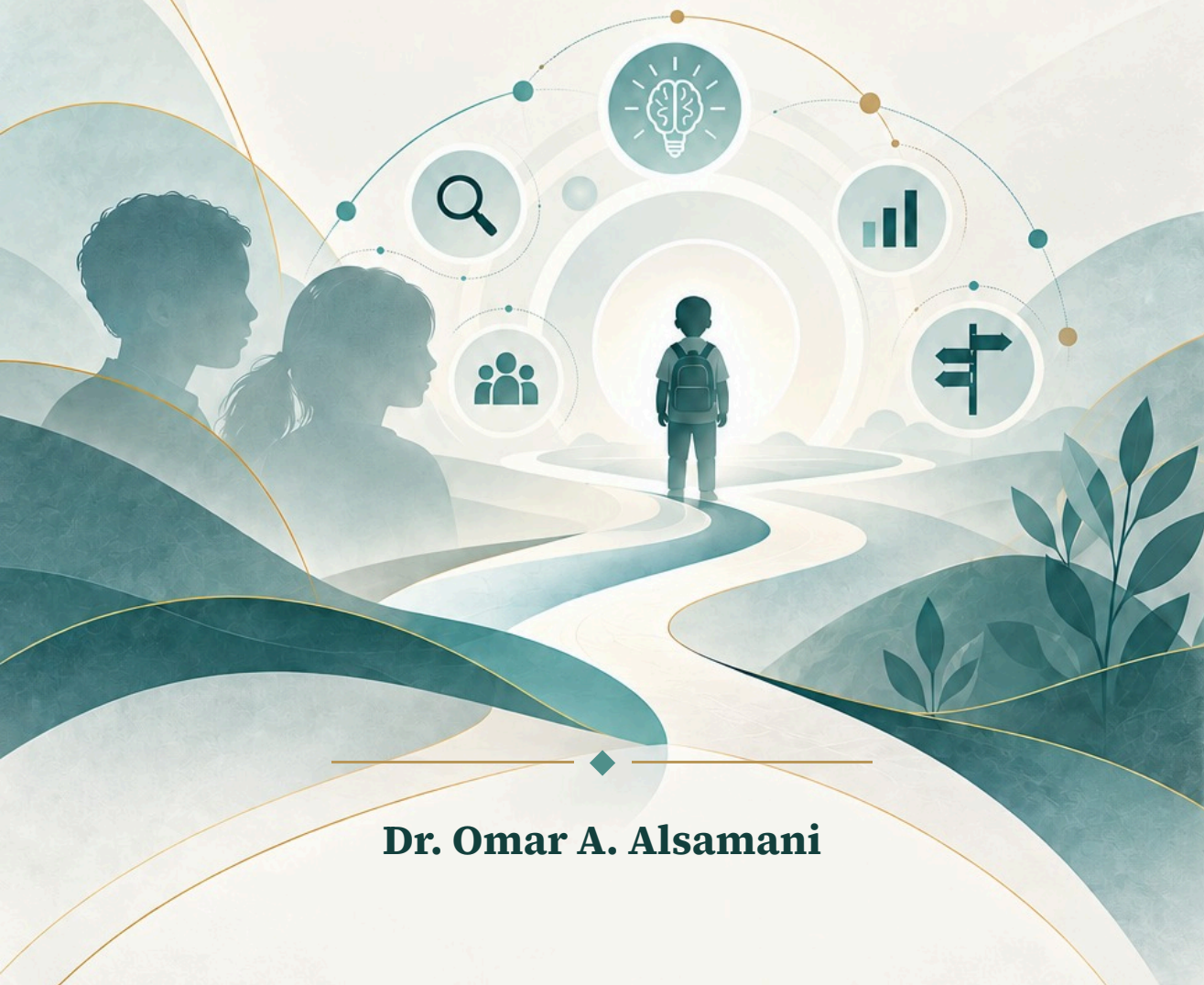


Diagnosis and Identification of Exceptional Learners

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

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.

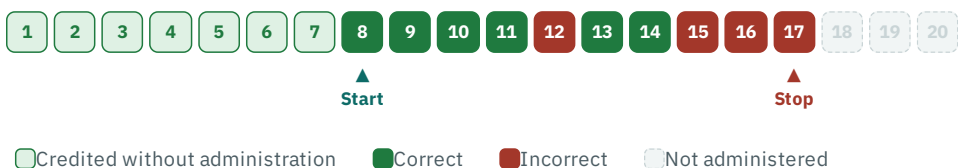


Figure 3.1: A hypothetical 20-item subtest with a basal of two consecutive passes and a discontinue rule of three consecutive errors. The student started at item 8 according to his age and passed items 8 and 9, so items 1 through 7 were credited. The examiner then stopped after three errors on items 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” (**Gl**) 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:

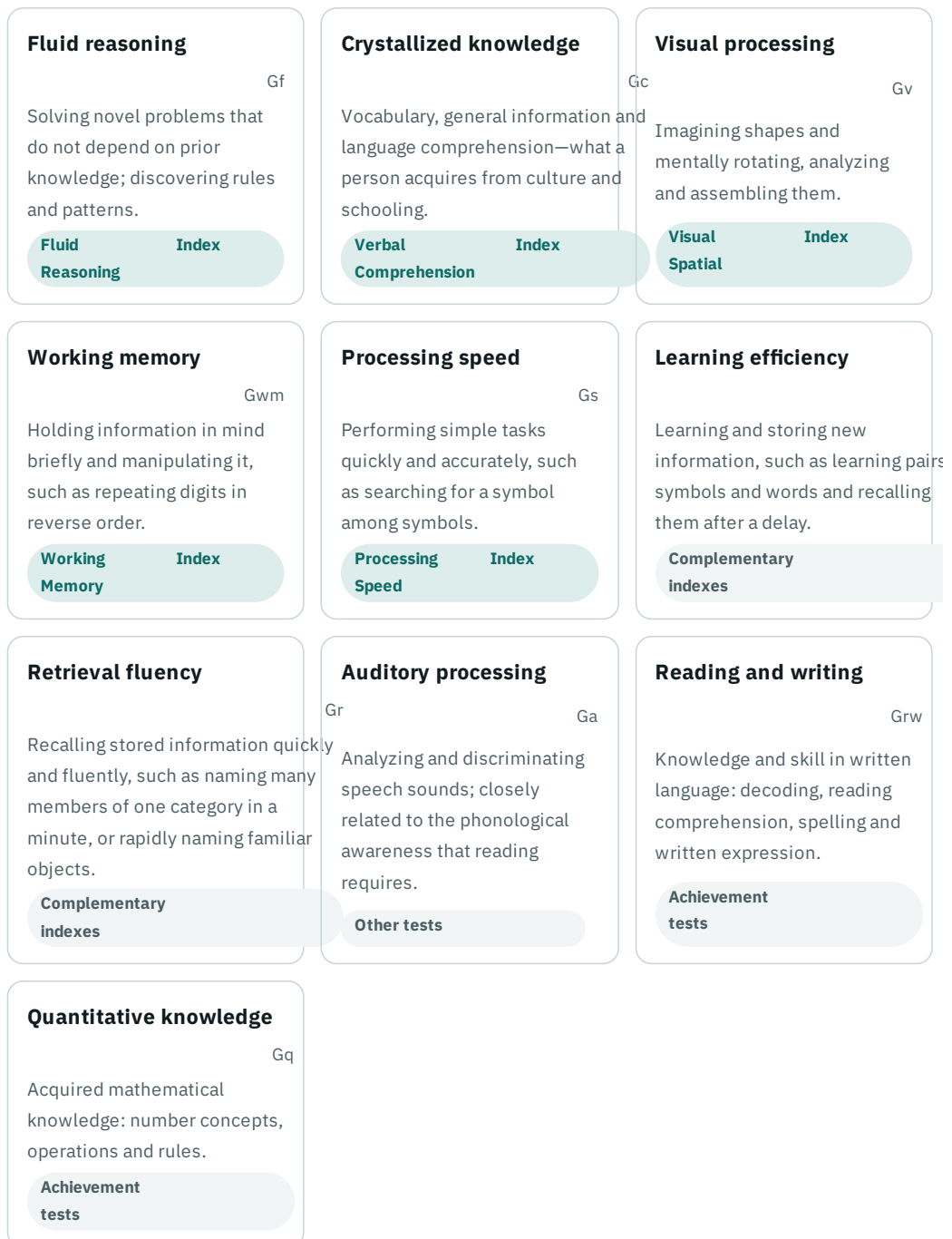


Figure 3.2: Ten broad abilities in CHC theory as updated by Schneider and McGrew (Schneider & McGrew, 2018), with the corresponding indexes of the fifth edition of the Wechsler scale for children (Wechsler Intelligence Scale for Children, WISC-V). The last five abilities are not measured by its primary

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 (Schrang 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):

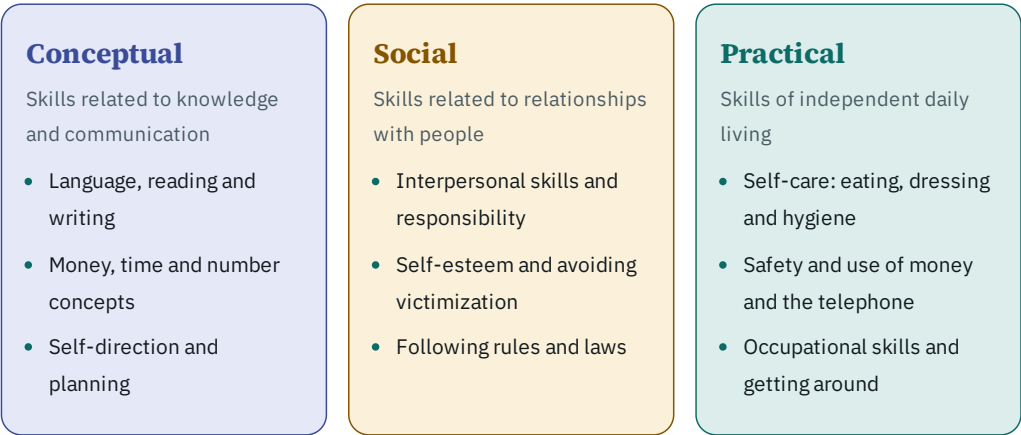


Figure 3.3: The three domains of adaptive behavior, with examples of the skills in each.

💡 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).

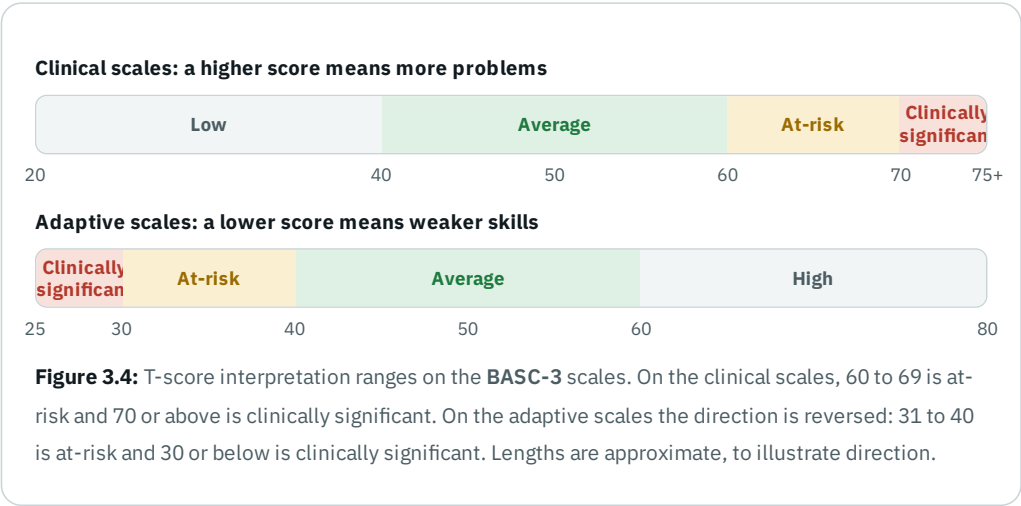
Table 3.7: Common rating scales and the state of their Arabic versions.

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:



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).

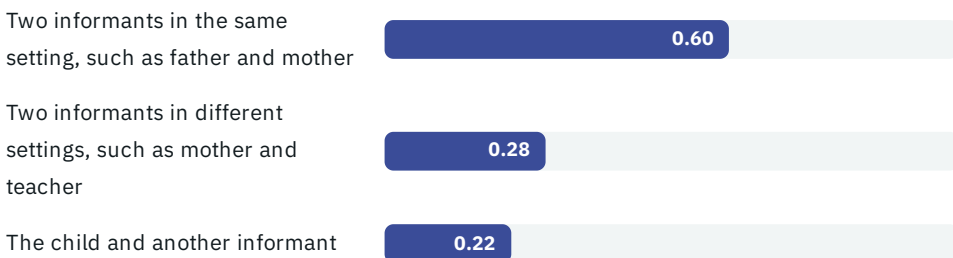


Figure 3.5: Mean correlations between different informants' ratings of a child's behavioral and emotional problems, from Achenbach and colleagues (Achenbach et al., 1987). The more the settings differ, the lower the agreement.

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:



□ 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.

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/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>

- 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>

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.

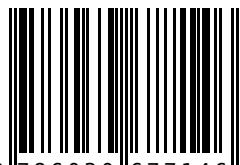
The book brings together two fields that the literature usually treats separately: disabilities and disorders, and giftedness and twice-exceptionality. It reads international standards and the American experience in light of published Saudi regulations, and describes the status of the Arabic version 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