dropping a syllable from long words" is clearer than "weakness in phonological processing" (that is, the processing of speech sounds). If you need the term, explain it the first time it appears.
- **Present numbers with their meaning:** Pair "his standard score is 76" with what it means: "he performs below roughly 95% of peers his age." This is what the percentile rank expresses: the percentage of the student's age peers in the norming sample who scored lower than he did. A percentile rank of 84 means he outperformed 84% of them, and a standard score of 76 corresponds to a percentile rank of about 5. Report the confidence interval as well, that is, the range within which the student's true score most likely falls, since every score carries some measurement error.
- **Begin with strengths:** Strengths are an integral part of the picture, and they are often the key to the plan.
- **Separate what was observed from what was inferred:** "I observed that he left the task four times in ten minutes" is an observation. "The task appeared to be above his level" is an inference. Both are useful, provided the reader can tell one from the other.
- **State the limits of the assessment:** If the norms were developed elsewhere, if the student was ill on the day of testing, or if the assessment was conducted in the student's second language, say so and note its likely effect.

**Table 15.1: Statements from diagnostic reports in professional language, rephrased in language parents and teachers can understand.**

| In professional language | In language everyone can understand |
|---|---|
| **He obtained a composite score of 72 on the Working Memory Index (percentile rank 3).** | He has clear difficulty holding information in mind while using it, such as remembering the steps of a problem as he solves it. His performance in this area is lower than that of most of his peers. |
| **He showed indicators of executive function weakness.** | He starts tasks without planning, forgets what he was asked to do when instructions have more than two steps, and needs reminders to check his work. |
| **Findings are consistent with autism spectrum disorder, Level 1.** | She meets the criteria for autism spectrum disorder. She needs support in understanding social situations and in coping with sudden change, while her language and achievement are on par with her peers. |

## Recommendations: from general to specific

A good recommendation answers four questions: What will be done? Who will do it? How much and how often? How will we know it has worked? It is tied to a specific finding in the report: no recommendation without a finding to support it, and no important finding without a recommendation that addresses it.

**Table 15.2: From general to specific recommendations, with the finding that supports each one.**

| General recommendation | Specific recommendation | Supporting finding |
|---|---|---|
| **He needs remedial reading instruction.** | Structured phonics instruction in a small group of no more than three students, 30 minutes a session, five days a week, delivered by the learning disabilities teacher, with weekly fluency measurement. | Weak decoding and phonological awareness |
| **Allowances should be made for him on tests.** | Test questions in science and social studies are read aloud to him, and he is given extended time on written tests. | Average oral comprehension, with slow reading that masks what he knows |
| **Communication with the family is advised.** | Every two weeks the teacher sends the family a graph of his fluency progress, along with one ten-minute home reading activity. | A cooperative family asking what they can do |

**Common misconceptions**

- **"A long report is a thorough report."** Length often comes from copying test descriptions out of their manuals. A useful report selects what answers the question and places the rest in an appendix.
- **"General recommendations are safer because they commit no one."** That is precisely their flaw: a recommendation that does not specify who does what will be carried out by no one.
- **"Stating the limits of the assessment weakens the report."** The opposite is true. A report that acknowledges its limits earns more trust and protects the student from a decision built on evidence weaker than it appears.

## Sample report: Mazen's report organized around the question

What follows is the report presented in the opening scene, rewritten on the right-hand structure in Figure 15.1. The wording is the author's, and the data are hypothetical, for teaching purposes. The report has been trimmed to what the parents and teacher need, and the scores have been moved to an appendix at the end. The numbered notes explain the function of each part.

Psychoeducational Assessment Report (teaching sample)

Name: (withheld) · Age: 8 years, 6 months · Grade 3

**Referral question.** Mazen's classroom teacher referred him with a question she framed together with his parents: Why does he struggle with reading even though he understands well what is read to him? And what will help him?[1] Before the referral, he had received twelve weeks of small-group support, in which he made slow progress.

**Sources of information.** Interviews with the parents and the teacher, observation in two lessons, samples from his notebooks, the record of previous support, an individual intelligence test, an individual achievement test, phonological awareness and rapid naming tasks, and reading of passages from his school textbook. His hearing and vision were screened before the assessment.

1. **The referral question** is framed with the family and the teacher, and it remains the organizing thread for everything that follows.
2. **Strengths first**, with a specific example from the student himself.
3. **The number with its meaning.** The score and its confidence interval go in the appendix; the meaning goes in the body.
4. **The judgment statement** names the category and the domain, ties them to the eligibility requirement in the applicable regulations, and then states the degree of confidence and the reason for it.
5. **The limits** are stated explicitly, along with their bearing on interpretation.
6. **Each recommendation** is tied to a finding and specifies what will be done, who will do it, how much and how often, and how success will be judged.

**What are his strengths?** Mazen understands what is read to him and explains it clearly.[2] When the teacher read the class a lesson about bees, he explained how the bees divide the work of the hive more precisely than many of his classmates. His performance in listening comprehension and in verbal and nonverbal reasoning fell in the average range. His parents describe him as curious, full of questions, and fond of being told stories.

**Where does the difficulty lie?** In reading words, and in reading speed. He reads 28 words correct per minute from a grade-level passage, whereas his schoolmates read about 70. His single-word reading is below that of roughly 95% of peers his age[3], and weaker still on pseudowords, a task in which guessing from context cannot help him. Most of his errors involve confusing letters that sound alike or look alike, and dropping a syllable from long words. His teacher noted that he avoids reading aloud and complains of tiredness after a single page.

**What explains it?** His performance was weak on phonological awareness tasks, such as deleting a sound from a word and blending sounds, and slow on rapid naming of letters. These two skills are among the strongest correlates of learning to decode.

The picture, then, is one of difficulty linking letters to their sounds, with intact comprehension.

**What was ruled out?** His hearing and vision are normal. His cognitive ability is in the average range, so intellectual disability does not explain his difficulty. His attendance has been regular since first grade, Arabic is the language of the home, and he has received classroom instruction followed by additional support, so the difficulty cannot be attributed to limited opportunity to learn. There were no indications of an emotional problem that would account for it, although his avoidance of reading merits monitoring.

**What does this mean?** Mazen meets the criteria for specific learning disorder with impairment in reading, affecting word reading accuracy and fluency. He also meets the requirements of the Special Education Organizational Guide for eligibility for the learning disabilities program: a clear discrepancy between his ability and his reading achievement, the exclusion of other causes, and the limited effectiveness of regular educational services alone.[4] Our confidence in this judgment is high, because the evidence from testing, classroom passages, the teacher, the family and the support record converges.

**Limits of this assessment.** No published local norms for reading fluency are available for his grade, so his fluency was compared with that of his schoolmates.[5] Sessions were held in the morning over two days, and he was cooperative in both.

**What do we do?**[6]

1. Structured, explicit phonics instruction in a group of no more than three students, 30 minutes daily, delivered by the learning disabilities teacher. *(Addresses weak phonological awareness and decoding.)*
2. Repeated reading of passages at his level for ten minutes daily, with immediate feedback. *(Addresses slow fluency.)*
3. Test questions read aloud to him in science and social studies, and extended time on written tests, with none of this applied to the reading test itself. *(Because his slow reading masks his good comprehension.)*
4. Weekly fluency measurement with a grade-level passage, reviewed every four weeks, with the graph sent to the family every two weeks along with a ten-minute home reading activity. *(So we know whether the services are sufficient.)*
5. Explaining his profile to Mazen in age-appropriate language: what he does well,

what he finds hard, and what will help him. *(Because of his avoidance of reading and the fatigue it causes him.)*

6. Rediagnosis at the time specified by the Organizational Guide, that is, six to eighteen months after this diagnosis, or earlier if needed.

**Table 15.3: Report appendix: Mazen's scores with confidence intervals (90%). Descriptive labels vary across tests; those used here are the Wechsler labels.**

| Area | Standard score | Confidence interval | Percentile rank | Description |
|---|---|---|---|---|
| **Verbal reasoning** | 106 | 99–112 | 66 | Average |
| **Nonverbal reasoning** | 102 | 94–109 | 55 | Average |
| **Listening comprehension** | 104 | 96–111 | 61 | Average |
| **Word reading** | 76 | 71–83 | 5 | Very low |
| **Pseudoword reading** | 73 | 68–81 | 4 | Very low |
| **Phonological awareness** | 78 | 72–86 | 7 | Very low |
| **—** | 79 | 72–88 | 8 | Very low |

| Area | Standard score | Confidence interval | Percentile rank | Description |
|---|---|---|---|---|
| **Fluency in grade-level text** | 28 words correct per minute (median of three passages); schoolmates about 70 | | | |

## 15.3 Meeting with the family

This may be the first time the family hears the word “disability” or “disorder” attached to their child's name. What is said in this meeting, and how it is said, stays with them for years. The task is to present the findings honestly, clearly and compassionately all at once, without softening them to the point that they are lost.

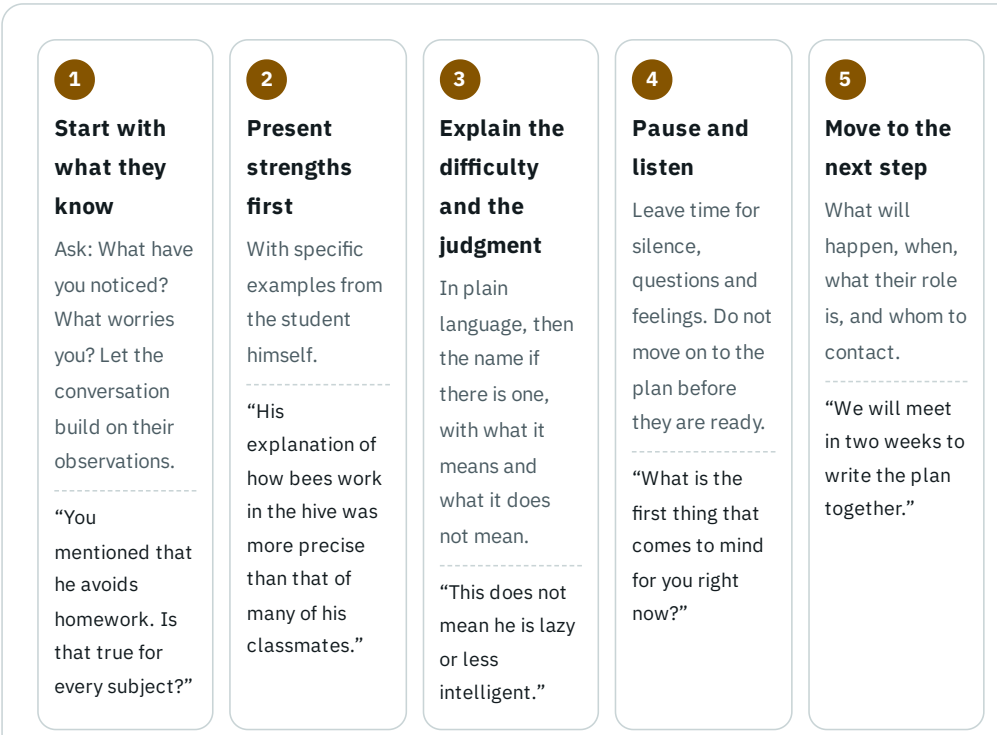



**Figure 15.2:** Five steps for a feedback meeting. Beneath each step is a sentence that might be said during it. The parents are given a copy of the report, ideally early enough before the meeting to read it, or a one-page summary in plain language.

**For the teacher: reactions to expect**

- **Denial:** "He's smart; he just doesn't want to study." Do not argue. Return to the shared evidence: samples of his work, his progress graph, and what the parents themselves have observed.
- **Guilt:** "We should have noticed long ago." Explain that many of these difficulties surface only with the demands of school, and that identifying them now is better than identifying them later.
- **Relief:** Some families feel relieved that the difficulty has a name and an explanation, after years of blaming the child or themselves. This, too, is a legitimate reaction.
- **Fear of stigma:** "Will this be written in his file?" Answer honestly about who will have access to the report and about their rights in this regard.

## 15.4 From the report to the individualized education program

The individualized education program (IEP) is the document that turns assessment findings into instruction. Its components are set by each country's regulations. In Saudi Arabia, the Regulations for Special Education Institutes and Programs specify eight components and make it one of the plan's foundations that "the family must participate in preparing, implementing, evaluating and following up the plan" (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Articles 83–85). Table 15.4 sets these alongside the components of the plan under US law (34 C.F.R. §300.320).

**Table 15.4: Components of the individualized education program in the Saudi Regulations and their counterparts in US law.**

| Component in the Regulations (1422 AH) | Counterpart in US law |
|---|---|
| **(a) Description of the present level of performance: strengths and areas of need** | Present levels of academic achievement and functional performance, and how the disability affects the student's involvement and progress in the general curriculum |
| **(b) Long-term and short-term goals, "provided that they are written in behavioral, measurable terms"** | Measurable annual goals. Short-term objectives or benchmarks are required only for students assessed on alternate assessments aligned to alternate achievement standards (§300.320(a)(2)(ii)) |
| **(c) Educational services and related services** | Special education, related services and accommodations, and any supports needed for school personnel |
| **(d) Start and end of services** | Projected start date of services, and their frequency, location and duration |
| **(e) Appropriate educational placement options** | A statement of the extent to which the student will not participate with peers in the general education class, with an explanation |
| **(f) Objective criteria for measuring performance, and procedures and periodic schedules for evaluating the plan (daily, weekly, monthly or annual)** | How progress toward annual goals will be measured, and when the family will be informed of it |
| **(g) Identification of those participating in implementing the plan** | Members of the IEP team are specified in a separate section and include the parents and, when appropriate, the student (§300.321) |

| Component in the Regulations (1422 AH) | Counterpart in US law |
|---|---|
| **(h) Identification of instructional materials** | No separate counterpart, although the team must consider the student's need for assistive technology |
| **No corresponding component among the eight** | The student's participation in statewide assessments and the accommodations provided in them, and transition services beyond school beginning no later than the first plan in effect when the student turns sixteen, or earlier if the team deems it appropriate (§300.320(b)) |

The comparison reveals a difference worth noting: the Saudi Regulations require short-term objectives for every student, whereas US law, since the 2004 amendments, has required only annual goals and limited short-term objectives to a narrow group. A reader who adopts an American IEP template may therefore omit a component required under their own system.

A good assessment feeds every one of these components. The following chain is what links the report to the plan. If any link breaks, the assessment becomes a document in a file, and the plan becomes a set of goals with no grounding.

| Link 1 | Link 2 | Link 3 | Link 4 | Link 5 |
|---|---|---|---|---|
| **Finding** | **Need** | **Goal** | **Service and accom-modation** | **Progress monitoring** |
| Mazen reads 28 words correct per minute from a grade-level passage, and most of his errors involve letters that look and sound alike. | He needs to master letter–sound correspondence and to have enough practice to build fluency. | By the end of the second semester, he will read 60 words correct per minute from a grade-level passage on three consecutive measurements. | Structured phonics instruction, 30 minutes daily, and test questions read aloud to him in science. | Weekly fluency measurement with a grade-level passage, graphed and reviewed every four weeks. |

**Figure 15.3:** The chain linking assessment to the plan, illustrated with Mazen's case from the opening scene. Each link rests on the one before it: the goal addresses a need, the need arises from a finding, and measurement tells us whether the service is achieving the goal.

## How is a measurable goal written?

A measurable goal specifies five things: **who**, **what** the student will do, expressed as observable behavior, **under what conditions**, **to what criterion**, and **by when** (Bateman & Linden, 2012). Compare:

**Non-measurable goal**

Mazen will improve his reading skills.

How will we know he has improved? By how much? When do we judge? At the end of the year, no one can say whether the goal was met.

**Measurable goal**

By the end of the second semester, Mazen will read aloud 60 words correct per minute from an unfamiliar grade-level passage on three consecutive weekly measurements.

## Where does the number 60 come from?

The criterion is not chosen at random. It is usually built on three sources: the **baseline** (Baseline), that is, the student's current performance, measured three times with the median taken; the **expected rate of growth** per week with good intervention; and **peer performance** or the local grade-level norm. In a well-known study of reading in English, the ambitious rate of growth in second grade was about two words correct per week (Fuchs et al., 1993). Such rates cannot be transferred to Arabic as they are, since the nature and length of Arabic words differ; local norms are therefore used where they exist, or data from students in the same school.

**Worked example: Mazen's goal**

1. Baseline: three measurements, 27, 28 and 30. Median: 28 words correct per minute.
2. Time remaining until the end of the second semester: 20 weeks.
3. An ambitious yet reasonable growth rate, estimated by the team from school data: about 1.6 words per week.
4. Goal: 28 + (1.6 × 20) = 60 words correct per minute.
5. The aimline is drawn from 28 at week zero to 60 at week twenty, and each measurement is compared against it, as in Figure 15.4.
6. Because the Saudi Regulations also require short-term objectives, the goal is divided into milestones that fall on the same aimline, for example reading 28 + (1.6 × 10) = 44 words correct per minute by the end of week ten.

## Accommodation and modification

An **accommodation** (Accommodation) changes how the student accesses or responds to a task without changing what is measured, such as reading the questions aloud to Mazen on a science test, since the aim is to measure his understanding of science independently of his reading. A **modification**

(Modification), by contrast, changes the content or the standard itself, as when a student is given a passage at a lower level. The distinction matters in the plan, because a modification changes the meaning of the score and is used only when the student genuinely needs it. Questions are not read aloud to Mazen on the reading test itself, because doing so would cancel out what is being measured.

## 15.5 Progress monitoring, reevaluation and exit from services

### Progress monitoring

A plan is a hypothesis: we hypothesize that this service will bring the student to this goal. Repeated measurement is what tests the hypothesis. Among its common tools is curriculum-based measurement (CBM), a brief, repeated measure of a basic skill using material from the curriculum itself, such as the number of words a student reads correctly in one minute (Deno, 2003; Chapter 4 discusses it in detail). One common decision rule is the consecutive data-point rule: if four consecutive points fall below the aimline, the intervention is changed; if four consecutive points fall above it, the goal is raised.

A systematic review of decision rules in reading fluency measurement, however, found that consecutive data-point rules are of low reliability when used alone, because a single measurement fluctuates with the difficulty of the passage and the day of testing. They are therefore supplemented by the **trend line** (**Trend line**) rule: the line of best fit is drawn through the student's data points, and its slope is compared with that of the aimline. The slope estimate cannot be trusted until enough measurements have accumulated, usually over six to ten weeks or more, depending on the quality of the passages and their equivalence in difficulty (Ardoin et al., 2013). Here the slope behaves like a score: **a slope is a range of plausible values**. A slope calculated from five

fluctuating measurements may in reality be much larger or much smaller than it appears, and judging it "insufficient" requires more data or other evidence.

In Saudi schools, a recent study described teachers applying practices close to the response to intervention model, but in a way that was "largely informal," lacking systematic progress monitoring, diagnostic tools and regulatory guidance (Almulla & Mahama, 2026). In another study, learning disabilities teachers cited lack of data, limited time and the absence of formal adoption of the model among the barriers to implementation (Altheiabi et al., 2025). Without repeated data, the plan comes to be judged by the teacher's impression at the end of the year.

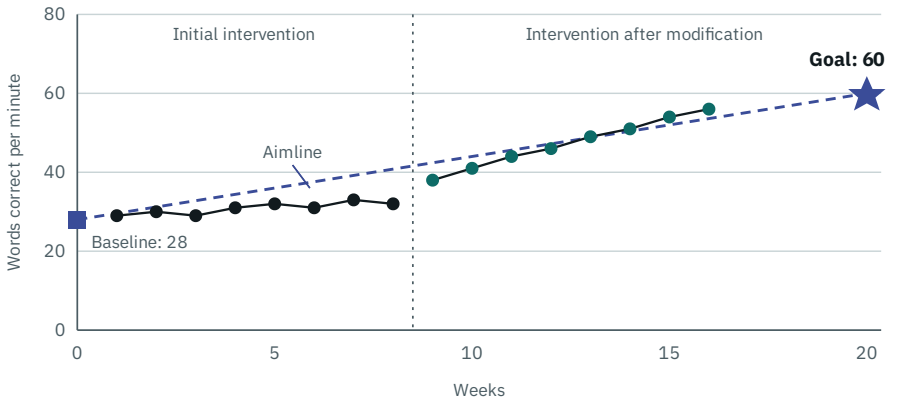



**Figure 15.4:** Mazen's progress graph with hypothetical values. The square represents the baseline, the dashed line the aimline (Aimline), and the star the goal. In the first eight weeks, all data points fell below the aimline, and their slope was clearly lower than that of the aimline, so the team modified the intervention: it increased the time and added repeated reading practice. Eight measurements are usually enough to judge the slope, whereas four alone are not. After the modification, the slope of the progress line grew steeper, and the points approached the aimline and then rose above it. The decision here was not "Mazen is not learning" but "this intervention is not enough for Mazen."

## Reevaluation

A diagnosis is a judgment tied to the moment it is made. Children grow, intervention has effects, and instruments improve. For this reason, most systems require reevaluation. In Saudi Arabia, the Special Education Organizational Guide states that “students with disabilities must be rediagnosed no less than 6 months and no more than 18 months after the initial diagnosis,” and it permits further reassessment and rediagnosis later “if the need arises” (Ministry of Education, 1436 AH [Arabic]). We found no fixed periodic schedule in it beyond the first rediagnosis. The Regulations add a right for the family: “rediagnosis of the pupil if the family considers that he was not placed in the appropriate program based on his previous diagnosis, should it request this” (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 76). Under US law, by contrast, reevaluation takes place at least once every three years unless the family and the school agree it is unnecessary, no more than once a year unless both agree, and also whenever the student's needs change or the parents or teacher request it (34 C.F.R. §300.303). The Saudi system thus brings the first review forward and places it close to the diagnosis, which suits judgments made about young children or on limited evidence, and it leaves what follows to the team's discretion.

Reevaluation does not necessarily mean repeating every test. The question it asks is different: Does the student still belong to the category? Does the student still need the services? Does the plan need to change? Answering these may require no more than reviewing progress data and teachers' observations, with targeted testing where necessary.

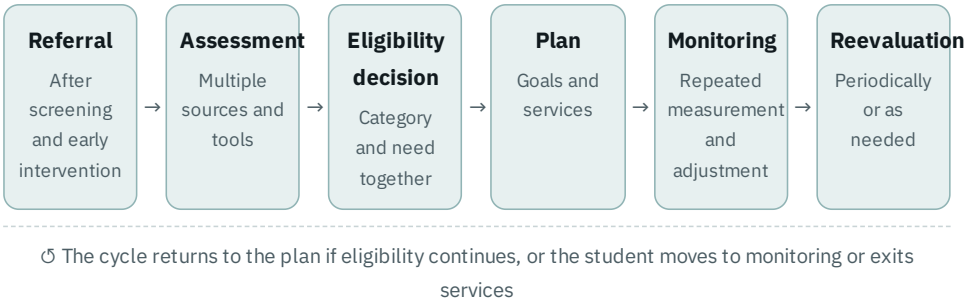



**Figure 15.5:** The cycle of diagnosis and services. Diagnosis is one link in a continuous cycle, and each link feeds the next.

## What might reevaluation reveal?

**Table 15.5: What reevaluation may reveal, and the usual decision in each case.**

| Result | Example | Usual decision |
|---|---|---|
| **Category and need continue** | A student with intellectual disability who still needs a modified curriculum | Eligibility continues, and the plan is updated |
| **The category has changed** | A child diagnosed with global developmental delay before age six, for whom later assessment showed a picture of autism spectrum disorder with average abilities | The category is changed, and the plan is rebuilt on the new profile |
| **The need has ended** | A student with a reading difficulty who has reached the level of her peers and maintains it with minor accommodations | She may exit services, with monitoring or accommodations in the general education class |

| Result | Example | Usual decision |
|---|---|---|
| **The initial judgment was wrong** | A student diagnosed with intellectual disability while learning Arabic, whose scores rose substantially as he mastered the language | The judgment is corrected, and the source of the error is reviewed to protect other students (Chapter 14 discusses the sources of this kind of error) |

### Exit from services

A student's exit from services is a diagnostic decision in its own right, and it requires assessment just as entry did. Under US law, a student may not be determined to no longer have a disability without an evaluation, except when the student graduates with a regular diploma or reaches the upper age limit (34 C.F.R. §300.305(e)). We found no provision in the Special Education Organizational Guide specific to exit procedures, so rediagnosis is the gateway through which this decision passes. A successful exit is planned: When will support be tapered? Who will monitor the student afterward? What signs would call for a return? Exit from gifted programs follows a similar logic, taken up in the next section.

## 15.6 Decisions for gifted learners: from identification to placement

Most of this chapter so far has been built on the disability model: a category, a need, and a plan that addresses a deficit. Yet the value of a judgment of giftedness, too, depends on the service decision that follows it, and the question is essentially the same: What does this student need, and how will we know that what we provide fits? The label “gifted” does not determine the service any more than a disability label does. Of two students classified as gifted, one may need acceleration in mathematics and the other a program in

creative writing. A growing strand of the literature therefore calls for asking whether the service matches the student's need in a particular domain before asking about the label (Peters et al., 2014; Worrell et al., 2019).

## Types of decisions and the evidence each requires

**Table 15.6: Types of decisions for gifted learners, the evidence each decision requires, and what to ask before making it.**

| Decision | What it means | Evidence required | What to ask before deciding |
|---|---|---|---|
| **Enrichment** | Broadening and deepening the curriculum without changing the student's grade or pace of progress, either within the classroom or in a separate program | Ability or interest that exceeds what the classroom offers: an ability measure, teacher ratings, and the student's products | Is the enrichment tied to the student's area of strength, or is it a general activity suitable for any student? |
| **Subject acceleration** | The student studies a single subject using a higher grade's curriculum or alongside a higher grade, while remaining with his own grade for all other subjects | Advanced achievement in that same subject: above-level testing showing that the ceiling is masking the student's ability, followed by a test on the higher grade's curriculum showing what he already knows of it and what he does not, so that only the unknown material is taught | Can his timetable be arranged? And who will teach him the following year so that he does not repeat what he has already mastered? |

| Decision | What it means | Evidence required | What to ask before deciding |
|---|---|---|---|
| **Whole-grade acceleration** | Skipping an entire grade, or early entrance | High general ability, advanced achievement in more than one subject, social and emotional maturity, and the student's own wish together with the support of the family and teachers. A structured tool such as the Iowa Acceleration Scale is used | How does the student himself see it? How will the new class receive him? And who will follow up with him after the move? |
| **Talent-domain program** | A program specialized in a particular domain, such as science, mathematics, scientific research, or the arts | Evidence from the domain itself (a test, a product, or a performance in it) added to the overall score | Does the admission criterion measure what the program actually offers? |
| **Exit from the program or change of service** | Ending a student's enrollment in a program, or moving him to another service | Performance data from the program covering at least one semester after the service has been modified, and an assessment that rules out an undetected learning disability or a temporary circumstance, with the reason documented | Does the problem lie in the student, or in how well the program fits him? And do the student and his family know when and how he can return? |

The acceleration decision is among the most thoroughly studied decisions in gifted education, and among the least used. The report **A Nation Empowered** summarized decades of research and concluded that acceleration is one of the

most effective options available for academically advanced students; that accelerated students generally achieve at least as well as their older peers; that the evidence shows, on average, no social or emotional harm from it; and that the greatest obstacle to it is the beliefs of educators, which run contrary to the research findings (Assouline et al., 2015). This does not mean that acceleration suits every gifted student. The **Iowa Acceleration Scale** (**Iowa Acceleration Scale**) was therefore developed to guide the team in whole-grade acceleration decisions specifically. The team uses it to gather data on ability, achievement, aptitude, development, relationships, and attitudes, and the scale treats certain items as critical, among them the student's own unwillingness to be accelerated (Assouline et al., 2009).

## The gifted student's plan and goals

A student with a disability receives a written individual plan, whereas a gifted student is often placed in a program with no plan at all. Some systems address this through an individual document called an **Advanced Learning Plan** (**Advanced Learning Plan**), as in several US states. The plan specifies the student's areas of strength, measurable goals in those areas, social and emotional goals, programs and placement, and the way progress will be monitored. The essential difference between a goal for a gifted student and a goal for a student with a disability is that the former measures *growth beyond grade level*, that is, learning what the student has not yet mastered. Compare:

**Non-measurable goal**

The student's skills in advanced mathematics will grow, and his passion for science will be satisfied.

What content? What criterion?

**Measurable goal**

By the end of the second semester, after studying them through mathematics acceleration, the student will pass the end-of-unit test for the sixth-grade fractions and ratio units with a score of 85% or higher. He will also complete a

And how will we know he has learned something new to him?

science research project in which he defines a question and a method for collecting and analyzing data, earning 3 out of 4 or higher on every criterion of a rubric applied by two independent raters.

## Continuation and exit

A judgment of giftedness, like a judgment of disability, holds at a particular point in time. A longitudinal study that followed elementary students found that many of those identified as gifted in a given year no longer met the same criterion a few years later, while others met it only later (Lohman & Korb, 2006). Two consequences follow. First, identification should be repeated periodically, so that the door is not closed to a student who was not identified in third grade. Second, exit from a program should not rest on a single score or one poor semester. The decline may stem from an undetected learning disability, a program that does not fit the student's area of strength, an environment that fails to challenge him and so erodes his motivation, or a temporary family circumstance. A fair exit decision therefore passes through four steps: looking for the cause of the decline; modifying the service with clear goals for a set period; reviewing the data with the student and his family; and, finally, documenting the decision together with publicly stated criteria for return.

Twice-exceptional learners are a special case. Such a student needs a single plan with two tracks, one that builds on his strength and one that supports his difficulty, with neither sacrificed for the sake of the other. The case of Rakan in Section 15.9 offers an example of such a dual plan.

## 15.7 The regulatory framework in Saudi Arabia compared with US law

So far this chapter has presented general principles of eligibility, planning, and reevaluation. The procedures binding on the school and the family, however, are set by each country's own regulations. In this section we present what we found in the published Saudi documents, and we place US law in a parallel column for comparison, since it is among the most detailed systems and among the most prominent in the literature (Yell, 2019). The presentation rests on four Saudi documents: the Regulations for Special Education Institutes and Programs, issued by the Ministry of Education [Al-Ma'arif] in 1422 AH; the Special Education Organizational Guide, issued by the Ministry of Education in its first edition for the 1436–1437 AH school year, which names the Regulations among its foundations; the Law on the Rights of Persons with Disabilities, issued by Royal Decree No. (M/27) dated 11/2/1445 AH, which replaced the Disability Welfare Law of 1421 AH (Article 32); and the Personal Data Protection Law. At the time of writing we found no published edition of the Organizational Guide more recent than the first, so readers should verify which edition is in force at the time they consult it.

**Table 15.7: Principles of eligibility, planning, and family rights in the Saudi regulations, and their counterparts in US law. Sources: the Regulations (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]), the Organizational Guide (Ministry of Education, 1436 AH [Arabic]), the Law on the Rights of Persons with Disabilities (1445 AH), and the Personal Data Protection Law (1443 AH); and 34 C.F.R. pt. 300.**

| Principle | In the Saudi regulations | In US law (34 C.F.R. pt. 300) |
|---|---|---|
| **Who determines eligibility** | The assessment and diagnosis team: the school principal, the vice principal for student affairs, the special education teacher, the general education teacher, the student counselor or social worker, specialists according to the category, the physician or nurse where available, and the parent or guardian where possible (Organizational Guide). In the Regulations: the relevant special education teacher, the behavioral training teacher (psychologist), the parent or guardian, and anyone whose participation the case requires (Article 81) | A team that includes the parents and qualified professionals (§300.306(a)(1)) |
| **Assessment requirements** | "Not relying solely on the results of a single measure," using "the tools and measures approved by the Ministry of Education," administered and interpreted by "qualified specialists" (Organizational Guide), with tests that are "appropriate to the child's environment" (Regulations, Article 80) | No single procedure is used as the sole criterion; tools are valid for the purposes for which they are used and are administered by trained personnel (§300.304) |

| Principle | In the Saudi regulations | In US law (34 C.F.R. pt. 300) |
|---|---|---|
| **Informed consent** | Written consent from the parent or guardian after formal notification, with a response due within a period "not exceeding (14) days from the date of the written notification"; if no response is received, "this is considered permission granting consent" (Organizational Guide; the Regulations say "within two weeks," Article 80). The parent's permission is sought "for the purpose of diagnosis, program preparation, and approval of any modifications" (Article 76) | Informed written consent before the initial evaluation, and separate consent before services begin. Silence does not count as consent, except for reevaluation when the school can show that it made reasonable efforts and the parents did not respond (§300.300) |
| **Participation** | The family is invited "to take part in diagnosis, program preparation, and evaluation" (Article 76); the parent or guardian is a member of the assessment team (Article 81); and "the family must participate in preparing, implementing, evaluating, and monitoring the plan" (Articles 83–85) | The right to attend eligibility and plan meetings and to participate meaningfully in the decision (§300.322) |
| **Access and explanation** | The parent or guardian is enabled "to review everything related to his child's program," and the pupil's condition is explained "in understandable language" (Article 76). We found no explicit provision for giving the family a copy of the evaluation report | Access to records, and a copy of the evaluation report at no cost (§300.306(a)(2), §300.613) |

| Principle | In the Saudi regulations | In US law (34 C.F.R. pt. 300) |
|---|---|---|
| **Confidentiality and data protection** | "Maintaining the confidentiality of all information" (Articles 76 and 80). The Personal Data Protection Law, issued by Royal Decree No. (M/19) dated 9/2/1443 AH and amended by Royal Decree No. (M/148) dated 5/9/1444 AH, has been in force since 14 September 2023; it is overseen by the Saudi Data and Artificial Intelligence Authority (SDAIA), and it classifies health data as sensitive data | The confidentiality rules in the law (§300.610 et seq.), together with the Family Educational Rights and Privacy Act (FERPA) |
| — | Eight elements, including long-term and short-term goals written in measurable behavioral terms (Articles 83–85; see Table 15.4) | Measurable annual goals, with short-term objectives only for students assessed through alternate assessments (§300.320) |
| **Reevaluation** | No sooner than 6 months and no later than 18 months after the initial diagnosis, and thereafter "as needed" (Organizational Guide). Also at the family's request "if it considers that [the child] has not been placed in the appropriate program" (Article 76) | At least once every three years, and no more than once a year unless the family and the school agree otherwise (§300.303) |

| Principle | In the Saudi regulations | In US law (34 C.F.R. pt. 300) |
|---|---|---|
| **Disputes and a second evaluation** | The right to request rediagnosis (Article 76). We found no published provision granting the family the right to an independent evaluation by an outside party at the expense of the education authority, nor any specific procedure for appealing an eligibility decision | Parents may request an independent educational evaluation **at public expense**; the agency must then either pay for it or request a due process hearing to show that its own evaluation is appropriate. The agency must also consider any independent evaluation the family submits (§300.502) |
| **System-level entitlement** | The Law on the Rights of Persons with Disabilities: the right to "supportive educational and training services at all stages," including the provision and adaptation of "modern assessment and diagnostic tools," and the design and implementation of early intervention programs (Article 8), as well as "screening and diagnostic services" and "early detection of disability" (Article 9) | The right to a free appropriate public education (§300.101), and the duty to locate, identify, and evaluate every child with a disability (§300.111) |

Table 15.7 shows that the Saudi regulations converge with US law on most principles: multiple sources, the team, family consent and participation, confidentiality, and a written plan with measurable goals. They differ in three respects that assessors would do well to know. First, a parent's silence for fourteen days counts as consent, which means that the notification itself must be clear, understandable, and documented, since a family may "consent" without having understood. Second, the first reevaluation comes early, after

six to eighteen months. Third, we found no published provision for the right to an independent evaluation or for an appeals procedure; if such a right exists, it is not spelled out in the documents we reviewed.

## Gifted programs

The Special Education Organizational Guide does not address gifted students; it lists eleven categories, all of them disability categories (Ministry of Education, 1436 AH [Arabic]). The Regulations had earlier defined "superior ability and giftedness" and set alternative admission conditions combining academic achievement, IQ, teacher testimony, and creative thinking (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic], Article 8). Current practice is governed instead by the gifted program guides and the documents of the national program, as summarized in Table 15.8. US law has no binding federal counterpart to this: giftedness is not one of the IDEA categories, and it is regulated by the states.

**Table 15.8: Stages of the decision for gifted students in the published Saudi documents.**

| Stage | What we found |
|---|---|
| **Nomination and identification** | The National Program for Gifted Identification, launched in 2011 as a partnership among Mawhiba, the Ministry of Education, and the National Center for Assessment. It is based on the Mawhiba Multiple Cognitive Abilities Test and targets grades 3 through 10 at three levels, and "the student is entitled to sit the test once at each level." Registration is initiated by the family, with a financial contribution from the parent except for subsidized students (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]) |

| Stage | What we found |
|---|---|
| **Admission to Mawhiba programs** | Available to those who qualify: Mawhiba classes, the Mawhiba advanced program in science and mathematics, research and enrichment programs, and nomination for university scholarships (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]) |
| **Admission to Ministry of Education gifted classes** | Students are nominated "from among the students classified in the National Project for Gifted Identification"; if the quota is not filled, places are completed according to criteria (a mental abilities test, personal characteristics, academic achievement, and creative products) by meeting two or more criteria, one of which must be the mental abilities test with a score "of no less than 120" (Ministry of Education, n.d. [Arabic]) |
| **School program for the care of gifted girls** | Admission for "the top 5% of the outcomes of the project," that is, the national project, from grade 4 through grade 11 (Ministry of Education, 1437 AH [Arabic]) |
| **Individual plan for the gifted student** | In the documents we reviewed, we found no provision requiring a written individual plan for every gifted student |
| **Continuation and exit** | Among the materials we reviewed, we found no published criteria for continuation in, or exit from, these programs |
| **Combining with special education services** | We found no provision allowing a student to combine a gifted program with a special education service, nor any provision prohibiting it |

In light of Section 15.6, the Saudi regulations appear detailed at the stage of identification and admission and only general beyond it. For a student who has entered a program, the published texts tell us neither how goals are to be written, nor when the student's status is to be reviewed, nor how exit takes

place. A recent review of gifted education in the Kingdom described extensive efforts in identification and programming (Alfaiz et al., 2022), and an exploratory study estimated that about 5.9% of a sample of 6,875 students in four private schools in Riyadh may be twice-exceptional, calling for policies specific to them (Almulhim et al., 2026). In all these areas the decision is left largely to the team's professional discretion, which makes clear principles, such as those set out in the previous section, all the more necessary.

**Question for reflection:** A student in a gifted program saw her grades drop over a single semester, and the program administration proposed removing her to free her seat for someone else. In the absence of published exit criteria, what steps would you suggest to the team, and what evidence would you ask for before the decision is made?

## 15.8 Ethics and the student's voice

The following principles are established in the professional codes for assessors, among them the standards of the National Association of School Psychologists (National Association of School Psychologists [NASP], 2020) and the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (American Psychological Association [APA], 2017). In essence they are general principles applicable to any system. In Saudi Arabia, the Organizational Guide requires that tools be administered and interpreted “by qualified specialists” (Ministry of Education, 1436 AH [Arabic]), and psychologists practicing in the health sector are subject to professional classification and licensure by the Saudi Commission for Health Specialties, including a licensing examination whose content covers psychological assessment and diagnosis (Saudi Commission for Health Specialties, 2024 [Arabic]). We found no published text specifying the levels of qualification required to administer each category of standardized tests.

⚠ **Ethics for the assessor (NASP, 2020; APA, 2017)**

- **Boundaries of competence:** Assessors do not administer a tool they have not been trained on, nor do they issue a judgment in an area outside their specialty. Referral to another specialist is a sign of competence.
- **Test security and the right of access:** Items from standardized tests are not copied, nor are they given to the family or the teacher for practice, because doing so destroys their validity for that student and for others. This is why no actual items appear in this book. Test security, however, does not justify withholding results from the family: the family has the right to know what was measured, what the result was, and what it means, even if it does not see the items themselves.
- **Understood consent:** Consent given through silence, or through a signature on a form that was never explained, is not informed consent in the professional sense. Whatever may suffice from a regulatory standpoint, the assessor's duty is to make sure the family understands what will happen and why.
- **No pressure on the outcome:** A judgment is neither made harsher nor softened so that a student will receive a service or be denied one, or because the school wants a certain number. The judgment follows the evidence.
- **Independence and conflicts of interest:** If some party asks for a particular result, the assessor's duty is to write what was found and to make clear the limits of what is known. Assessors also disclose any interest that might influence their judgment, such as assessing a student at a center they own and then recommending its services.

### The student's voice

It is often forgotten that a diagnosis belongs to someone. A student who understands his own profile (what he does well, what he finds hard, and what helps him) is better able to ask for what he needs and less likely to explain his difficulty by deciding that he is "stupid." Students are involved in a way that fits their age: a young child hears a simple explanation of his strengths and of what will help him, an older child attends part of the plan meeting, and an adolescent takes part in writing his own goals. Under US law, the student is invited to the plan meeting whenever postsecondary goals are discussed, beginning no later than the first plan to be in effect when the student turns sixteen, or earlier if the team sees fit (34 C.F.R. §300.321(b), §300.320(b)). Among the eight plan elements in the Saudi Regulations, we found no explicit provision for the student's own participation, although the element "identifying the participants" is broad enough to accommodate it. Adolescents moving from one school level to the next need this participation more than any other students, as the case of Tamim in the next section shows.

## 15.9 Three cases

All three cases are hypothetical, written for teaching purposes, and their figures are illustrative.

### Case 1: Mazen, 8 years 6 months

From report to plan

**The story:** Mazen is the student from the opening scene, and his full report was presented in Section 15.2: he understands what is read to him and explains his ideas well orally, but he has struggled with reading since first grade. After the first meeting, at which his parents did not understand the report, the specialist went over it with them again. She began by asking them what they had noticed, then read them what the teacher had written about his explanation of how bees work in the hive, and then explained the difficulty in a single sentence: "Mazen understands well, but in his mind the letters don't link to their sounds

quickly enough, so reading tires him out. This has a name, and there is teaching that helps with it." The team then wrote the following plan together with the family.

**Table 15.9: Mazen's individualized education program.**

| Plan element | What was written |
|---|---|
| **Present level** | Strengths: oral comprehension and reasoning in the average range, and clear intellectual curiosity. Needs: reads 28 words correct per minute in grade-level text, whereas his schoolmates read about 70. Avoids reading aloud and complains of fatigue. |
| **Long-term goal** | Will read 60 words correct per minute in grade-level text, on three consecutive measurements, by the end of the second semester. |
| **Short-term goals** | (1) Will read 44 words correct per minute in grade-level text by the end of week 10. (2) Will read two-syllable words containing letters that are similar in sound or in shape with 90% accuracy, on a list of twenty words, by the end of week 8. |
| **Service and setting** | Structured phonics instruction with the learning disabilities teacher in the resource room, 30 minutes daily; he remains in his regular class for the rest of the day. |
| **Accommodations** | Questions are read aloud to him in science and social studies. He is not asked to read aloud in front of the class without preparation. |
| **Progress monitoring** | Weekly fluency measurement, review of the graph every four weeks, and a graph sent to the family every two weeks. |
| **Participants** | The learning disabilities teacher, the classroom teacher, the parents, and Mazen himself, through a simple explanation of his plan. |

**What we learned:** The same report, with the same numbers, became useful once it was organized around the family's question and led to goals that could be judged. We saw in Figure 15.4 how progress monitoring led to a change in the intervention rather than to blaming Mazen.

> **What evidence could change the decision?** If, after the adjustment, the slope of his trend line remained clearly below the slope of the aimline across eight or more measurements, the team would review the intensity and type of service and reconsider the picture of the difficulty itself. If his oral comprehension proved to weaken as the language demands of later grades increased, the assessment would be broadened to oral language, as happened in Tamim's case.

## Case 2: Tamim, 14 years 9 months, grade 9

Reevaluation changes the picture

**The story:** Tamim was diagnosed in fourth grade with specific learning disorder in reading: he struggled with decoding, and his nonverbal ability was in the average range. Rediagnosis a year later confirmed his continued eligibility, and he received resource room services for five years; the decoding goals in his plan were met two years ago. In the second semester of ninth grade, before his move to secondary school, his parents requested a reevaluation: he was struggling in science and social studies and said, “I read the whole page, and I don't know what it said.” The special education teacher requested it as well, because he no longer knew what the service was offering Tamim now that he had mastered decoding. Rather than simply renewing eligibility, the team asked the question afresh and added to the assessment what had not been asked the first time: listening comprehension; oral language, through an assessment conducted by the speech-language pathologist; and developmental history.

**Table 15.10: Tamim's evidence at the initial diagnosis and at reevaluation.**

| Source | At initial diagnosis (grade 4) | At reevaluation (grade 9) |
|---|---|---|
| **Word reading (standard score)** | 74 | 93 |
| **Pseudoword reading** | 71 | 88 |
| **Fluency in grade-level text** | Below the 10th percentile among his schoolmates | At about the 30th percentile among his schoolmates |
| **Reading comprehension** | 78 | 79 |
| **Listening comprehension (text read aloud to him)** | Not measured | 77 |
| **Oral language (language sample and vocabulary and syntax tasks)** | Not assessed | Short sentences; errors in plural forms and attached pronouns; limited academic vocabulary; difficulty understanding complex sentences and long instructions |
| **Nonverbal reasoning** | 98 | 101 |
| **Developmental history (mother)** | Not asked about | Delayed speech; did not combine words into sentences until after age three |
| **Teachers** | “Struggles with reading” | “Understands when things are explained with examples, but gets lost in long questions even when they are read to him” |

| Source | At initial diagnosis (grade 4) | At reevaluation (grade 9) |
|---|---|---|
| **Tamim himself** | — | Knows he is "slower than others at understanding," and asks: Will he be able to choose what he wants in secondary school? |

**Analysis of the evidence:** The "reading difficulty" explanation was appropriate in fourth grade: decoding was weak, and it responded to intervention until it approached the average. His comprehension, however, remained low to roughly the same degree, and it then emerged that his understanding of what he hears is low as well. The remaining difficulty, now that turning letters into words has improved, lies in understanding language itself, whether read or heard. This is supported by his history of delayed speech and by what the speech-language pathologist found in his oral language. The better explanation now is a language disorder (developmental language disorder), of which the decoding difficulty was the most visible face in his early years; that face has been treated, but the underlying basis remains.

**Decision:**

- **Category:** The primary category is changed to "speech and language disorders," one of the categories listed in the Organizational Guide, with documentation of the history of reading difficulty and its response to intervention.
- **Plan:** Phonics instruction is discontinued, and speech-language services begin, with goals in academic vocabulary and in understanding complex sentences. The special education teacher and the subject teachers work with the speech-language pathologist to pre-teach new vocabulary before lessons and to break long instructions into parts.

- **Accommodations:** Extra time; long questions broken into parts without changing what they measure; and a glossary of science terms.
- **Transition:** Before the school year begins, a meeting is held between the intermediate and secondary school teams, and his plan moves with him. With his teacher, Tamim writes a "one-page profile" introducing himself to his new teachers: what he does well, what he finds hard, and what helps him. The data are reviewed after his first semester in secondary school, and rediagnosis takes place on the schedule the Organizational Guide sets following a change of category.
- **Tamim's voice:** He attended the last part of the meeting, chose to tell his new teachers himself, and discussed secondary school options with the school counselor in light of his strength in nonverbal reasoning.

**What we learned:** Reevaluation is a new question that goes beyond renewing eligibility: What is the student's picture now? A label that was right at nine may not be the best explanation at fourteen, because the school's demands shift from learning to read to reading to learn, and because the initial assessment never asked about oral language. Transitions between school levels are moments of risk: the new school knows nothing about the student unless his plan moves with him, and unless he himself can speak about who he is.

> **What evidence could change the decision?** If his listening and reading comprehension improved quickly after a few weeks of pre-teaching academic vocabulary, that would suggest his weakness reflects limited vocabulary and experience rather than a developmental language disorder, and the category would be reviewed. And if a recent hearing screening revealed an undetected hearing loss, the explanation of both the speech delay and the weak comprehension would change.

## Case 3: Rakan, 12 years, grade 7 One plan, two tracks

**The story:** Rakan is twice-exceptional; that is, he combines giftedness with a learning disability. He was diagnosed in elementary school with specific learning disorder in reading, then nominated for the gifted program after qualifying on the national test. He joined the program and received reading support along with testing accommodations. In seventh grade his teacher asked for his plan to be reviewed, because he had begun refusing accommodations in front of his classmates, saying, "I don't want anyone to know." The specialist found that his reading accuracy had improved considerably, but his speed was still low. Because we found no provision governing the combination of the two services (Section 15.7), the gifted program team and the school's learning disabilities team agreed on a single plan with two tracks, instead of two plans, each unaware of the other.

**A session with Rakan alone:** The specialist showed him his profile in a simple graph: reasoning at the top, speed at the bottom. He said, "So it means I'm not stupid, I'm just slow at reading?" He then chose for himself the accommodations that suited him and helped draft his goals.

**Table 15.11: Rakan's dual plan: the strength track and the support track.**

| Plan element | Strength track (gifted program) | Support and accommodations track (learning disabilities) |
|---|---|---|
| **Present level** | Very high verbal and nonverbal reasoning (General Ability Index 126), original ideas in science projects, and fluent oral explanation | Reading accuracy of 96% in grade-level texts, speed roughly half that of his peers, and long completion times on extended written assignments |

| Plan element | Strength track (gifted program) | Support and accommodations track (learning disabilities) |
|---|---|---|
| **Objective** | By the end of the year, he will complete a research project in physics and present it orally with visuals, earning 3 out of 4 or higher on every criterion of the program rubric, as scored by two raters | By the end of the second semester, he will use text-to-speech and audiobooks independently on four out of every five extended reading tasks, and will request extra time himself on every test |
| **Service** | The gifted program, with enrichment in science | A weekly session with the learning disabilities teacher to practice reading strategies and assistive technology |
| **Accommodations** | Completing some tasks orally or through a visual presentation in place of long written reports, without lowering the level of the content | Extra time on tests without announcing it in class, and digital texts with text-to-speech |
| **Progress monitoring** | Project rubric at mid-semester and at the end of the semester | A monthly log of technology use, and a sample of assignment completion times |
| **Responsible** | Gifted program teacher | Learning disabilities teacher, in coordination with the program teacher |

**What we learned:** A twice-exceptional student needs a single plan that reads his strength and his difficulty together, so that he is neither removed from the gifted program because of his difficulty nor left without support because of his giftedness. An accommodation the

student rejects does not work. Once Rakan understood his profile, he began asking for what he needed on his own, a skill he will carry with him beyond school.

> **What evidence could change the decision?** If he continued to refuse accommodations even after they were no longer visible to others, and his grades fell in reading-heavy subjects, the team would review the accommodations with Rakan himself, looking for alternatives he would accept. And if the samples of assignment completion times showed that his reading speed had approached that of his peers, the support track would be gradually scaled back while monitoring continued.

## 15.10 Principles to take with you from this book

Across this book we have covered many categories, a variety of tools, and numerous cases. Even if the names of instruments and the numbers in the criteria fade from memory, these are the principles we hope will endure:

1 **No single measure settles the question**

Judgment rests on converging evidence from multiple sources, tools, and settings.

2 **A score is a range, not a point**

Every score carries measurement error, and a cutoff is a wide zone.

3 **Rule out before you rule in**

Check hearing, vision, language, and opportunities to learn before attributing a difficulty to the student.

4 **Functional impact is a requirement**

A low score is not enough without a real effect on the student's life and learning.

5 **Differential diagnosis is the heart of the work**

Alongside the question “Are the criteria met?” comes the question “What is the best explanation?”

6 **Giftedness and difficulty can coexist**

Look at the student in comparison with peers and in comparison with himself.

7 **A tool is a product of its culture**

Norms, items, and language carry assumptions, and fairness needs someone to ask about them.

8 **The family is a source and a partner**

Its knowledge of the child is evidence, and its decision is part of the decision.

9 **Diagnosis serves instruction**

Its value lies in the measurable goals and services that change after it.

10 **Every judgment is provisional**

The child grows, the evidence is renewed, and the judgment is revisited.

## Conclusion

The central argument we have sought to develop in this chapter is that a diagnosis is measured by what follows from it, and that it is the student's profile, more than the label, that determines the service. A connected chain follows from this: a report organized around the referral question, describing behavior before terminology and presenting numbers with their meaning; recommendations that specify what is to be done, by whom, to what extent, and how success will be recognized; and an individualized education program that moves from finding to need to goal to service to measurement. The plan is a hypothesis tested by repeated measurement: if the student does not progress, the intervention is changed and the student is not blamed. Reevaluation, in turn, is a new question about the student's picture now, one that goes beyond

renewing eligibility. The same logic applies to gifted students: each placement decision has its own evidence, and no gifted student should be removed from a program without a search for the cause of the decline.

Within the Saudi regulatory framework, however, there remain issues for which we found no published text, such as the right to an independent evaluation and criteria for continuation in and exit from gifted programs; in these matters judgment rests on professional discretion until they are settled by regulation. The student, for his part, retains a right to understand his own profile and to take part in the decision about him, a right that grows in importance at transitions between school levels and a reminder that the purpose of all assessment is a decision that serves the student at its center.

## 15.11 Summary, key terms, and activities

### Chapter summary

- The diagnostic judgment, the eligibility decision, and the service decision are three different decisions. It is the profile, more than the label, that determines the service. In Saudi Arabia, the eligibility decision is made by the school's assessment and diagnosis team.
- A useful report is organized around the referral question, begins with strengths, describes behavior before terminology, presents numbers with their meaning, states its own limitations, and places tables in an appendix.
- A good recommendation specifies what is to be done, by whom, to what extent, and how success will be recognized, and it is tied to a finding in the report.
- The feedback meeting with the family begins with what the family knows, presents strengths, explains the judgment clearly, and leaves time for listening before moving on to the plan.

- A chain running from finding to need to goal to service to measurement links assessment to the individualized education program. The Saudi Regulations specify eight elements for the plan, including long-term and short-term goals.
- The plan is a hypothesis tested by repeated measurement. Consecutive data-point rules are complemented by the trend line, and a slope is a range that requires enough data points. Lack of progress calls for changing the intervention rather than blaming the student.
- Reevaluation is a new question about the student's picture now, going beyond renewal of eligibility. The Organizational Guide sets the first reevaluation at six to eighteen months after diagnosis. Exit from services is a decision that requires assessment and planning.
- Decisions for gifted students take several forms (enrichment, subject acceleration, whole-grade acceleration, and talent-domain programs), each with its own evidence. Gifted students have goals that measure growth beyond their grade level, and they are not removed from a program without a search for the cause of their decline. A twice-exceptional student has a single plan with two tracks.
- Under the Saudi regulations, the family has rights to consent, participation, access, confidentiality, and requesting rediagnosis, and its silence for fourteen days counts as consent. We found no published provision for the right to an independent evaluation, nor for criteria for continuation in and exit from gifted programs.
- Students have a right to understand their own files and to take part in the decisions made about them, and this right matters all the more at transitions between school levels.

## Glossary of terms

| | |
|---|---|
| **Eligibility determination** | The decision on whether a student is entitled to services; it usually requires both a qualifying category and a need for specially designed instruction. |
| **Assessment and diagnosis team** | The school team that determines eligibility under the Saudi regulations; it includes members from school administration, special education, general education and counseling, along with relevant specialists and the parent or guardian. |
| **Individualized Education Program (IEP)** | A document that sets out the student's present level, goals, services and accommodations, and how progress will be measured. |
| **Present levels of performance** | A description of what the student can do now, with both strengths and areas of need; it is the starting point for writing goals. |
| **Measurable goal** | A goal that specifies the behavior, the conditions, the criterion and the time frame, so that one can judge whether it has been met. |
| **Baseline** | The level of performance before intervention, usually measured three times, with the median taken. |
| **Aimline** | A line on the progress graph connecting the baseline to the goal, against which data points are compared. |

| | |
|---|---|
| **Trend line** | The line of best fit through a student's data points on the progress graph; its slope is compared with the slope of the aimline. |
| **Decision rule** | A rule agreed upon in advance for changing the intervention or the goal, such as four consecutive data points below the aimline, or a trend-line slope lower than the aimline slope. |
| **Reevaluation** | A periodic or as-needed evaluation that asks whether the category still applies, whether the student still needs services, and what the student's current profile looks like. |
| **Accommodation** | A change in how a task is presented or responded to that does not change what is being measured, such as reading science test questions aloud to the student. |
| **Modification** | A change in the content or the standard itself, such as providing a lower-level text; it changes the meaning of the score. |
| **Enrichment** | Broadening and deepening the curriculum for a gifted student without changing the student's grade placement or pace of progress through it. |
| **Acceleration** | Moving a student through the curriculum at a pace matched to the student's level, whether in a single subject, by skipping a whole grade, or through early admission. |

| | |
|---|---|
| **Advanced Learning Plan (ALP)** | An individual document for a gifted student that identifies areas of strength, measurable goals in those areas, social and emotional goals, and how progress will be monitored. |
| **Independent educational evaluation** | An evaluation conducted by a party outside the school. Under US law, families may request one at public expense if they disagree with the school's evaluation. |

## Review questions

1. Classify each of the following as a diagnostic judgment, an eligibility decision or a service decision, and justify your answer: (a) "She meets the criteria for autism spectrum disorder." (b) "Science questions will be read aloud to him." (c) "He is admitted to the learning disabilities program." Then explain: can a student be diagnosed and yet not be eligible for services?
2. Rephrase the following statement in language that parents can understand: "He obtained a standard score of 68 on the Processing Speed Index, in the extremely low range."
3. Turn the recommendation "needs support with attention" into a specific recommendation, and state the finding that could support it.
4. A student's baseline is 40 words correct per minute, 15 weeks remain, and a reasonable growth rate is 1.5 words per week. What is the goal? What is her short-term goal at week 8? After four measurements, all of which fell below the aimline, should the intervention be changed now? Justify your answer in light of what you have learned about the number of data points and the trend line.
5. What is the difference between an accommodation and a modification? And why are the questions not read aloud to Mazen on the reading test?

6. Compare the Saudi regulations with US law on three points: the effect of a parent's silence on a request for consent, the timing of reevaluation, and the right to an independent evaluation. What does the first difference imply for how the notice should be prepared?
7. Why does a decision to exit a student from services require an evaluation? And what steps should precede removing a student from a gifted program?
8. In Tamim's case: what question did the first evaluation fail to ask, and why did answering it change both the category and the plan?
9. Choose one of the cases in this chapter (Mazen, Tamim or Rakan) and write a paragraph explaining the student's profile to the student in age-appropriate language.

## Case for application

### Aseel, age 10, fifth grade

**The story:** Aseel qualified through the National Program for Gifted Identification and attends a general enrichment activity at her school once a week. Her mathematics teacher says that she "finishes the class exercises in a quarter of the time and then distracts her classmates," and Aseel says she "knows the lesson before it is taught." Her parents have asked that she be moved straight to seventh grade, but Aseel told the counselor that she does not want to leave her friends. With the family's consent, the teacher gathered the following data.

**Table 15.12: Aseel's data.**

| Source | Result |
|---|---|
| **Above-level mathematics test** | Performance close to the average for eighth-grade students |

| Source | Result |
|---|---|
| **End-of-year sixth-grade mathematics test** | 92%, with all of her errors in the geometry unit |
| **Arabic language and science** | High scores within the range of her grade level |
| **Ability test** | Quantitative reasoning clearly higher than verbal reasoning |
| **Teachers and counselor** | She gets along well with her classmates, and becomes frustrated when asked to redo what she has already mastered |

**Tasks:** (1) Which of the decision types in Table 15.6 would you propose for Aseel, and why not the others? (2) What evidence is still missing before you can be confident in your decision? (3) Write one measurable goal for her. (4) How would you explain your decision to her parents, who asked for something else, and to Aseel herself?

## Learning activities

Any activity that uses data from a real student is carried out with the consent of the family and the school administration, after removing everything that could identify the student, and under the supervision of a specialist. The products of these activities are for learning purposes only, and no decision is to be based on them.

1. **Rewriting a report (about two hours):** Take a report organized test by test, either a real report used under the conditions above or one you write yourself for Tamim from the data in Table 15.10, and reorganize it around the referral question, following the model report in Section 15.2. *Product:* a two-page report and a one-page summary for the family.

2. **Role play (45 minutes):** In groups of three, one of you plays the specialist and the other two play the parents. The specialist presents the findings of one of the cases in this chapter, following the steps in Figure 15.2. The parents choose one of the expected reactions without telling the specialist. Then switch roles. *Product:* a list of what worked in each meeting and what did not.
3. **Building a chain (one hour):** For Tamim's case after reevaluation, or for a student you know (under the conditions above), build the five-link chain in Figure 15.3 for two different goals, giving each long-term goal a short-term objective. *Product:* a two-column table.
4. **The regulatory framework in your school (two hours):** Consult the current edition of the Special Education Organizational Guide and the guides for gifted programs, and check every cell in Tables 15.7 and 15.8: has the text changed? Then compare the texts with practice in a school you know: who actually attends the planning meeting? Do parents receive a copy of the report? How is the consent notice worded? How does a student exit services, or leave the gifted program? *Product:* a one-page table showing where practice matches the text and where it falls short.
5. **A two-track plan (one hour):** Write a plan, using the template in Table 15.11, for Aseel (the case for application) or for a twice-exceptional student from one of the book's chapters. *Product:* a plan in a single table.
6. **Reviewing the book:** Return to the ten principles in Section 15.10 and choose an example to illustrate each one, drawn from the cases in this chapter or elsewhere in the book. *Product:* ten short paragraphs.

## Chapter references

King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/

Law on the Rights of Persons with Disabilities, Royal Decree No. (M/27) (11/2/1445 AH) [In Arabic]. https://ymplatform.com/sdys/

Personal Data Protection Law, Royal Decree No. (M/19) (9/2/1443 AH), as amended by Royal Decree No. (M/148) (5/9/1444 AH) [In Arabic].

Saudi Commission for Health Specialties. (2024). *Saudi Psychology Licensure Examination (SPSLE): Applicant guide*. https://scfhs.org.sa/sites/default/files/2024-03/SPSLE%20Applicant%20Guide%202024.pdf

Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf

Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf

Ministry of Education. (1437 AH). *Organizational and procedural guide for the school program for gifted girls* [In Arabic]. General Directorate for Gifted Girls. https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20(1).pdf

Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf

Alfaiz, F. S., Alfaid, A. A., & Aljughaiman, A. M. (2022). Current status of gifted education in Saudi Arabia. *Cogent Education, 9*(1), Article 2064585. https://doi.org/10.1080/2331186X.2022.2064585

Almulhim, N., Aljughaiman, A., Alnaim, F., Alsayed, A., Alfawzan, S., & Alarfaj, A. (2026). Hidden potential: An exploratory study of the prevalence of twice-exceptional students in Saudi Arabia. *Journal of Intelligence, 14*(5), Article 83. https://doi.org/10.3390/jintelligence14050083

Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications, 13*(1), Article 974. https://doi.org/10.1057/s41599-026-07376-9

Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties, 30*(1), 75–92. https://doi.org/10.1080/19404158.2025.2473425

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18. https://doi.org/10.1016/j.jsp.2012.09.004

Assistance to States for the Education of Children With Disabilities, 34 C.F.R. pt. 300 (2006) (§300.8, §300.101, §300.111, §300.300, §300.303, §300.304, §300.305, §300.306, §§300.320–300.322, §300.502, §§300.610–300.613).

Assouline, S. G., Colangelo, N., Lupkowski-Shoplik, A., Lipscomb, J., & Forstadt, L. (2009). *Iowa Acceleration Scale manual* (3rd ed.). Great Potential Press.

Assouline, S. G., Colangelo, N., VanTassel-Baska, J., & Lupkowski-Shoplik, A. (Eds.). (2015). *A nation empowered: Evidence trumps the excuses holding back America's brightest students*. Belin-Blank Center.

Bateman, B. D., & Linden, M. A. (2012). *Better IEPs: How to develop legally correct and educationally useful programs* (5th ed.). Attainment Company.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184–192.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 27–48.

Harvey, V. S. (2006). Variables affecting the clarity of psychological reports. *Journal of Clinical Psychology, 62*(1), 5–18.

Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted, 29*(4), 451–484.

National Association of School Psychologists. (2020). *The professional standards of the National Association of School Psychologists*.

Peters, S. J., Matthews, M. S., McBee, M. T., & McCoach, D. B. (2014). *Beyond gifted education: Designing and implementing advanced academic programs*. Prufrock Press.

Schneider, W. J., Lichtenberger, E. O., Mather, N., & Kaufman, N. L. (2018). *Essentials of assessment report writing* (2nd ed.). Wiley.

Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixson, D. D. (2019). Gifted students. *Annual Review of Psychology, 70*, 551–576.

Yell, M. L. (2019). *The law and special education* (5th ed.). Pearson.

End of the book

# Appendices and lists

Quick references

Appendix A: Guide to instruments

Appendix B: Glossary

List of abbreviations

List of figures and tables

Index

Appendix A

# A Guide to instruments

A concise reference guide to the instruments presented in this book (54 tools and procedures), grouped by purpose, with a pointer to where each is discussed in the chapters.

**Before using this guide**

This appendix brings together the instruments presented in the book's instrument profiles and in its tables of instruments and their Arabic versions, arranged in eight tables by purpose. For each instrument it records only what the book itself reports: purpose, age range, examiner qualification level and the status of its Arabic version, followed by the chapter in which it is presented. Age ranges, qualification levels and Arabic versions are all given as they stood when the book was prepared; they change as new editions, new Arabic standardizations and new publisher requirements appear, so the publisher and the instrument's own manual should be consulted before any professional use. This guide is no substitute for the full instrument profile in the relevant chapter.

Purpose:

**Screening**

**Rating**

**Diagnosis**

**Identification**

**Monitoring**

. A dash (—) means the book does not give this information for the instrument. An asterisk (*) marks an instrument mentioned in the body of a chapter rather than in a profile or table. The chapter number points to the profile or table in which the instrument first appears in that chapter.

# A.1 Intelligence and abilities

**Table A.1: Individual intelligence and ability tests and nonverbal ability tests.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Wechsler Intelligence Scale for Children, Fifth Edition** WISC-V | Diagnosis | 6 to 16 years 11 months | — | We know of no published Saudi standardization of the fourth or fifth edition. The older revised edition (**WISC-R**) was standardized in Saudi Arabia for ages 6–16 (Al-Nafi et al., 2000 [Arabic], as cited in Ejbara, 2024) | **Ch. 3 Ch** |
| **Wechsler Preschool and Primary Scale of Intelligence, Third Edition** WPPSI-III | Diagnosis | The book does not give the range of the original; the Saudi version was administered from 4 years to 7 years 3 months | — | A Saudi version derived from a Jordanian version, administered to 788 children in the Jeddah education region between 2018 and 2022 (Ejbara, 2024); its generalizability to other regions of the Kingdom has not been established | **Ch. 3 Ch** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Stanford-Binet Intelligence Scales, Fifth Edition** SB5 | Diagnosis | 2 to 85+ years | Publishers restrict it to those with advanced training in administering and interpreting individual intelligence tests; check the publisher's current designation for the level | Two Gulf versions are in use in the Kingdom, one prepared by Mahmoud Abu Al-Nil and colleagues and the other by Safwat Farag; a study in the Northern Borders region compared them (Mohammed, 2016 [Arabic]). We found no published data on their norming samples or the year the data were collected | **Ch. 3 Ch** |
| **Kaufman Assessment Battery for Children, Second Edition, Normative Update** KABC-II NU | Diagnosis | 3 to 18 years | — | No published standardized Arabic version is known | **Ch. 3** |
| **Differential Ability Scales, Second Edition** DAS-II | Diagnosis | 2 years 6 months to 17 years 11 months | — | No published standardized Arabic version is known | **Ch. 3** |
| **Cognitive Assessment System, Second Edition** CAS2 | Diagnosis | 5 to 18 years 11 months | — | No published standardized Arabic version is known | **Ch. 3** |
| **Leiter International Performance Scale, Third Edition** Leiter-3 | Diagnosis | 3 to 75+ years | — | No published standardized Arabic version is known | **Ch. 3 Ch** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Universal Nonverbal Intelligence Test, Second Edition** UNIT2 | Diagnosis | 5 to 21 years 11 months | — | No published standardized Arabic version is known | **Ch. 3 Ch** |
| **Test of Nonverbal Intelligence, Fourth Edition** TONI-4 | Screening | 6 to 89 | — | No published standardized Arabic version is known | **Ch. 3** |
| **Raven's Progressive Matrices** Raven's | Screening | A wide range, depending on the form | — | No published Arabic standardization of its current forms has been verified | **Ch. 3** |

## A.2 Achievement and academic skills

**Table A.2: Comprehensive achievement batteries and tests of specific skills in reading, mathematics and writing.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Wechsler Individual Achievement Test, Fourth Edition** WIAT-4 | Diagnosis | 4 to 50 years 11 months | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements | No published standardized Arabic version is known | **Ch. 3 Ch** |
| **Kaufman Test of Educational Achievement, Third Edition** KTEA-3 | Diagnosis | 4 to 25 years 11 months | — | No published standardized Arabic version is known | **Ch. 3** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Woodcock-Johnson IV Tests of Achievement**<br>WJ IV ACH | Diagnosis | 2 to 90+ years | — | No published standardized Arabic version is known. The publisher refers to a fifth edition (**WJ V**), so check for the most recent edition | **Ch. 3** |
| **KeyMath-3 Diagnostic Assessment**<br>KeyMath-3 | Diagnosis | 4 years 6 months to about 21 years | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements | No published standardized Arabic version is known | **Ch. 8** |
| **Test of Early Mathematics Ability, Third Edition**<br>TEMA-3 | Diagnosis Screening | 3 years to 8 years 11 months | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements | No published standardized Arabic version is known, nor any standardized Arabic instrument for number sense | **Ch. 8** |
| **Test of Written Language, Fourth Edition**<br>TOWL-4 | Diagnosis | 9 years to 17 years 11 months | An examiner trained in administering standardized tests and scoring writing samples, in accordance with the publisher's requirements | No published standardized Arabic version is known | **Ch. 8** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Comprehensive Test of Phonological Processing, Second Edition** CTOPP-2 | Diagnosis | 4 to 24 years, in two age-based forms | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements | No published standardized Arabic version is known. Research has used Arabic tasks built on the same idea (Taibah & Haynes, 2011), but they may not be interpreted against the original norms | **Ch. 8 Ch** |
| **Test of Word Reading Efficiency, Second Edition** TOWRE-2 | Diagnosis Monitoring | 6 to 24 years | An examiner trained in administering and interpreting individually administered standardized tests, in accordance with the publisher's requirements | No published standardized Arabic version is known. The closest we found is a Jordanian measure of word recognition fluency for monitoring first graders (Abu-Hamour, 2014b), which is a monitoring tool, not a diagnostic test | **Ch. 8** |
| **Curriculum-based measures: oral reading fluency, maze, math computation and written expression** CBM | Screening Monitoring | Passages and problems drawn from the student's grade-level curriculum | — | Arabic studies of curriculum-based reading fluency measures (Jordan; United Arab Emirates) with small samples; they do not provide norms (Abu-Hamour, 2014a; Mahfouz & Mohamed, 2023) | **Ch. 4 Ch** |

# A.3 Adaptive behavior

**Table A.3: Adaptive behavior scales used in diagnosing intellectual disability and determining its severity.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Vineland Adaptive Behavior Scales, Third Edition** Vineland-3 | Rating Diagnosis | Birth to 90 years for the interview and parent form; 3 to 21 years for the teacher form | The interview requires an examiner trained in semi-structured interviewing; check the level specified by the publisher | Arabic catalogs list a title, "Vineland Adaptive Behavior Scale: Saudi Norms," but its author, year and sample, and the edition on which it was based, could not be verified. No published Arabic standardization of the third edition is known | **Ch. 3 Ch** |
| **Adaptive Behavior Assessment System, Third Edition** ABAS-3 | Rating Diagnosis | Birth to 89 years, depending on the form | Completed by a parent or teacher; scored and interpreted by a qualified professional; check the level specified by the publisher | The teacher form was adapted into Arabic in Oman and studied with 410 children with intellectual disability; the study supported a single general factor and high reliability (Emam et al., 2020). No published Saudi standardization is known | **Ch. 3 Ch** |
| **Diagnostic Adaptive Behavior Scale** DABS | Diagnosis | 4 to 21 years | — | Nothing published was found | **Ch. 7** |

# A.4 Development and early screening

**Table A.4: Developmental screening and developmental assessment instruments in early childhood.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Ages & Stages Questionnaires, Third Edition** ASQ-3 | Screening | Questionnaire intervals from 2 to 60 months; usable from 1 to 66 months | Completed by parents; scored and interpreted by a professional trained in early childhood or health care; no psychological assessment license required | No published standardized Arabic version is known, and translation alone does not carry over the U.S. cutoff scores | **Ch. 7** |
| **Bayley Scales of Infant and Toddler Development, Fourth Edition*** Bayley-4 | Diagnosis | 16 days to 42 months | — | No published standardized Arabic version is known | **Ch. 3 Ch** |

# A.5 Autism

**Table A.5: Screening, rating and diagnostic instruments for autism spectrum disorder.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Modified Checklist for Autism in Toddlers, Revised with Follow-Up** M-CHAT-R/F | Screening | 16 to 30 months | Completed by parents; the follow-up interview is conducted by a physician, nurse or specialist trained in it; no psychological assessment license required | The first version was studied in nine Arab countries with 228 children: sensitivity 0.86, specificity 0.80 (Seif Eldin et al., 2008). Approved translations of the revised version are available on the official website, but no published Arabic standardization study of it is known | **Ch. 9** |
| **Social Communication Questionnaire** SCQ | Screening | 4 years and older, provided mental age exceeds 2 years | Completed by a parent; scored and interpreted by a trained professional in accordance with the publisher's requirements | No published standardized Arabic version is known | **Ch. 9** |
| **Social Responsiveness Scale, Second Edition** SRS-2 | Rating | 2 years 6 months to adulthood | — | No published standardized Arabic version is known | **Ch. 9 Ch** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Childhood Autism Rating Scale, Second Edition**<br>CARS-2 | Rating Diagnosis | 2 years and older | A specialist trained in autism assessment, after observing the child and gathering information about the child | An Arabic edition prepared by Leila Akoury-Dirani and Mayssam Alameddine (Akoury-Dirani & Alameddine, 2016 [Arabic]), for which no published description of the sample was found; a factor-structure study with 301 children (Alotaibi & Alotaibi, 2021); and an adapted Iraqi version (Al-Shibly et al., 2025). No published Arabic norms for the T-scores are known | **Ch. 9** |
| **Gilliam Autism Rating Scale, Third Edition**<br>GARS-3 | Rating | 3 to 22 years | Completed by parents, a teacher or a specialist; interpreted by a qualified specialist | An Arabic version studied in Saudi Arabia with 178 participants, with a Cronbach's alpha of 0.971 (Alasmari et al., 2024a, 2024b). No published Saudi norms for the index are known, so scores are read against the U.S. norms | **Ch. 9** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Autism Diagnostic Interview–Revised**<br>ADI-R | Diagnosis | Individuals with a mental age above 2 years | A specialist with formal training in administering and coding the interview | No approved or published standardized Arabic version is known | **Ch. 9** |
| **Autism Diagnostic Observation Schedule, Second Edition**<br>ADOS-2 | Diagnosis | 12 months to adulthood, in five modules | A specialist with formal training in the instrument; the publisher restricts purchase to those with advanced qualifications | No Arabic translation approved by the publisher is known, nor any published standardized Arabic version | **Ch. 9** |

## A.6 Attention, behavior and emotion

**Table A.6: Broadband screening tools, narrowband and broadband rating scales, and universal behavioral screening systems in schools.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Strengths and Difficulties Questionnaire**<br>SDQ | Screening | Parent and teacher forms from 2 to 17 years, with age-specific versions; a self-report form from 11 to 17 years | Completed by parents, teachers and students; interpreted by a trained professional | An Arabic translation is available on the official website; its validity was examined with Yemeni children aged 5 to 12 in a clinical sample (87) and a community sample (100) (Alyahri & Goodman, 2006). No published Saudi norms are known | **Ch. 3 Ch** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Vanderbilt ADHD Diagnostic Rating Scales** VADRS | Rating | Primarily elementary school (about 6 to 12 years); the Saudi study included children aged 5 to 12 | Completed by parents and teachers; interpreted by a physician or qualified specialist | An Arabic version (**VADRS-A**) with parent and teacher forms, studied in Saudi Arabia with 1,812 raters (504 parents and 1,308 teachers) of children aged 5 to 12 (Alqahtani et al., 2024) | **Ch. 3 Ch** |
| **Conners 4th Edition** Conners 4 | Rating | 6 to 18 years for the parent and teacher forms, with a self-report form for older youth | A professional trained in interpreting standardized rating scales, in accordance with the publisher's requirements | No standardized Arabic version published in a peer-reviewed journal is known | **Ch. 3 Ch** |
| **Achenbach System of Empirically Based Assessment: Child Behavior Checklist (CBCL), Teacher's Report Form (TRF) and Youth Self-Report (YSR)** ASEBA | Rating | 6 to 18 years for the parent and teacher forms and 11 to 18 years for the self-report; separate forms for preschoolers | Completed by parents, teachers and students; interpreted by a professional trained in standardized rating scales | No published standardized Arabic version with local norms is known; ask the publisher about licensed translations | **Ch. 3 Ch** |
| **Behavior Assessment System for Children, Third Edition** BASC-3 | Rating | 2 to 21 years 11 months for the parent and teacher forms | — | No published standardized Arabic version is known | **Ch. 3 Ch** |
| **Revised Child Anxiety and Depression Scale** RCADS | Rating Screening | Students in roughly Grades 3 through 12 | Completed by the student and parents; interpreted by a trained professional | No published standardized Arabic version is known | **Ch. 10** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Computerized continuous performance tests**<br>CPT | Diagnosis | Vary by test; most start at school age | A trained examiner administers it and a qualified professional interprets it, in accordance with the publisher's requirements | The task is usually nonverbal, but no published standardized Arabic norms are known for any of these tests; they are a supplementary measure on which a diagnosis cannot rest alone | **Ch. 10** |
| **Student Risk Screening Scale and its extended version for internalizing problems**<br>SRSS · SRSS-IE | Screening | — | The teacher rates every student on a single list; no qualification level is specified | No published standardized Arabic version of universal behavioral screening systems is known | **Ch. 10** |
| **Behavioral and Emotional Screening System**<br>BESS | Screening | — | Brief teacher, parent and student forms; no qualification level is specified | No published standardized Arabic version of universal behavioral screening systems is known | **Ch. 10** |
| **Systematic Screening for Behavior Disorders**<br>SSBD | Screening | — | The teacher ranks the class and rates the highest-ranked students, followed by direct observation in the original version; no qualification level is specified | No published standardized Arabic version of universal behavioral screening systems is known | **Ch. 10** |

# A.7 Hearing, vision and communication

**Table A.7: Procedures for assessing hearing and functional vision, and language tests.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Pure-tone audiometry, tympanometry and objective measures (otoacoustic emissions and auditory brainstem response)** PTA · OAE · ABR | Screening Diagnosis | Objective measures from birth; behavioral testing with age-appropriate methods from about 6 months, and by the conventional method from about age 5 | A licensed audiologist; school screening is conducted by trained personnel under professional supervision | Pure-tone audiometry requires no language. Speech discrimination testing, however, requires equivalent Arabic word lists, and no recent published standardized Saudi lists have been verified | **Ch. 11** |
| **Functional vision assessment and learning media assessment** FVA · LMA | Rating | All ages; LEA Symbols and four-letter charts (**HOTV**) are used with young children from about age three | A teacher of students with visual impairments, together with an orientation and mobility specialist | LEA Symbols do not depend on language. No published Arabic guide to learning media assessment is known, so graded Arabic texts are prepared locally | **Ch. 11** |
| **Clinical Evaluation of Language Fundamentals, Fifth Edition** CELF-5 | Diagnosis | 5 to 21 years 11 months | A speech-language pathologist or qualified examiner, in accordance with the publisher's requirements | No published standardized Arabic version is known, and translating its items does not yield a valid instrument, because the structures it measures are English | **Ch. 11** |

# A.8 Giftedness and creativity

**Table A.8: Group ability tests, rating scales and observation tools, creativity tests, and identification procedures.**

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Cognitive Abilities Test, Form 7**<br>CogAT | Screening<br>Identification | Kindergarten through Grade 12, in grade-based levels | Administered in class by the teacher under standardized directions; results interpreted by a trained professional | No published standardized Arabic version is known | **Ch. 3 Ch** |
| **Naglieri Nonverbal Ability Test, Third Edition**<br>NNAT3 | Screening<br>Identification | Kindergarten through Grade 12 | — | No published standardized Arabic version is known | **Ch. 3 Ch** |
| **Mawhiba Multiple Cognitive Abilities Test**<br>MMCAT | Identification | Grade 3 of elementary school through Grade 10 (first year of secondary school), at three levels | No qualification level is specified; administered by the National Center for Assessment at the Education and Training Evaluation Commission, in partnership with Mawhiba and the Ministry of Education | Originally developed in Arabic for the Saudi context and used nationally since 2011. No published technical manual was found, nor information on its score scale or norming year | **Ch. 3 Ch** |
| **Gifted Rating Scales**<br>GRS | Rating<br>Identification | School Form for Grades 1–8 (ages 6–13); Preschool Form from 4 years to 6 years 11 months | Completed by a teacher who has known the student long enough; interpreted by a professional in accordance with the publisher's requirements | No published standardized Arabic version is known, nor any Saudi standardization study | **Ch. 12** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Scales for Rating the Behavioral Characteristics of Superior Students** SRBCSS | Rating Identification | — | Rating scales completed by the teacher; no qualification level is specified | No published Saudi standardization study was found; the scales have no national norms and are read against local norms built by the program | **Ch. 12** |
| **Teacher's Observation of Potential in Students** TOPS | Identification | Designed for young students | The teacher observes students over several weeks during enriched learning activities; no qualification level is specified | No published Saudi standardization study was found | **Ch. 12** |
| **Traits, Aptitudes, and Behaviors** TABs | Identification | — | Observed by the teacher; no qualification level is specified | No published Saudi standardization study was found | **Ch. 12** |
| **Torrance Tests of Creative Thinking** TTCT | Identification | Kindergarten to adulthood | Administered by a teacher or specialist after training; scored by a trained scorer or by the publisher's scoring service | Figural Form B was standardized in Saudi Arabia in 1416 AH with 1,227 male and female students aged 9 to 16 (Al-Nafi, 1428 AH [Arabic]); the original standardization report could not be consulted, and the norms are nearly thirty years old | **Ch. 12** |

| Instrument | Purpose | Ages | Examiner qualification | Arabic version | Chapter |
|---|---|---|---|---|---|
| **Above-level testing**<br>Above-level | Identification | Typically Grades 3 through 9 in talent search programs | Administered by whoever is qualified to administer the original test under its conditions; interpreted by a professional familiar with the norms of the higher grade | No published Arabic program with reported norms is known. The idea can be applied with Arabic achievement tests for higher grades, provided this is stated explicitly and the result is read against the norms of the grade for which the test was designed | **Ch. 12** |
| **Two-lens assessment protocol for a student suspected of being twice-exceptional**<br>2e | Diagnosis<br>Identification | School age; instruments are selected according to the student's age | A team that includes expertise in both giftedness and disability; individual measures are administered by a qualified psychologist in accordance with each publisher's requirements | No standardized Arabic protocol for twice-exceptional learners is known | **Ch. 13** |

# B Appendix B Glossary

The terms from the chapter glossaries, gathered into a single alphabetical glossary that defines each term and lists the chapters that address it, followed by an English–Arabic index.

## B.1 Terms in alphabetical order (218 terms)

| | |
|---|---|
| **Above-level testing** | Administering a test designed for older students to an advanced student to gauge how far ahead the student is. (Ch. 12) |
| **Acceleration** | Delivering content to a student at a younger age or a faster pace than usual, such as taking a subject with a higher grade. (Ch. 12 Ch. 15) |

| | |
|---|---|
| **Accommodation** | A change in presentation or response mode that does not alter what a test measures. (Ch. 3 Ch. 15) |
| **Adaptive behavior** | The conceptual, social and practical skills a person performs in everyday life. (Ch. 3 Ch. 7) |
| **Advanced Learning Plan (ALP)** | An individual document for a gifted student that specifies areas of strength, measurable goals in those areas, social-emotional goals, and how progress will be monitored. (Ch. 15) |

| | |
|---|---|
| **Age equivalent** | The age at which the average raw score of children equals the student's raw score; it does not mean that the student's skills resemble those of children of that age. (Ch. 2) |
| **Aimline** | A line on a progress-monitoring graph that connects the starting level to the goal, against which the student's data points are compared. (Ch. 4 Ch. 15) |
| **Alternate assessment** | In the U.S. system: an achievement test based on alternate achievement standards for students with the most significant cognitive disabilities; it is not what Chapter 5 means by “alternative methods.” (Ch. 5) |
| **Anchoring** | Allowing the first idea that occurs to the evaluator to govern the interpretation of everything that follows. (Ch. 14) |
| **Arithmetic fluency** | Performing basic arithmetic facts and operations both accurately and quickly. (Ch. 8) |
| **Assessment and diagnosis team** | The school team that determines eligibility under Saudi regulations; it includes members from the school leadership, special education, general education and counseling, along with specialists and the parent. (Ch. 15) |
| **Assessment protocol** | A preplanned set of tools and procedures, each of which answers a specific assessment question. (Ch. 13) |

| | |
|---|---|
| **Attention-Deficit/Hyperactivity Disorder (ADHD)** | A neurodevelopmental disorder marked by a persistent pattern of inattention, hyperactivity-impulsivity, or both. (Ch. 10) |
| **Audiogram** | A graph showing the softest sound a person can hear at different frequencies in each ear. (Ch. 11) |
| **Auditory neuropathy spectrum disorder** | A disorder in which the inner ear functions but the signal is not transmitted to the brain in a synchronized way, so speech understanding is poorer than the audiogram would predict. (Ch. 11) |
| **Augmentative and alternative communication** | Means that supplement or replace speech, such as signs, pictures and speech-generating devices. (Ch. 11) |
| **Authentic assessment** | Assessment that asks students to perform tasks resembling real-life tasks, scored with rubrics. (Ch. 5) |
| **Basal** | The number of consecutive correct responses after which the student is assumed to know the preceding items and is credited with them. (Ch. 3) |
| **Base rate** | The proportion of the norming sample showing a given difference or pattern, which indicates whether the difference is rare or common. (Ch. 2) |
| **Baseline** | The initial measurements gathered before new instruction begins, against which later progress is compared. (Ch. 4 Ch. 15) |

| | |
|---|---|
| **Basic Interpersonal Communicative Skills (BICS)** | The language of everyday conversation, in which a second-language learner reaches fluency close to that of peers in about two years. (Ch. 14) |
| **Behavior function** | What a behavior achieves for the person, such as gaining attention or escaping a task. (Ch. 5) |
| **Behavior intervention plan** | A plan based on the results of a functional assessment that teaches the student a replacement behavior serving the same function and changes the conditions that trigger the problem behavior. (Ch. 5) |
| **Biopsychosocial model** | A model that views disability as the outcome of an interaction between an individual's health condition and environmental and personal factors. (Ch. 1) |
| **Bone conduction** | Transmission of sound to the inner ear through the bones of the skull, bypassing the outer and middle ear. (Ch. 11) |
| **Borderline intellectual functioning** | Intellectual functioning below average yet above the range of intellectual disability. The DSM lists it among conditions that may be a focus of clinical attention rather than as a disorder, and “slow learning” has no counterpart category in current classification systems. (Ch. 6 Ch. 7) |
| **Broadband scale** | A rating scale that covers many behavioral and emotional areas at once. (Ch. 3) |

| | |
|---|---|
| **Camouflaging** | A person's attempts to conceal their traits by imitating others and memorizing ready-made phrases for social situations; it is more common among females. (Ch. 9 Ch. 13) |
| **Catatonia** | Freezing of movement, holding fixed postures, extreme slowing, or purposeless movements; it may accompany autism and calls for urgent medical referral. (Ch. 9) |
| **Ceiling** | The point at which administration stops once the discontinue rule is met; the student is assumed to fail all items beyond it. (Ch. 3) |
| **Ceiling effect** | A test's inability to distinguish among high performers because it contains too few difficult items. (Ch. 12) |
| **Cerebral visual impairment** | Impairment in the brain's processing of what the eyes see, with the eyes themselves intact or nearly so. (Ch. 11) |
| **CHC theory** | A hierarchical model of cognitive abilities comprising a general factor, broad abilities and narrow abilities. (Ch. 3) |
| **Chromosomal microarray** | A genetic test that detects small deletions or duplications of chromosomal material that conventional karyotyping cannot see. (Ch. 7) |
| **Co-occurrence** | The presence of another disorder alongside autism at the same time, such as intellectual disability, |

| | |
|---|---|
| | ADHD or anxiety; each is diagnosed and planned for. (Ch. 9) |
| **Cognitive Academic Language Proficiency (CALP)** | The language of learning and abstract concepts; it takes roughly five to seven years to acquire. (Ch. 14) |
| **Cognitive Proficiency Index, CPI** | A composite score derived from working memory and processing speed subtests, compared with the General Ability Index to describe an uneven profile. (Ch. 12 Ch. 13) |
| **Cohen's kappa** | An index of agreement between observers that removes agreement expected by chance; it is more accurate than simple percentage agreement when the behavior is very rare or very frequent. (Ch. 4) |
| **Combination rule** | The way scores on multiple criteria are combined into a single decision, including the “and” rule, the “or” rule and the averaging rule. (Ch. 6) |
| **Compensatory strategies** | Methods a student uses to work around a difficulty, such as memorization or reliance on context; they may also mask it. (Ch. 13) |
| **Computerized adaptive test** | A test in which the computer selects each item on the basis of the student's previous responses and stops once the ability estimate reaches a specified level of precision. (Ch. 3) |
| **Conduct disorder** | A repetitive and persistent pattern of behavior that violates the rights of others or major age-appropriate social norms. (Ch. 10) |

| | |
|---|---|
| **Conductive hearing loss** | Hearing loss originating in the outer or middle ear, appearing as a gap between air and bone conduction; it may be temporary or fluctuating. (Ch. 11) |
| **Confidence interval** | A range of scores within which the true score is expected to fall with a specified probability. (Ch. 2) |
| **Confirmation bias** | The evaluator's tendency to notice evidence that supports a prior hypothesis and to overlook evidence that contradicts it. (Ch. 1 Ch. 14) |
| **Consensual Assessment Technique** | Assessing the creativity of a product through independent ratings by judges familiar with the domain, then examining their agreement. (Ch. 12) |
| **Continuous performance test** | A computerized task that measures sustained attention and impulsivity through responses to a series of rapidly presented stimuli. (Ch. 10) |
| **Contrast sensitivity** | The ability to distinguish an object from its background when their colors are close. (Ch. 11) |
| **Creative-productive giftedness** | The ability to produce original ideas and products for a real audience; Renzulli contrasts it with schoolhouse giftedness, that is, skill at learning lessons and taking tests. (Ch. 12) |
| **Criterion-referenced test** | A test on which a student's score is interpreted against a predetermined standard of performance or mastery. (Ch. 2) |

| | |
|---|---|
| **Culture-Language Interpretive Matrix (C-LIM)** | A method for interpreting a score pattern according to the cultural loading and linguistic demands of the tests, by comparing it with the range expected for students with similar language exposure. (Ch. 14) |
| **Curriculum-based measurement** | Brief, repeated measures with standardized procedures, drawn from the curriculum, that monitor progress in basic skills. (Ch. 4) |
| **Cut-off score** | The dividing score set by an instrument or a system for making a decision. (Ch. 1 Ch. 2 Ch. 3 Ch. 9) |
| **Decibel** | The unit of sound intensity; the higher the value, the louder the sound. (Ch. 11) |
| **Decision rule** | A rule agreed in advance for changing the intervention or the goal, such as four consecutive data points below the aimline or a trend line slope lower than the aimline slope. (Ch. 15) |
| **Decoding** | Converting a written word into its spoken form by linking letters to their sounds. (Ch. 8) |
| **Descriptive assessment** | Recording what precedes and follows a behavior in its natural settings without altering conditions; it shows what accompanies the behavior, not what causes it. (Ch. 5) |
| **Developmental coordination disorder** | A neurodevelopmental disorder affecting the acquisition and execution of coordinated motor skills; it may appear as poor handwriting and slow writing. (Ch. 8) |

| | |
|---|---|
| **Developmental language disorder** | A persistent, impactful language difficulty not associated with a known condition such as hearing loss, autism or intellectual disability; no discrepancy with nonverbal intelligence is required. (Ch. 11) |
| **Developmental learning disorder** | The ICD-11 counterpart of specific learning disorder. (Ch. 8) |
| **Diagnosis** | The judgment, based on a comprehensive evaluation, that a student meets the criteria for a particular category. (Ch. 1) |
| **Diagnostic overshadowing** | Attributing new symptoms to a known disability, so that another condition is overlooked. (Ch. 1 Ch. 7 Ch. 14) |
| **Dialectal difference** | Pronunciation that reflects the dialect of the child and the child's community; it is not a disorder, even when it departs from Modern Standard Arabic. (Ch. 11) |
| **Differential diagnosis** | Verifying that the symptoms are not better explained by another condition. (Ch. 9) |
| **Differential item functioning, DIF** | A difference between two groups of equal ability in the probability of answering an item correctly. (Ch. 2 Ch. 14) |
| **Diglossia** | The coexistence of two varieties of a language within one community, such as Standard Arabic and the colloquial dialect, each with its own uses. (Ch. 8) |

| | |
|---|---|
| **Discontinue rule** | The number of consecutive errors after which the examiner stops administering a subtest. (Ch. 3) |
| **Discrepancy model** | Identifying a learning disability by a large gap between ability and achievement scores. (Ch. 6) |
| **Disorder of intellectual development, provisional** | An ICD-11 category for a child younger than four, or for an individual who cannot be validly assessed. (Ch. 7) |
| **Disproportionality** | A marked difference between a group's share of a category or program and its expected share. (Ch. 14) |
| **Disruptive mood dysregulation disorder** | Severe, recurrent temper outbursts with persistently irritable mood between outbursts, beginning before age 10. (Ch. 10) |
| **Divergent thinking** | The ability to generate many varied and original ideas in response to a single problem. (Ch. 12) |
| **DT→PI** | A model used in talent search programs that pinpoints what a student does not yet know of an advanced course, teaches only that material, and then tests the student at the end of the course. (Ch. 12) |
| **Dual differentiation** | Instruction that provides challenge in areas of strength and support in areas of difficulty at the same time. (Ch. 13) |
| **Dual discrepancy** | A student's falling below peers in both level of performance and rate of growth. (Ch. 6 Ch. 8) |

| | |
|---|---|
| **Dynamic assessment** | Assessment in which the examiner intervenes and deliberately teaches, then measures the effect of that teaching and how the student responds to it. (Ch. 5) |
| **Dyscalculia** | A pattern of learning difficulty marked by weaknesses in number sense, number facts, and calculation accuracy and fluency. (Ch. 8) |
| **Dysgraphia** | A term used for difficulty with handwriting or with spelling and written expression; its DSM counterpart is specific learning disorder with impairment in written expression. (Ch. 13) |
| **Dyslexia** | A pattern of learning difficulty marked by weaknesses in the accuracy and fluency of word recognition and in spelling, usually rooted in a phonological weakness. (Ch. 8) |
| **Early intervention** | Services for young children who are developmentally delayed or at risk of delay, and for their families; each country's system sets the age range and eligibility criteria. (Ch. 5) |
| **Echolalia** | Repeating words or phrases the child has heard from others, immediately or after a delay, often with the same intonation. (Ch. 9) |
| **Eligibility** | The determination that a student qualifies for special services, based on the diagnosis and its impact on the student's learning. (Ch. 1) |

| | |
|---|---|
| **Eligibility determination** | The decision that a student qualifies for services; it usually requires both the presence of a category and a need for specially designed instruction. (Ch. 15) |
| **Emotional disturbance** | An educational category in U.S. law for students whose emotional or behavioral characteristics impair their learning markedly and over a long period. (Ch. 10) |
| **Enrichment** | Broadening and deepening the curriculum for a gifted student without changing the student's grade placement or pace of progress. (Ch. 15) |
| **Equity formula** | A rule that sets the minimum acceptable representation of a group in a gifted program at its share of the student population multiplied by 0.8. (Ch. 14) |
| **Error analysis** | Examining patterns in a student's errors to uncover the faulty rules the student is applying. (Ch. 4) |
| **Excellence gaps** | Differences among groups in the proportions of students who reach advanced levels of performance. (Ch. 12 Ch. 14) |
| **Executive functions** | Abilities that regulate behavior toward a goal, such as planning, working memory and self-control. (Ch. 10) |
| **Exome sequencing** | A genetic test that reads the protein-coding portions of nearly all genes in search of a causal variant. (Ch. 7) |

| | |
|---|---|
| **Expanded core curriculum** | Skills a student with visual impairment needs to be taught explicitly, in nine areas that include braille, orientation and mobility, and independent living. (Ch. 11) |
| **Externalizing problems** | Problems directed outward toward others, such as aggression, defiance and rule-breaking. (Ch. 10) |
| **Floor and ceiling** | A test's capacity to discriminate among very low and very high levels of performance. (Ch. 2) |
| **Flynn effect** | Generational change in mean performance on intelligence tests—an upward drift through most of the twentieth century—which, where present, causes outdated norms to inflate scores. (Ch. 2) |
| **Functional analysis** | Systematically manipulating conditions to test a hypothesis about the function of a behavior experimentally. (Ch. 5) |
| **Functional behavioral assessment** | A process for identifying the function of a problem behavior and the conditions that trigger and maintain it. (Ch. 5 Ch. 10) |
| **Functional language** | A child's ability to use language for everyday needs and communication; it is one of the two classification specifiers in **ICD-11**. (Ch. 9) |
| **Functional vision assessment** | An assessment of how a student actually uses vision in learning situations and daily life. (Ch. 11) |
| **General Ability Index** | A Wechsler composite score that measures reasoning and comprehension without working |

| | |
|---|---|
| | memory and processing speed. (Ch. 8 Ch. 12) |
| **Gift** | High natural ability not yet refined through learning and practice. (Ch. 12) |
| **Gifted underachiever** | A student of very high ability whose achievement falls clearly below it. (Ch. 12) |
| **Global developmental delay** | A provisional diagnosis for a child under five who is delayed in two or more developmental domains. (Ch. 7) |
| **Graduated prompts** | Aids the examiner offers, from the most general to the most specific, recording how many the student needed to reach the solution. (Ch. 5) |
| **Halo effect** | The tendency for a single salient trait, positive or negative, to color judgments of many other qualities. (Ch. 14) |
| **Horn effect** | The negative counterpart of the halo effect: a salient negative trait lowers judgments of unrelated qualities. (Ch. 14) |
| **Hybrid model** | A model for identifying learning disabilities that combines inadequate response to intervention, low achievement and the exclusion of other explanations. (Ch. 6) |
| **Identification** | In this book: the stage of gathering initial evidence and trying classroom supports in preparation for a referral decision. Much of the literature uses the term for the entire process through to the final decision; in gifted education it denotes a multiple- |

| | |
|---|---|
| | criteria educational process for determining who needs gifted services, not a clinical diagnosis. (Ch. 1) |
| **Identification; detection** | A term that overlaps in the Arabic literature with “identification” and is commonly used for the process of searching for and identifying gifted students or students with disabilities, as in the “National Program for Gifted Identification” and the “early detection of disability.” (Ch. 1) |
| **In partial remission** | A specifier used when full criteria were previously met but have not been fully met over the past six months, while impairment in functioning persists. (Ch. 10) |
| **Independent educational evaluation** | An evaluation conducted by a party outside the school. Under U.S. law, a family that disagrees with the school's evaluation may request one at public expense. (Ch. 15) |
| **Individualized Education Program (IEP)** | A document that sets out the student's present levels, goals, services and accommodations, and how progress will be measured. (Ch. 15) |
| **Informant** | The person who provides information about the student on a rating scale or in an interview, such as a parent, a teacher or the student. (Ch. 3) |
| **Informant discrepancy** | Differences among parents', teachers' and students' ratings of the same problems; it is read as information about context, not as measurement error. (Ch. 10) |

| | |
|---|---|
| **Informed consent** | The family's consent given after learning what will be done, why, and who will have access to the results. (Ch. 1) |
| **Intellectual disability** | Significant limitations in both intellectual functioning and adaptive behavior, with onset during the developmental period. (Ch. 7) |
| **Interdisciplinary team** | A team whose members each assess the student in their own field, then share their findings and reach a joint decision and plan together. (Ch. 1) |
| **Internalizing problems** | Problems directed inward toward the self, such as anxiety, sadness, withdrawal and somatic complaints. (Ch. 10 Ch. 14) |
| **Interobserver agreement** | The proportion of agreement between independent observers recording the same behavior. (Ch. 4) |
| **Interval recording** | Dividing the observation period into short intervals and recording whether the behavior occurs in each. (Ch. 4) |
| **Interview-informed synthesized contingency analysis** | A streamlined form of functional analysis that combines the conditions identified in an open-ended interview into a single test condition, compared against a control condition. (Ch. 5) |
| **Intra-individual comparison** | Comparing a student's performance in one area with the student's abilities in other areas. (Ch. 13) |
| **Item Response Theory, IRT** | A measurement framework that describes each item with a curve linking ability to the probability |

| | |
|---|---|
| | of a correct response; measurement precision varies with ability level. (Ch. 2) |
| **Joint attention** | When a child directs another person's attention to something, or follows that person's attention to it, so that both share an interest in it—for example, pointing at an airplane and looking at their mother. (Ch. 9) |
| **Language deprivation** | A child's reaching an advanced age without a fully acquired language, spoken or signed, because the child never had full access to one. (Ch. 11) |
| **Learning disabilities** | The educational term for the category of students who qualify for special education services because of a specific learning difficulty, as defined by the education system. (Ch. 8) |
| **Learning efficiency (Gl)** | A broad ability in **CHC** theory concerning the speed of storing and consolidating new information. (Ch. 3) |
| **Learning media assessment** | An assessment that determines the most suitable medium through which a student with low vision or blindness learns to read and write. (Ch. 11) |
| **Limited prosocial emotions** | A specifier for conduct disorder describing lack of remorse or guilt, callousness and lack of empathy, unconcern about performance, and shallow or deficient affect. (Ch. 10) |
| **Local norms** | Norms that compare a student with peers in the same school or district rather than with a national |

| | |
|---|---|
| | sample. (Ch. 6 Ch. 12) |
| **Masking** | The giftedness compensates for the disability and the disability depresses the giftedness, so that overall performance appears average. (Ch. 13) |
| **Measurable goal** | A goal that specifies the behavior, the conditions, the criterion and the time frame, so that its attainment can be judged. (Ch. 15) |
| **Measurement invariance** | Evidence that a measure assesses the same construct in the same way in two groups, so their scores can be validly compared. (Ch. 2 Ch. 14) |
| **Mediated learning experience** | Learning in which an adult mediates between the child and the task, directing the child's attention, organizing information and helping the child transfer what has been learned. (Ch. 5) |
| **Meta-analysis** | A statistical method that pools the results of multiple studies on a single topic and analyzes them together to reach a more general conclusion. (Ch. 6) |
| **Miscue analysis** | Analysis of oral reading errors in terms of whether they preserve meaning, fit the sentence structure and resemble the printed word. It rests on a model that contemporary reading science has abandoned. (Ch. 4) |
| **Mode equivalence** | Evidence that a test's digital format yields scores with the same meaning as its paper format, so the paper norms remain valid for it. (Ch. 3) |

| | |
|---|---|
| **Modifiability** | The extent to which a student benefits from mediated instruction and is able to transfer what has been learned. (Ch. 5) |
| **Modification** | A change that alters what a test measures or its difficulty, so comparison with the norms is no longer valid. (Ch. 3 Ch. 15) |
| **Momentary time sampling** | Recording whether the behavior occurs only at the moment each interval ends. (Ch. 4) |
| **Multi-tiered system of supports (MTSS)** | A comprehensive schoolwide framework for delivering academic and behavioral support at increasing levels of intensity. (Ch. 6) |
| **Multidisciplinary team** | A group of specialists from different fields who share in assessing the student, each usually working independently. (Ch. 1) |
| **Narrowband scale** | A rating scale that focuses on a single disorder or domain and measures it in depth. (Ch. 3) |
| **Negative predictive value** | The proportion of those with a negative result who truly do not have the problem; it is affected by the prevalence of the problem in the group. (Ch. 1 Ch. 2) |
| **Nomination** | A teacher, parent or student's proposal of a student's name for evaluation in a gifted identification program. (Ch. 1) |
| **Nonresponse criterion** | The rule the team sets in advance for judging that a student has not responded to intervention, such as |

| | |
|---|---|
| | final status, slope or dual discrepancy, with a specified threshold. (Ch. 6) |
| **Norm sample** | The group to which an instrument was administered to derive its norms, and against which each student's score is later compared. (Ch. 2) |
| **Norm-referenced test** | A test on which a student's score is interpreted against the scores of same-age or same-grade peers in the norming sample. (Ch. 2) |
| **Normative comparison** | Comparing a student's performance with that of same-age peers. (Ch. 13) |
| **Number sense** | An intuitive grasp of quantities and their relations, such as comparing and estimating them and placing numbers on a number line. (Ch. 8) |
| **Operational definition** | Describing a behavior in terms of what can be seen and heard, so that two observers would agree on whether it occurred. (Ch. 4) |
| **Opportunity to display** | The situations a learning environment offers in which a student can show their abilities. (Ch. 12) |
| **Oppositional defiant disorder** | A persistent pattern of angry or irritable mood, argumentative and defiant behavior, and vindictiveness. (Ch. 10) |
| **Patterns of strengths and weaknesses (PSW)** | A model that looks for a specific cognitive weakness that explains an academic weakness alongside generally sound abilities; it has three well-known variants. (Ch. 6) |

| | |
|---|---|
| **Percentile rank** | The percentage of the student's peers in the norming sample (of the same age or grade) who scored below the student; a percentile rank of 84 means the student scored higher than 84% of them. It is not the percentage of correct answers. (Ch. 2) |
| **Phonological awareness** | The ability to perceive the sounds in words and to manipulate them through deletion, blending and segmentation. (Ch. 8) |
| **Phonological processing** | The brain's handling of the sounds of language, including phonological awareness, phonological memory and rapid naming. (Ch. 8) |
| **Play-based assessment** | Observing a young child at play to assess the various domains of development. (Ch. 5) |
| **Portfolio** | A selected collection of the student's work gathered over time according to a purpose and criteria. (Ch. 4) |
| **Positive predictive value** | The probability that the condition is truly present in someone with a positive result; it is affected by how prevalent the condition is in the group tested. (Ch. 2 Ch. 9) |
| **Pre-symbolic communication** | A stage in which the child communicates through body movements, facial expressions and actions, before words, signs and pictures. (Ch. 11) |
| **Present levels of performance** | A description of what the student can do now, with strengths and areas of need; it is the starting point for goals. (Ch. 15) |

| | |
|---|---|
| **Presentation** | The current pattern of symptoms: predominantly inattentive, predominantly hyperactive/impulsive, or combined. (Ch. 10) |
| **Pseudowords** | Made-up words that follow the sound rules of the language, used to measure decoding in isolation. (Ch. 8) |
| **Rapid automatized naming** | The speed of naming a series of familiar symbols; it is linked to reading fluency. (Ch. 8) |
| **Rating scale** | An instrument on which someone who knows the child rates the extent to which specific behaviors occur, using a graded scale. (Ch. 9) |
| **Raw score** | The number of correct responses or total points before conversion; it has no meaning on its own. (Ch. 2) |
| **Reevaluation** | A periodic or as-needed evaluation that asks whether the category still applies, whether services are still needed, and what the student's current profile looks like. (Ch. 15) |
| **Referral** | A formal request for a comprehensive evaluation of a student, accompanied by supporting evidence and the family's consent. (Ch. 1) |
| **Regression toward the mean** | The tendency of extreme scores to move closer to the mean when measurement is repeated. (Ch. 2) |
| **Regression-based discrepancy** | Comparing actual achievement with the achievement predicted from the ability score by a |

| | |
|---|---|
| | regression equation, rather than with the ability score itself. (Ch. 6) |
| **Relative age effect** | The higher likelihood that the youngest students in a class will be diagnosed, because they are compared with older classmates. (Ch. 10) |
| **Reliability/precision** | The consistency of a measure's results across time, items or raters, and the precision of the score in general. (Ch. 2) |
| **Representation index** | A group's share of a program's students divided by its share of the student population; a value of 1 indicates proportional representation. (Ch. 14) |
| **Response to intervention (RTI)** | A model that provides tiered support and monitors the student's progress, treating inadequate response as evidence of the need for a comprehensive evaluation. (Ch. 6) |
| **Retrieval fluency (Gr)** | A broad ability in **CHC** theory concerning the speed of retrieving stored information. (Ch. 3) |
| **Review, Interview, Observe, Test / Instruction, Curriculum, Environment, Learner** | A matrix for gathering evidence through four methods (review, interview, observation and testing) across four domains (instruction, curriculum, environment and learner). (Ch. 1) |
| **Risk index** | The proportion of a group's members who have been diagnosed. (Ch. 14) |
| **Risk ratio** | A group's risk index divided by the risk index of a comparison group. (Ch. 14) |

| | |
|---|---|
| **Routing subtest** | A brief subtest that determines the starting level for the remaining subtests of a battery. (Ch. 7) |
| **Rubric** | A tool that sets out the criteria for evaluating a completed piece of work and describes, in words, the levels of performance on each criterion. It should not be confused with a **Rating scale**, which estimates the frequency or intensity of behavior. (Ch. 4 Ch. 5) |
| **Sandwich and cake formats** | Two formats of dynamic assessment: the first places instruction between a pretest and a posttest; the second delivers prompts within each item. (Ch. 5) |
| **Saudi Sign Language** | A complete natural language used by deaf people in the Kingdom, with its own vocabulary and grammar. (Ch. 11) |
| **Scaled score** | A standard score with a mean of 10 and a standard deviation of 3, used on subtests to compare a student's performance with the norming sample. (Ch. 9) |
| **Screening** | A quick procedure applied to an entire group to identify who needs more in-depth evaluation. (Ch. 1 Ch. 9) |
| **Sensitivity** | A tool's ability to detect those who actually have the problem, a property strongly required of screening tools. (Ch. 1 Ch. 2 Ch. 9) |

| | |
|---|---|
| **Sensorineural hearing loss** | Hearing loss originating in the inner ear or the auditory nerve; it is usually permanent. (Ch. 11) |
| **Simple view of reading** | A model that defines reading comprehension as the product of decoding and language comprehension; Chapter 8 also refers to it as the simple model of reading. (Ch. 8) |
| **Social maladjustment** | A term left undefined in U.S. regulations; students who are socially maladjusted are excluded from the emotional disturbance category unless they are also shown to have an emotional disturbance. (Ch. 10) |
| **Specific learning disorder** | The DSM term for persistent difficulty learning basic academic skills that is not better explained by other factors. (Ch. 8) |
| **Specificity** | An instrument's ability to rule out those who do not have the problem. (Ch. 1 Ch. 2 Ch. 9) |
| **Speech sound disorder** | Persistent difficulty producing speech sounds that is inappropriate for the child's age and interferes with the intelligibility of the child's speech. (Ch. 11) |
| **Standard deviation** | A measure of how widely scores are spread around their mean. (Ch. 2) |
| **Standard error of estimate** | The spread of actual scores around the scores predicted by a regression equation; it is used to judge how far achievement falls from what is expected. (Ch. 6) |

| | |
|---|---|
| **Standard Error of Measurement, SEM** | An estimate of how much an individual's score fluctuates around the true score because reliability is imperfect. (Ch. 2) |
| **Standard error of the difference** | The expected amount of error in the difference between two scores, from which the smallest statistically significant difference is calculated. (Ch. 2) |
| **Standard score** | A transformed score with a fixed mean and standard deviation, most commonly a mean of 100 and a standard deviation of 15. (Ch. 2) |
| **Standardized administration** | Administering an instrument with the same directions, materials, timing and scoring procedures used with the norming sample. (Ch. 3) |
| **Stopping** | Replacing a fricative sound, such as /s/, with a stop sound, such as /t/. (Ch. 11) |
| **Support needs intensity** | The amount and type of support a person needs to participate in life activities. (Ch. 7) |
| **T-score** | A standard score with a mean of 50 and a standard deviation of 10, used in most rating scales. (Ch. 3) |
| **Tactile communication** | Communication that relies on touch, such as signs the child feels with a hand placed over the signer's hand. (Ch. 11) |
| **Talent** | A well-developed skill in a specific domain, grown out of a gift through learning and practice. (Ch. 12) |

| | |
|---|---|
| **Talent pool** | A broad group of students, roughly 15–20%, who receive enrichment services in Renzulli's model, with flexible entry for anyone who displays gifted behaviors. (Ch. 12) |
| **Teleassessment** | Administering an assessment instrument by video connection, with the examiner in one location and the student in another. (Ch. 3) |
| **Test adaptation** | Modifying a measure's items and examples to fit a new culture and studying the result empirically; it is broader than translation and narrower than local norming. It differs from “accommodation” in the sense of supports provided to a student during testing. (Ch. 2) |
| **Three-Ring Conception** | Renzulli's conception of gifted behavior as the intersection of above-average ability, task commitment and creativity. (Ch. 12) |
| **Transdisciplinary team** | A team that assesses the child together in a single session, with one member working directly with the child under the guidance of the others. (Ch. 1) |
| **Treatment fidelity** | The extent to which an intervention is delivered as designed: the number and length of sessions and adherence to its steps. (Ch. 6) |
| **Trend line** | The line of best fit through a student's data points on a progress graph; its slope is compared with that of the aimline. (Ch. 15) |

| | |
|---|---|
| **Twice-exceptional** | Students who combine giftedness with a disability or learning difficulty, either of which may mask the other. (Ch. 8 Ch. 9 Ch. 12 Ch. 13) |
| **Tympanometry** | A measure of eardrum movement; a flat tracing usually indicates fluid behind the eardrum. (Ch. 11) |
| **Universal behavior screening** | Periodically rating every student in a class with a brief tool to detect those who need support or evaluation, including students who have not been referred. (Ch. 10) |
| **Universal screening** | Assessing all students in a grade or school level, rather than relying on nomination or family registration. (Ch. 6 Ch. 12 Ch. 14) |
| **Validity** | The degree to which evidence supports the interpretation of test scores for a particular purpose; validity evidence is drawn from five sources. (Ch. 2) |
| **Visual acuity** | Clarity of vision, expressed as a fraction such as 6/6 and measured at distance and near. (Ch. 11) |
| **Visual field** | The area a person can see without moving their eyes. (Ch. 11) |
| **Voice disorders** | Departures in vocal quality, pitch, loudness or resonance from what is appropriate to the speaker's age and sex. (Ch. 11) |
| **Zone of proximal development** | The distance between what a child can do alone and what the child can do with help from someone |

more experienced. (Ch. 5)

## B.2 English–Arabic index

Above-level testing — الاختبار فوق مستوى الصف

Acceleration — الإسراع

Accommodation — التكييف

Adaptive behavior — السلوك التكيفي

Advanced Learning Plan (ALP) — خطة التعلم المتقدم

Age equivalent — العمر المكافئ

Aimline — خط الهدف

Alternate assessment — التقييم البديل

Anchoring — التثبيت على الفرضية الأولى

Arithmetic fluency — الطلاقة الحسابية

Assessment and diagnosis team — فريق القياس والتشخيص

Assessment protocol — بروتوكول التقييم

Attention-Deficit/Hyperactivity Disorder (ADHD) — اضطراب ضعف الانتباه وفرط الحركة

Audiogram — مخطط السمع

Auditory neuropathy spectrum disorder — اعتلال العصب السمعي

Augmentative and alternative communication — التواصل البديل والمعزز

Authentic assessment — التقييم الأصيل

Basal — القاعدة

Base rate — معدل الأساس

Baseline — خط الأساس

Basic Interpersonal Communicative Skills (BICS) — مهارات التواصل الأساسية

Behavior function — وظيفة السلوك

Behavior intervention plan — خطة التدخل السلوكي

Biopsychosocial model — النموذج البيولوجي النفسي الاجتماعي

Bone conduction — التوصيل العظمي

Borderline intellectual functioning — الأداء العقلي الحدي

Broadband scale — مقياس واسع النطاق

Camouflaging — الإخفاء

Catatonia — الجمود الحركي

Ceiling — السقف

Ceiling effect — أثر السقف

Cerebral visual impairment — ضعف البصر الدماغي

CHC theory — نظرية كاتل وهورن وكارول

Chromosomal microarray — التحليل الصبغي الدقيق

Co-occurrence — المصاحبة

Cognitive Academic Language Proficiency (CALP) — الكفاءة اللغوية الأكاديمية

Cognitive Proficiency Index, CPI — مؤشر الكفاءة المعرفية

Cohen's kappa — معامل كابا لكوهين

Combination rule — قاعدة الجمع

Compensatory strategies — الاستراتيجيات التعويضية

Computerized adaptive test — الاختبار التكيفي المحوسب

Conduct disorder — اضطراب المسلك

Conductive hearing loss — فقدان السمع التوصيلي

Confidence interval — فترة الثقة

Confirmation bias — التحيز التأكيدي

Consensual Assessment Technique — أسلوب التقدير التوافقي

Continuous performance test — اختبار الأداء المستمر

Contrast sensitivity — حساسية التباين

Creative-productive giftedness — الموهبة الإبداعية المنتجة

Criterion-referenced test — الاختبار محكي المرجع

Culture-Language Interpretive Matrix (C-LIM) — مصفوفة تفسير الثقافة واللغة

Curriculum-based measurement — القياس المبني على المنهج

Cut-off score — درجة القطع

Decibel — الديسيبل

Decision rule — قاعدة القرار

Decoding — فك الرموز

Descriptive assessment — التقييم الوصفي

Developmental coordination disorder — اضطراب التناسق النمائي

Developmental language disorder — اضطراب اللغة النمائي

Developmental learning disorder — اضطراب التعلم النمائي

Diagnosis — التشخيص

Diagnostic overshadowing — الحجب التشخيصي

Dialectal difference — الاختلاف اللهجي

Differential diagnosis — التشخيص الفارقي

Differential item functioning, DIF — الأداء التفاضلي للبند

Diglossia — الازدواجية اللغوية

Discontinue rule — قاعدة التوقف

Discrepancy model — نموذج التباعد

Disorder of intellectual development, provisional — اضطراب النمو الفكري، مؤقت

Disproportionality — التمثيل غير المتناسب

Disruptive mood dysregulation disorder — اضطراب خلل التنظيم المزاجي التخريبي

Divergent thinking — التفكير التباعدي

DT→PI — الاختبار التشخيصي ثم التدريس الموجه

Dual differentiation — التمايز المزدوج

Dual discrepancy — التباعد المزدوج

Dynamic assessment — التقييم الديناميي

Dyscalculia — عسر الحساب

Dysgraphia — عسر الكتابة

Dyslexia — عسر القراءة

Early intervention — التدخل المبكر

Echolalia — المصداة

Eligibility — الأهلية

Eligibility determination — قرار الأهلية

Emotional disturbance — الاضطراب الانفعالي

Enrichment — الإثراء

Equity formula — معادلة العدالة

Error analysis — تحليل الأخطاء

Excellence gaps — فجوات التميز

Executive functions — الوظائف التنفيذية

Exome sequencing — تسلسل الإكسوم

Expanded core curriculum — المنهج الأساسي الموسع

Externalizing problems — المشكلات الخارجية

Floor and ceiling — الأرضية والسقف

Flynn effect — أثر فلين

Functional analysis — التحليل الوظيفي

Functional behavioral assessment — التقييم الوظيفي للسلوك

Functional language — اللغة الوظيفية

Functional vision assessment — التقييم الوظيفي للبصر

General Ability Index — مؤشر القدرة العامة

Gift — الاستعداد

Gifted underachiever — الموهوب متدني التحصيل

Global developmental delay — التأخر النمائي العام

Graduated prompts — التلميحات المتدرجة

Halo effect — أثر الهالة

Horn effect — أثر القرون

Hybrid model — النموذج الهجين

Identification — التعرف

Identification; detection — الكشف

In partial remission — في حالة هدأة جزئية

Independent educational evaluation — التقييم التربوي المستقل

Individualized Education Program (IEP) — الخطة التربوية الفردية

Informant — المقدّر

Informant discrepancy — تباين المقدّرين

Informed consent — الموافقة المستنيرة

Intellectual disability — الإعاقة الفكرية

Interdisciplinary team — الفريق متداخل التخصصات

Internalizing problems — المشكلات الداخلية

Interobserver agreement — الاتفاق بين الملاحظين

Interval recording — تسجيل الفترات

Interview-informed synthesized contingency analysis — التحليل الوظيفي المركب المبني على المقابلة

Intra-individual comparison — المقارنة داخل الفرد

Item Response Theory, IRT — نظرية الاستجابة للمفردة

Joint attention — الانتباه المشترك

Language deprivation — الحرمان اللغوي

Learning disabilities — صعوبات التعلم

Learning efficiency (GI) — كفاءة التعلم

Learning media assessment — تقييم وسيلة التعلم

Limited prosocial emotions — الانفعالات الاجتماعية الإيجابية المحدودة

Local norms — المعايير المحلية

Masking — الإخفاء المتبادل

Measurable goal — الهدف القابل للقياس

Measurement invariance — تكافؤ القياس

Mediated learning experience — خبرة التعلم الموسَّطة

Meta-analysis — التحليل البعدي

Miscue analysis — تحليل الانحرافات القرائية

Mode equivalence — تكافؤ الصيغ

Modifiability — القابلية للتعديل

Modification — التعديل

Momentary time sampling — العينة اللحظية

Multi-tiered system of supports (MTSS) — نظام الدعم متعدد المستويات

Multidisciplinary team — الفريق متعدد التخصصات

Narrowband scale — مقياس محدد النطاق

Negative predictive value — القيمة التنبؤية السالبة

Nomination — الترشيح

Nonresponse criterion — محك عدم الاستجابة

Norm sample — عينة التقنين

Norm-referenced test — الاختبار معياري المرجع

Normative comparison — المقارنة المعيارية

Number sense — إحساس العدد

Operational definition — التعريف الإجرائي

Opportunity to display — فرصة الإظهار

Oppositional defiant disorder — اضطراب العناد المتحدي

Patterns of strengths and weaknesses (PSW) — أنماط القوة والضعف

Percentile rank — الرتبة المئينية

Phonological awareness — الوعي الصوتي

Phonological processing — المعالجة الصوتية

Play-based assessment — التقييم القائم على اللعب

Portfolio — ملف الإنجاز

Positive predictive value — القيمة التنبؤية الموجبة

Pre-symbolic communication — مرحلة ما قبل الرمزية

Present levels of performance — المستوى الحالي للأداء

Presentation — نمط الظهور

Pseudowords — كلمات لا معنى لها

Rapid automatized naming — التسمية السريعة

Rating scale — مقياس التقدير

Raw score — الدرجة الخام

Reevaluation — إعادة التقييم

Referral — الإحالة

Regression toward the mean — الارتداد نحو المتوسط

Regression-based discrepancy — التباعد القائم على الانحدار

Relative age effect — أثر العمر النسبي

Reliability/precision — الثبات/الدقة

Representation index — مؤشر التمثيل

Response to intervention (RTI) — الاستجابة للتدخل

Retrieval fluency (Gr) — طلاقة الاسترجاع

Review, Interview, Observe, Test / Instruction, Curriculum, Environment, Learner — إطار RIOT/ICEL

Risk index — مؤشر الخطر

Risk ratio — نسبة الخطر

Routing subtest — اختبار التوجيه

Rubric — سلم التقدير اللفظي

Sandwich and cake formats — صيغة الشطيرة وصيغة الكعكة

Saudi Sign Language — لغة الإشارة السعودية

Scaled score — الدرجة الموزونة

Screening — المسح

Sensitivity — الحساسية

Sensorineural hearing loss — فقدان السمع الحسي العصبي

Simple view of reading — النظرة البسيطة للقراءة

Social maladjustment — سوء التوافق الاجتماعي

Specific learning disorder — اضطراب التعلم المحدد

Specificity — النوعية

Speech sound disorder — اضطراب أصوات الكلام

Standard deviation — الانحراف المعياري

Standard error of estimate — الخطأ المعياري للتقدير

Standard Error of Measurement, SEM — الخطأ المعياري للقياس

Standard error of the difference — الخطأ المعياري للفرق

Standard score — الدرجة المعيارية

Standardized administration — التطبيق المقنن

Stopping — الإيقاف

Support needs intensity — شدة الحاجة إلى الدعم

T-score — الدرجة التائية

Tactile communication — التواصل اللمسي

Talent — الموهبة المتحققة

Talent pool — مجمع المواهب

Teleassessment — التطبيق عن بعد

Test adaptation — تكييف المقياس

Three-Ring Conception — نموذج الحلقات الثلاث

Transdisciplinary team — الفريق عابر التخصصات

Treatment fidelity — أمانة التطبيق

Trend line — خط الاتجاه

Twice-exceptional — مزدوجو الاستثنائية

Tympanometry — قياس ضغط الأذن الوسطى

Universal behavior screening — المسح السلوكي الشامل

Universal screening — المسح الشامل

Validity — الصدق

Visual acuity — حدة البصر

Visual field — مجال الإبصار

Voice disorders — اضطرابات الصوت

Zone of proximal development — منطقة النمو القريبة

# List of figures

**Figure 1.1** From screening to eligibility. 7
**Figure 1.2** Three models of teamwork. 31
**Figure 2.1** The normal curve, with the common score types aligned beneath it by position. 57
**Figure 2.2** What each move is worth on a standard score scale with a mean of 100 and a standard deviation of 15. 61
**Figure 2.3** The relationship between reliability and validity, illustrated with target shooting. 63
**Figure 2.4** Majed's score and the confidence interval around it. 64
**Figure 2.5** Two hypothetical items. 73
**Figure 3.1** A hypothetical 20-item subtest with a basal of two consecutive passes and a discontinue rule of three consecutive errors. 109
**Figure 3.2** Ten broad abilities in CHC theory as updated by Schneider and McGrew (Schneider & McGrew, 2018), with the corresponding... 112
**Figure 3.3** The three domains of adaptive behavior, with examples of the skills in each. 123
**Figure 3.4** T-score interpretation ranges on the BASC-3 scales. 129
**Figure 3.5** Mean correlations between different informants' ratings of a child's behavioral and emotional problems, from Achenbach and... 130
**Figure 3.6** Student characteristics and assessment conditions that guide the choice of instrument type. 138
**Figure 4.1** One minute divided into six intervals. 158
**Figure 4.2** A hypothetical graph monitoring Ghadeer's reading fluency. 166
**Figure 4.3** Error analysis in arithmetic and spelling. 171
**Figure 5.1** The test–teach–retest model. 202
**Figure 5.2** A hypothetical example of two children who both started with low scores (8 and 7 out of 20). 203
**Figure 5.3** A graduated prompting scale, from most general to most specific. 204
**Figure 5.4** Four common functions of behavior. 208
**Figure 5.5** The four developmental domains the team observes during a single play-based assessment session. 218
**Figure 6.1** A hypothetical example of the simple-difference discrepancy model. 238
**Figure 6.2** The tiers of response to intervention, from the broad base (Tier 1) to the narrow peak (Tier 3). 242
**Figure 6.3** A hypothetical example of dual discrepancy in reading fluency (words correct per minute). 243
**Figure 6.4** The pattern sought by the patterns of strengths and weaknesses model: a discrepancy between the cognitive strength and the two... 246
**Figure 6.5** Nayef's reading fluency across the three tiers (hypothetical data). 250
**Figure 6.6** Three rules for combining two criteria. 254
**Figure 7.1** The three criteria for diagnosing intellectual disability. 274
**Figure 7.2** The decision pathway for diagnosing intellectual disability. 282
**Figure 7.3** A hypothetical adaptive behavior profile for Nawaf on the Adaptive Behavior Assessment System, as rated by his mother, with... 296
**Figure 7.4** Where the four cases fall. 299
**Figure 8.1** DSM-5-TR criteria for specific learning disorder. 336

**Figure 8.2** The four patterns of the simple view of reading. 339
**Figure 8.3** Four features of Arabic reading that any assessment of dyslexia in Arabic should take into account. 341
**Figure 8.4** The decision pathway for learning disabilities. 347
**Figure 8.5** Ahmed's profile with hypothetical values. 361
**Figure 8.6** Dima's profile with hypothetical values. 366
**Figure 9.1** The five DSM-5-TR criteria. 396
**Figure 9.2** Early red flags that call for evaluation without delay, from Filipek et al. 405
**Figure 9.3** The decision pathway in diagnosing autism spectrum disorder. 408
**Figure 9.4** Outcomes of evaluation for children who still screened positive on the M-CHAT-R/F after the Follow-Up interview: 47.5% were... 411
**Figure 9.5** An original illustration of the checklist's page layout. 415
**Figure 9.6** An original schematic for choosing the right booklet. 421
**Figure 9.7** Total score ranges on the Standard Version for children under 13: below 30, minimal to no symptoms; 30 to 36.5, mild to... 423
**Figure 9.8** An illustrative profile of a teacher rating, using hypothetical values. 425
**Figure 9.9** An original schematic of the path from session to classification in the observation schedule. 430
**Figure 9.10** A suggested layout of the session room, seen from above. 431
**Figure 9.11** Choosing the right module. 432
**Figure 9.12** A hypothetical excerpt from a diagnostic report, worded by the author, with annotations showing what the teacher takes from... 436
**Figure 10.1** DSM-5-TR criteria for attention-deficit/hyperactivity disorder, in abbreviated form. 473
**Figure 10.2** The three presentations. 473
**Figure 10.3** Rate of attention-deficit/hyperactivity disorder diagnosis per 10,000 children by birth month, in states where September 1 is... 476
**Figure 10.4** Externalizing and internalizing problems. 478
**Figure 10.5** The decision pathway for attention-deficit/hyperactivity disorder and behavioral and emotional disorders. 489
**Figure 10.6** Illustrative drawing of the layout of the Strengths and Difficulties Questionnaire form. 493
**Figure 10.7** Illustrative drawing of the layout of the Vanderbilt teacher form. 496
**Figure 10.8** Hypothetical summary of the mother's and teacher's Vanderbilt ratings. 509
**Figure 11.1** A hypothetical audiogram for Khalid. 544
**Figure 11.2** An illustration of a visual acuity chart, with the categories of distance vision impairment in the International... 552
**Figure 11.3** One scene viewed with two visual fields. 553
**Figure 11.4** A child's pronunciation is judged by the standard of his dialect and community. 562
**Figure 11.5** Decision pathway for hearing impairment. 576
**Figure 11.6** Decision pathway for visual impairment. 577
**Figure 11.7** Decision pathway for communication disorders. 579
**Figure 12.1** The Three-Ring Conception. 625
**Figure 12.2** The Differentiated Model of Giftedness and Talent in its second version, simplified by the author (after Gagné, 2009, 2013). 627
**Figure 12.3** A hypothetical illustration of the ceiling effect. 643
**Figure 12.4** An illustrative activity designed by the author in the style of repeated-figures tasks; it is not taken from any published test. 644
**Figure 12.5** A general pathway for identifying gifted learners. 666
**Figure 13.1** A hypothetical example of masking. 693

**Figure 13.2** The three groups of twice-exceptional learners in schools. 694
**Figure 13.3** The two lenses in the debate over identifying twice-exceptional learners. 696
**Figure 13.4** A hypothetical profile for Yasser on the WISC-V indexes. 699
**Figure 13.5** The decision pathway for a student suspected of being twice-exceptional. 709
**Figure 14.1** The risk index made visible. 745
**Figure 14.2** Four points at which bias enters the diagnostic pathway. 748
**Figure 14.3** Share of village school students among the directorate's students and in the gifted program, in the worked example. 752
**Figure 14.4** Approximate time, in years, that a second language learner needs to approach native-speaking peers, after Cummins (Cummins, 1981). 756
**Figure 14.5** A simplified hypothetical example of the Culture-Language Interpretive Matrix. 758
**Figure 15.1** Two structures for the same report. 787
**Figure 15.2** Five steps for a feedback meeting. 797
**Figure 15.3** The chain linking assessment to the plan, illustrated with Mazen's case from the opening scene. 801
**Figure 15.4** Mazen's progress graph with hypothetical values. 804
**Figure 15.5** The cycle of diagnosis and services. 806

# List of tables

**Table 1.1** Purposes of assessment, the question each purpose answers, and an example of a tool for it. 5
**Table 1.2** The difference among testing, measurement, assessment and evaluation. 12
**Table 1.3** Medical and educational classification systems and their purposes. 15
**Table 1.4** The principal regulatory documents for diagnosis and identification in Saudi Arabia. 18
**Table 1.5** The categories of US law (IDEA) and their counterparts in the Special Education Organizational Guide, with the main… 21
**Table 1.6** Principles of fair assessment, what they mean in practice, and their counterparts in Saudi regulations. 26
**Table 1.7** The RIOT/ICEL matrix, with examples in each cell. 30
**Table 1.8** Steps of the general pathway in the US system and their counterparts in Saudi regulations. 35
**Table 1.9** Common errors in diagnosis, how they occur and how to avoid them. 37
**Table 1.10** Sultan's file at the outset: what we have and what is missing. 39
**Table 1.11** Alignment of chapter objectives with review questions and learning activities. 46
**Table 2.1** Common types of standard scores. 58
**Table 2.2** Descriptive classifications of composite scores on the Wechsler Intelligence Scale for Children, Fifth Edition. 59
**Table 2.3** Types of reliability and the question each one answers. 62
**Table 2.4** The five sources of validity evidence in the Standards for Educational and Psychological Testing, with examples from adapting… 75
**Table 2.5** The same tool (sensitivity 85%, specificity 95%) in two groups with different prevalence. 77
**Table 2.6** Forms of measurement bias and how they are examined. 79
**Table 2.7** Three levels in transferring an instrument from one culture to another (the author's simplification). 81
**Table 2.8** Hypothetical scores for a ten-year-old student, modeled on the indexes of the Wechsler Intelligence Scale for Children, Fifth… 85
**Table 2.9** Who administers what in the Kingdom, according to the published documents we reviewed at the time of writing. 91
**Table 3.1** Comprehensive individual intelligence tests and the state of their Arabic versions. 113
**Table 3.2** Nonverbal ability tests and the state of their Arabic versions. 116
**Table 3.3** Three group ability tests used in gifted screening. 119
**Table 3.4** Academic areas and their component skills in individual achievement tests. 120
**Table 3.5** Three comprehensive individual achievement tests and the state of their Arabic versions. 121
**Table 3.6** A comparison of the Vineland scales and the Adaptive Behavior Assessment System. 123
**Table 3.7** Common rating scales and the state of their Arabic versions. 128
**Table 3.8** Accommodation versus modification during testing. 131
**Table 3.9** When teleassessment is suitable and when in-person administration is preferable. 134
**Table 4.1** Formal standardized methods and informal methods. 154
**Table 4.2** Observation recording methods and the behaviors each one suits. 156
**Table 4.3** Distribution of the two observers' records across twenty intervals. 160

**Table 4.4** The best-known curriculum-based measures. 163
**Table 4.5** A classification of Arabic oral reading errors by decoding and structure. 173
**Table 4.6** A brief analytic rubric for written expression. 175
**Table 4.7** Form A, recording three passages and classifying their errors. 180
**Table 4.8** Form B, interval grid. 181
**Table 4.9** Form C, event log. 183
**Table 4.10** Form D, evidence matrix. 185
**Table 4.11** Bandar's informal data. 191
**Table 5.1** Alternative methods and the question each one asks. 200
**Table 5.2** The two formats of dynamic assessment and what each measures. 205
**Table 5.3** Levels of evidence in functional behavioral assessment. 209
**Table 5.4** Examples from Badr's antecedent-behavior-consequence record. 211
**Table 5.5** Traditional testing and authentic assessment. 215
**Table 5.6** Characteristics of assessment from birth to three and what they mean for the assessor. 219
**Table 5.7** Sources of evidence when standardized instruments are not suitable. 223
**Table 5.8** Summary of Hattan's antecedent-behavior-consequence record (eight incidents over eight days). 230
**Table 5.9** Mapping chapter objectives to questions and activities. 230
**Table 6.1** Judgments for three hypothetical students using the simple difference (cutoff 22.5) and regression (cutoff 18 below expected,... 241
**Table 6.2** Common nonresponse criteria, example thresholds and their limitations. 243
**Table 6.3** Strengths and limitations of the response to intervention model. 245
**Table 6.4** The three versions of the patterns of strengths and weaknesses model. 247
**Table 6.5** Nayef’s curriculum-based measurement data across the three tiers (words correct per minute; hypothetical figures). 250
**Table 6.6** Results of Nayef’s comprehensive evaluation (hypothetical standard scores with a mean of 100 and SD of 15; 95% confidence... 251
**Table 6.7** Nayef’s evidence against the criteria of the hybrid model, the DSM and the Saudi regulatory text. 252
**Table 6.8** Applying the three combination rules to seven hypothetical students (cutoff 120). 254
**Table 6.9** Comparison of identification models by their central question, their evidence, and what they may miss. 256
**Table 6.10** Identification rules in published Saudi texts, read through the concepts of this chapter. 257
**Table 6.11** Evidence available on Lujain at referral (hypothetical data). 263
**Table 6.12** Mapping chapter objectives to questions and activities. 264
**Table 7.1** Examples of adaptive behavior deficits in the three domains. 276
**Table 7.2** Severity levels of intellectual disability by adaptive behavior domain (adapted from APA, 2022). 277
**Table 7.3** ASQ-3 at a glance. 287
**Table 7.4** SB5 at a glance. 289
**Table 7.5** Vineland-3 at a glance. 292
**Table 7.6** ABAS-3 at a glance. 295
**Table 7.7** Known Arabic versions of instruments for diagnosing intellectual disability. 298
**Table 7.8** Evidence in Nawaf's case. 300
**Table 7.9** Evidence in Hamad's case. 302
**Table 7.10** Evidence in Maryam's case. 303
**Table 7.11** Evidence in Al-Anoud's case. 307

**Table 7.12** Differential diagnosis of intellectual disability. 309
**Table 7.13** Common co-occurring conditions in intellectual disability and what their assessment requires. 312
**Table 7.14** Criteria for intellectual disability in the DSM and ICD-11, in U.S. 314
**Table 7.15** Mapping chapter objectives to review questions and learning activities. 324
**Table 8.1** Domains of specific learning disorder and their subskills in the DSM. 338
**Table 8.2** Early warning signs in reading, writing, and mathematics by stage. 344
**Table 8.3** Comprehensive Test of Phonological Processing (CTOPP-2) at a glance. 348
**Table 8.4** Test of Word Reading Efficiency (TOWRE-2) at a glance. 350
**Table 8.5** Wechsler Individual Achievement Test (WIAT-4) at a glance. 352
**Table 8.6** KeyMath-3 Diagnostic Assessment (KeyMath-3) at a glance. 354
**Table 8.7** Profile of the Test of Early Mathematics Ability (TEMA-3). 357
**Table 8.8** Profile of the Test of Written Language (TOWL-4). 358
**Table 8.9** Evidence in Lama's case. 362
**Table 8.10** Evidence in Dima's case. 364
**Table 8.11** Evidence in Yazan's case. 368
**Table 8.12** Differential diagnosis of learning disabilities. 370
**Table 8.13** Learning disabilities in the diagnostic manuals, US law, and Saudi regulations. 373
**Table 8.14** Published Arabic instruments and studies in the domains of learning disabilities. 378
**Table 8.15** Data for Muadh's case. 385
**Table 9.1** Forms of Criterion B and examples at school. 399
**Table 9.2** Support levels in autism spectrum disorder (adapted from American Psychiatric Association, 2022). 402
**Table 9.3** A sample observation record for teachers to use before referral, with two entries from Yousef's record. 409
**Table 9.4** M-CHAT-R/F information card. 416
**Table 9.5** M-CHAT-R/F score ranges and the action for each. 417
**Table 9.6** SCQ information card. 418
**Table 9.7** CARS-2 information card. 421
**Table 9.8** GARS-3 information card. 425
**Table 9.9** ADI-R information card. 428
**Table 9.10** ADOS-2 information card. 433
**Table 9.11** Evidence in Sara's case. 437
**Table 9.12** Evidence in Fahd's case, in the order it was gathered. 440
**Table 9.13** Conditions that can resemble autism and what distinguishes them. 445
**Table 9.14** Common co-occurring conditions in autism spectrum disorder and what their assessment requires. 447
**Table 9.15** Autism spectrum disorder in the diagnostic manual and the international classification, in U.S. 449
**Table 9.16** Arabic instruments available for autism assessment, based on what we found published at the time of writing. 453
**Table 9.17** Mapping chapter objectives to review questions and learning activities. 460
**Table 10.1** Severity levels and other specifiers for attention-deficit/hyperactivity disorder in the DSM. 474
**Table 10.2** Abbreviated criteria for the main behavioral and emotional disorders of childhood and adolescence in the DSM (adapted and… 479
**Table 10.3** Signs that warrant attention, by school level. 486
**Table 10.4** Examples of universal behavior screening systems in schools. 490

**Table 10.5** Strengths and Difficulties Questionnaire (SDQ) at a glance. 494
**Table 10.6** Vanderbilt scales at a glance. 497
**Table 10.7** Conners, Fourth Edition (Conners 4) at a glance. 499
**Table 10.8** Achenbach System, school-age forms (ASEBA) at a glance. 502
**Table 10.9** Revised Child Anxiety and Depression Scale (RCADS) at a glance. 504
**Table 10.10** Computerized continuous performance tests (CPT) at a glance. 506
**Table 10.11** Evidence in Haya's case. 510
**Table 10.12** Evidence in Saud's case. 512
**Table 10.13** Evidence in Lamia's case. 514
**Table 10.14** Differential diagnosis of ADHD and behavioral and emotional disorders. 517
**Table 10.15** Attention-deficit/hyperactivity disorder and behavioral and emotional disorders in the diagnostic manuals, US law and Saudi... 521
**Table 10.16** Published Arabic versions of the instruments in this chapter. 526
**Table 10.17** Data in Walid's case. 533
**Table 11.1** Degrees of hearing loss and what a child may miss at each (Clark, 1981). 545
**Table 11.2** Degrees of hearing loss in the World Health Organization classification (WHO, 2021). 546
**Table 11.3** Types of communication disorders and their distinguishing features. 555
**Table 11.4** Areas to examine in a child's Arabic when developmental language disorder is suspected. 560
**Table 11.5** Categories of multiple, physical and health impairments and brain injury in U.S. 563
**Table 11.6** Evidence in Salman's case. 569
**Table 11.7** Warning signs in hearing, vision and communication. 572
**Table 11.8** Profile of pure-tone audiometry, tympanometry and objective measures. 580
**Table 11.9** Profile of functional vision assessment and learning media assessment. 583
**Table 11.10** Profile of the Clinical Evaluation of Language Fundamentals (CELF-5). 585
**Table 11.11** Options for assessing ability in deaf or blind students, and their limits. 588
**Table 11.12** Evidence in Khalid's case. 590
**Table 11.13** Results of Rahaf's educational assessments. 593
**Table 11.14** Observations in Abdulrahman's case and their interpretation. 594
**Table 11.15** Evidence in Mashael's case. 596
**Table 11.16** Differential diagnosis in sensory disabilities and communication disorders. 598
**Table 11.17** This chapter's categories in the diagnostic classifications, US law and the Saudi regulations. 602
**Table 11.18** Published Arabic instruments and studies in sensory disabilities and communication. 607
**Table 11.19** Data in Hadi's case. 615
**Table 12.1** Levels of giftedness in the Differentiated Model (Gagné, 2009) and their approximate equivalents on an intelligence scale with... 628
**Table 12.2** Four models of giftedness from the standpoint of identification. 631
**Table 12.3** Principles of fair identification of gifted learners and what they mean in practice. 633
**Table 12.4** Common nomination and observation tools in identifying gifted learners. 639
**Table 12.5** Profile of the Cognitive Abilities Test. 646
**Table 12.6** Profile of the Gifted Rating Scales. 649
**Table 12.7** Profile of the Torrance Tests of Creative Thinking. 651
**Table 12.8** Profile of above-level testing. 654
**Table 12.9** Criteria for identifying gifted learners in the published Saudi documents we were able to review. 657
**Table 12.10** Profile of the Mawhiba Multiple Cognitive Abilities Test, based on what we found published at the time of writing. 661

**Table 12.11** Arabic instruments available for identifying gifted learners, based on what we found published at the time of writing. 663
**Table 12.12** Evidence in Jana's case. 666
**Table 12.13** Evidence in Mishari's case. 668
**Table 12.14** Evidence in Dana's case. 670
**Table 12.15** Situations in which giftedness may be wrongly suspected or hidden, and how the team handles them. 672
**Table 12.16** Alignment of chapter objectives with review questions and learning activities. 680
**Table 13.1** Forms of twice-exceptionality across the book's chapters, and what each place adds. 691
**Table 13.2** The usual identification models, what each requires, and why each may miss the twice-exceptional student. 695
**Table 13.3** Components of the two-lens assessment protocol, and the question each component answers. 701
**Table 13.4** Protocol information profile. 702
**Table 13.5** The three possibilities in differential diagnosis, what favors each, and what is not sufficient on its own. 710
**Table 13.6** Evidence in Yasser's case through the two lenses. 712
**Table 13.7** Evidence in Muhannad's case. 717
**Table 13.8** Evidence in Joud's case. 720
**Table 13.9** Evidence in Faris's case. 724
**Table 13.10** Twice-exceptionality in the diagnostic manuals, US law and Saudi regulations. 729
**Table 13.11** Data in Arwa's case. 735
**Table 14.1** Sources of bias in the instrument and ways of detecting it. 748
**Table 14.2** Common assessor biases and how to reduce them. 761
**Table 14.3** Evidence in Zain's case. 764
**Table 14.4** Evidence in Ohoud's case. 767
**Table 14.5** Alignment of chapter objectives with review questions and learning activities. 777
**Table 15.1** Statements from diagnostic reports in professional language, rephrased in language parents and teachers can understand. 789
**Table 15.2** From general to specific recommendations, with the finding that supports each one. 790
**Table 15.3** Report appendix: Mazen's scores with confidence intervals (90%). 795
**Table 15.4** Components of the individualized education program in the Saudi Regulations and their counterparts in US law. 799
**Table 15.5** What reevaluation may reveal, and the usual decision in each case. 806
**Table 15.6** Types of decisions for gifted learners, the evidence each decision requires, and what to ask before making it. 808
**Table 15.7** Principles of eligibility, planning, and family rights in the Saudi regulations, and their counterparts in US law. 813
**Table 15.8** Stages of the decision for gifted students in the published Saudi documents. 817
**Table 15.9** Mazen's individualized education program. 822
**Table 15.10** Tamim's evidence at the initial diagnosis and at reevaluation. 824
**Table 15.11** Rakan's dual plan: the strength track and the support track. 827
**Table 15.12** Aseel's data. 836

# Index

Key terms of the book and the pages where they appear in the chapters. Each term is defined in Appendix B.


## A

Above-level testing, 622, 632, 642, 645, 692, 698, 710, 808
Acceleration, 623, 626, 630–631, 719, 807–809
Accommodation, 108, 131, 248, 252, 313, 337–338, 362, 367, 513, 548, 567, 583, 662
Adaptive behavior, 5, 32, 58, 60, 65, 75, 106, 108, 122, 220, 271, 276–277, 370
Advanced Learning Plan (ALP), 810
Age equivalent, 61, 221
Aimline, 165, 191, 363, 803, 823
Alternate assessment, 199, 201, 799, 815
Anchoring, 761
Arithmetic fluency, 356, 360
Assessment and diagnosis team, 4, 18, 317, 376, 516, 523, 569, 813
Assessment protocol, 688
Attention-Deficit/Hyperactivity Disorder (ADHD), 10, 37, 58, 371, 447, 468–469, 509, 564, 623, 667, 672, 690–691
Audiogram, 543, 548–549, 581
Auditory neuropathy spectrum disorder, 581
Augmentative and alternative communication, 541, 558
Authentic assessment, 200, 215–216

## B

Base rate, 68
Baseline, 191, 313, 568, 802
Basic Interpersonal Communicative Skills (BICS), 756
Behavior intervention plan, 200, 212
Biopsychosocial model, 14
Bone conduction, 544, 579–580
Borderline intellectual functioning, 238, 310
Broadband scale, 503

## C

Camouflaging, 692, 720–721, 723
Ceiling, 53, 71, 74, 133, 622, 642, 692, 698, 808
Ceiling effect, 53, 622, 642, 645, 692, 698
Cerebral visual impairment, 569, 571, 584, 599
CHC theory, 108
Chromosomal microarray, 282, 308, 448
Co-occurrence, 271, 311, 331, 446, 454, 468, 520, 542
Cognitive Academic Language Proficiency (CALP), 756
Cognitive Proficiency Index, CPI, 641, 663, 673, 698, 701, 713, 725
Combination rule, 237, 253, 255–256, 659, 674
Computerized adaptive test, 74, 133
Conduct disorder, 475, 480, 485, 496
Conductive hearing loss, 440
Confidence interval, 37, 39, 51–53, 64, 251, 258, 271, 275, 290, 469, 525, 659
Confirmation bias, 37, 761
Consensual Assessment Technique, 644, 653
Continuous performance test, 506
Contrast sensitivity, 551, 582, 593
Creative-productive giftedness, 625
Criterion-referenced test, 5, 55, 206
Culture-Language Interpretive Matrix (C-LIM), 743, 757, 771, 777
Curriculum-based measurement, 5, 54, 122, 154, 163, 167, 169, 242, 244, 263, 345, 360, 364, 379
Cut-off score, 21, 37, 64–67, 119, 258–259, 411–413, 416, 531

## D

Decibel, 21, 544, 581, 602, 735
Decision rule, 165, 167, 186, 244, 803
Decoding, 39, 120, 172, 174, 206, 246, 249–250, 330, 332, 338–339, 713, 765
Descriptive assessment, 208, 213
Developmental coordination disorder, 342, 360, 371, 599, 714, 716
Developmental language disorder, 372, 541, 548, 557, 825–826
Developmental learning disorder, 333, 373
Diagnosis, 2–4, 54–55, 60, 63, 110, 117–118, 120, 168, 172, 174
Diagnostic overshadowing, 37, 311, 319, 761
Dialectal difference, 594, 599

Differential diagnosis, 271, 311, 331, 342, 386, 392, 441, 446, 468, 542, 561, 688, 718
Differential item functioning, DIF, 79, 81, 749
Diglossia, 122, 561
Discontinue rule, 108, 133, 289, 292
Discrepancy model, 67, 70, 236–237, 248, 259, 337, 369, 383
Disorder of intellectual development, provisional, 570
Disproportionality, 743–744, 747
Disruptive mood dysregulation disorder, 475, 481, 483, 534
Divergent thinking, 644, 651, 653
Dual differentiation, 723
Dual discrepancy, 167, 236, 242, 244, 247, 361, 365
Dynamic assessment, 199–201, 561, 563, 587, 599, 630, 632, 759, 765
Dyscalculia, 332, 343, 355, 372
Dysgraphia, 332, 716, 766
Dyslexia, 136, 172, 206, 330–332, 599, 714

## E

Early intervention, 18–19, 23, 118, 219–221, 246, 280, 417–418, 434, 452, 605
Echolalia, 399
Eligibility, 2, 4–5, 67, 101, 108, 135, 176, 178, 200, 206, 219–220, 249
Eligibility determination, 4–5
Emotional disturbance, 17, 22, 214, 252, 334, 375, 451, 468, 483–485
Enrichment, 10, 444, 518, 623, 625–626, 723, 768–769, 808, 818, 828
Error analysis, 5, 154, 164, 170, 186, 355, 360
Excellence gaps, 750
Executive functions, 469, 726
Exome sequencing, 282, 308
Expanded core curriculum, 554
Externalizing problems, 478, 490, 492, 495

## F

Floor and ceiling, 53, 74
Functional analysis, 210, 212–213
Functional behavioral assessment, 199, 207
Functional vision assessment, 579

## G

General Ability Index, 368, 641–642, 692, 697–698, 701, 827
Gifted underachiever, 672
Global developmental delay, 277, 282, 319, 396, 400, 806
Graduated prompts, 204–205

## H

Halo effect, 650, 761, 766
Horn effect, 761, 766
Hybrid model, 236, 248–249, 253, 345

## I

Identification, 2, 4, 8, 80, 110, 113, 117, 216, 235, 248, 253, 255, 334, 370
In partial remission, 474
Independent educational evaluation, 816
Individualized Education Program (IEP), 18, 33, 36, 219, 252, 784, 798, 830
Informed consent, 28, 93, 814
Intellectual disability, 17, 20–21, 37, 56, 60, 65, 67, 110, 117, 122, 124, 222, 224
Internalizing problems, 478, 490, 746
Interobserver agreement, 154, 186
Interval recording, 157–158, 509, 531, 702, 717, 725
Interview-informed synthesized contingency analysis, 213
Intra-individual comparison, 728, 730
Item Response Theory, IRT, 72, 133, 297

## L

Language deprivation, 550, 588, 598
Learning disabilities, 21, 35, 70, 110, 120, 178, 191–192, 237, 249, 280, 302–304
Learning efficiency (GI), 111
Learning media assessment, 579, 582–583
Limited prosocial emotions, 475, 480
Local norms, 10, 37, 122, 169–170, 237, 256, 358, 360, 379, 414, 495, 503, 526

## M

Masking, 172, 372, 519, 688, 693, 808
Measurable goal, 252, 784, 801, 810, 816
Measurement invariance, 81, 749
Mediated learning experience, 201
Meta-analysis, 130, 241, 306, 447, 469, 492, 525
Miscue analysis, 172
Modifiability, 202
Modification, 32, 131, 142, 802–803, 814
Momentary time sampling, 158, 182, 191
Multi-tiered system of supports (MTSS), 242, 489
Multidisciplinary team, 27, 138, 214, 218, 220, 411

## N

Negative predictive value, 8, 77–78

Nomination, 9, 20, 76, 137, 255, 622, 634, 696, 728, 762, 766–767, 769, 817
Nonresponse criterion, 256
Norm sample, 107–109, 243, 287, 290, 703
Norm-referenced test, 206
Normative comparison, 730
Number sense, 338, 343, 357

## O

Operational definition, 154–155, 186
Opportunity to display, 622, 636–637, 639
Oppositional defiant disorder, 475, 479, 481

## P

Patterns of strengths and weaknesses (PSW), 70, 88, 236, 246, 258–259, 334, 369, 383
Percentile rank, 51–52, 169, 191, 243, 353, 357, 422, 500, 567, 642, 647, 652, 655
Phonological awareness, 348–349, 360, 509, 697, 702, 790–792, 795
Phonological processing, 246, 340, 348, 378, 702, 782
Play-based assessment, 199–200, 218
Portfolio, 154, 174, 176, 253, 645, 669–670, 735
Positive predictive value, 8, 77–78, 417
Pre-symbolic communication, 223, 565
Present levels of performance, 36
Pseudowords, 120, 792

## R

Rapid automatized naming, 782
Rating scale, 58–59, 62, 75, 106, 108, 129–130, 175, 215, 252–253, 312, 392
Raw score, 57, 61, 71, 287, 425, 505
Reevaluation, 18, 36, 70, 301, 307, 309, 474, 517, 524, 605, 692, 784, 805
Referral, 8, 10, 30, 80, 91, 152, 168, 177, 221, 282, 284, 301, 304, 352
Regression toward the mean, 53, 70, 207, 241
Relative age effect, 476–477
Reliability/precision, 53, 62, 78, 88
Representation index, 743, 777
Response to intervention (RTI), 21, 35, 168, 177–178, 186, 199, 205, 236, 241, 245, 345, 360, 364
Retrieval fluency (Gr), 111
Risk index, 746–747, 750
Risk ratio, 746–747
Routing subtest, 288–289
Rubric, 154, 175, 215, 217, 360, 828

## S

Saudi Sign Language, 550, 595, 608
Scaled score, 58, 71, 290, 295, 307, 349, 355, 359, 425–426, 586
Screening, 2–4, 63–64, 74, 76, 117, 119, 129, 136, 163, 191, 221
Sensitivity, 7, 53, 76–77, 214, 256, 259, 287, 290, 293, 295, 410, 417, 453
Sensorineural hearing loss, 735
Simple view of reading, 331, 338, 345, 352
Social maladjustment, 485, 515, 522
Specific learning disorder, 237, 249, 252, 264, 331–332, 338, 362, 691, 695, 715, 782, 785, 823
Specificity, 7, 53, 76–77, 259, 287, 290, 293, 295, 410, 417, 433, 531
Speech sound disorder, 555, 557, 595
Standard deviation, 23, 53, 55, 124, 126–127, 221, 237, 240, 244, 275, 280, 287, 337
Standard error of estimate, 66
Standard Error of Measurement, SEM, 53, 62–63, 65, 290, 659
Standard error of the difference, 69, 101, 247, 258, 264
Standard score, 12, 52–53, 55, 57, 129, 206, 221, 237, 243, 254, 306, 308–309
Standardized administration, 109, 136
Support needs intensity, 278

## T

T-score, 58–59, 108, 129, 420, 422–423, 500, 502, 505, 639, 649, 670, 702
Tactile communication, 565
Talent pool, 625, 631
Teleassessment, 108, 134–135
Test adaptation, 53, 81, 131
Three-Ring Conception, 624
Treatment fidelity, 245, 251–252
Trend line, 165, 803, 823
Twice-exceptional, 11, 88, 175, 216, 241, 245, 256, 258, 369–370, 444, 641, 659, 673
Tympanometry, 410, 440, 579–580

## U

Universal behavior screening, 468, 489, 514, 527
Universal screening, 77–78, 237, 249, 255, 361, 364, 417, 453, 516, 527, 634, 640, 646

## V

Validity, 26, 28, 53, 74, 118, 137, 142, 206, 216, 245, 313, 410, 414, 427

Visual acuity, 551–552
Visual field, 22, 552, 592, 603
Voice disorders, 555–556

## Z

Zone of proximal development, 201

## When is it sound to conclude that a student has a disability, a disorder, giftedness or twice-exceptionality, and on what grounds?

This book traces the stages of diagnosis and identification systematically, from referral and the gathering of evidence to the interpretation of that evidence and the decisions built on it. It goes beyond defining the categories and setting out their criteria to examine the nature of measurement instruments, the procedures for administering them, the qualifications required of those who administer them, the principles for interpreting their results, and the point at which the evidence is sufficient to support a judgment.

The book brings together two fields that the literature usually treats separately: disabilities and disorders, and giftedness and twice-exceptionality. It reads international standards and the American experience in light of published Saudi regulations, and describes the status of the Arabic version of each instrument it presents.

Its fifteen chapters address the foundations of measurement; formal, informal and alternative methods of gathering evidence; and identification models; then take up the categories of exceptionality one by one, supported by decision pathways, instrument profiles and applied cases. They close with fairness in diagnosis, report writing, and the development of the individualized education program.

---

Diagnosis and Identification of Exceptional Learners — Dr. Omar A. Alsamani

ISBN 978-603-06-7714-6

9 786030 677146