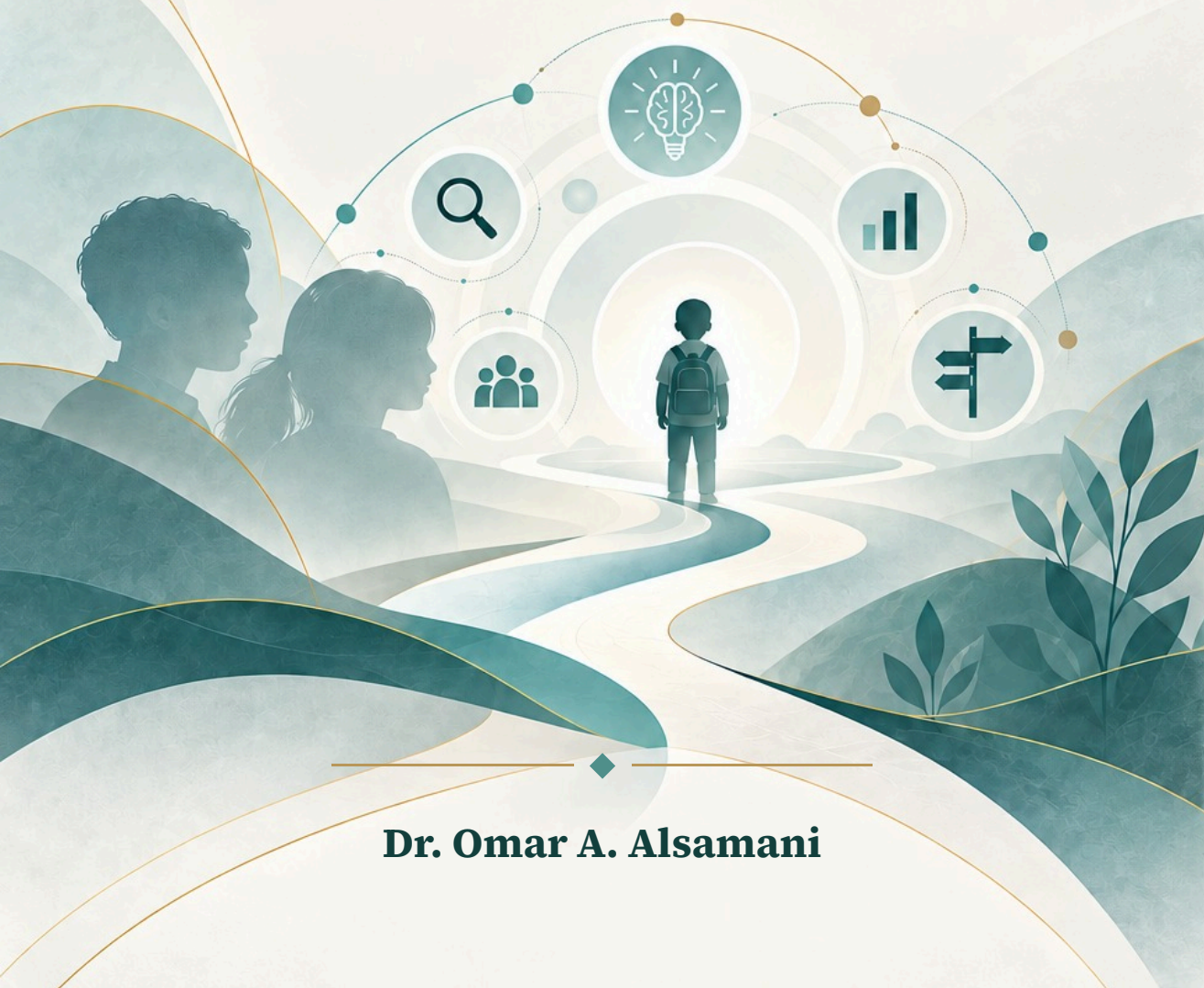


Diagnosis and Identification of Exceptional Learners

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

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.

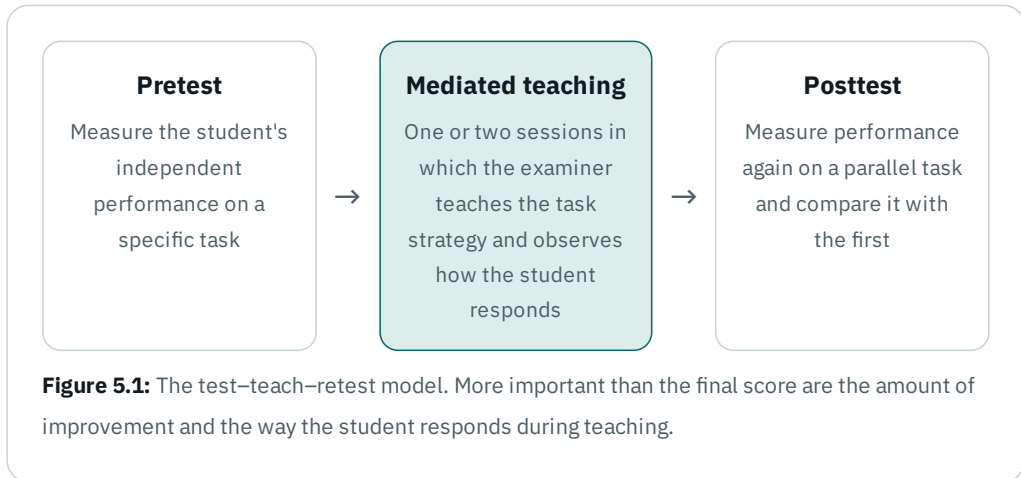
5.2 Dynamic assessment

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



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).

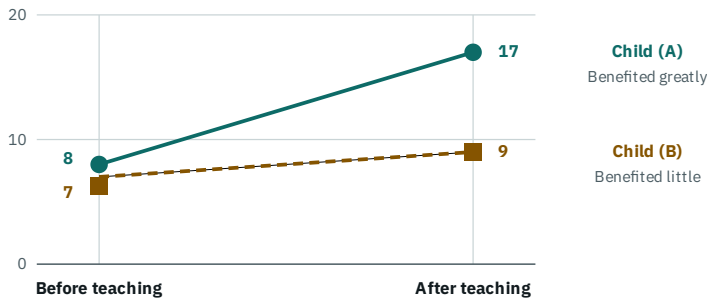


Figure 5.2: A hypothetical example of two children who both started with low scores (8 and 7 out of 20). Child (A), shown by the solid line and circular markers, jumped to 17 after two sessions, which suggests that the low initial score stemmed from a lack of opportunity. Child (B), shown by the dashed 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.

- 1 "Think about the problem again."
- 2 "What is the question asking for? Read it again."
- 3 "Do you remember a similar problem we solved a little while ago?"
- 4 "Look at these two numbers. How are they related?"
- 5 Show the first step of the solution and ask the student to complete the rest.
- 6 Show the full solution with an explanation, then give a parallel problem.

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.

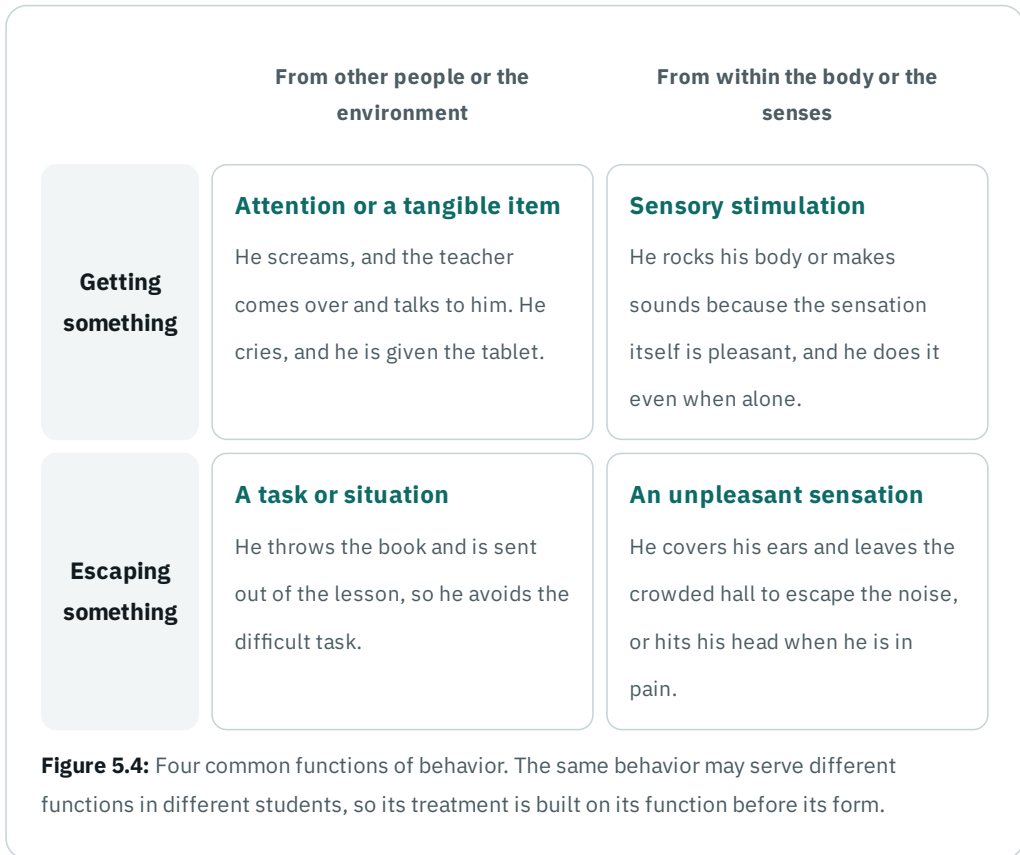
5.3 Functional behavioral assessment

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.



❓ **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

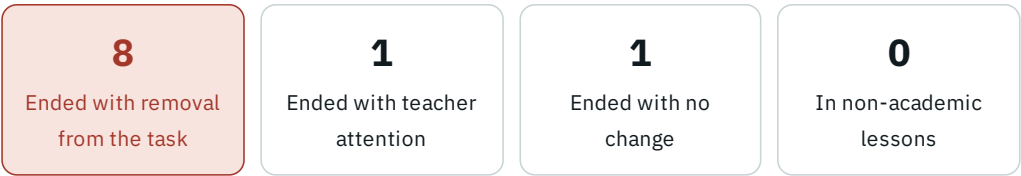
Level	What it provides	Its limits	Who usually carries it out
Descriptive (antecedent-behavior-consequence recording)	An actual pattern from the natural setting	Correlational: shows what accompanies the behavior, not what causes it	The teacher after training, and the specialist
Experimental (functional 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



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).

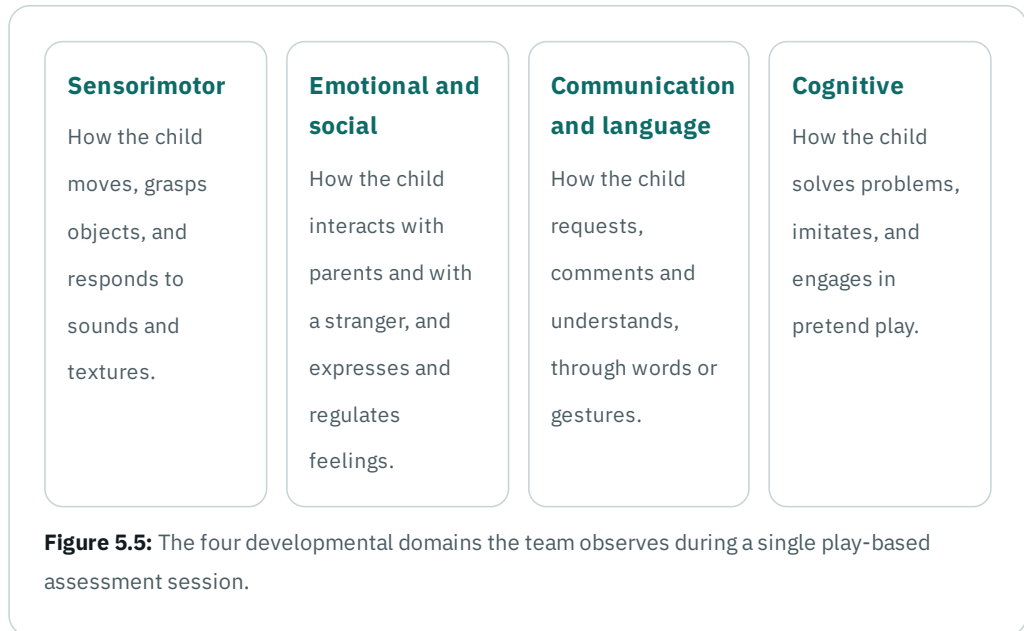
The practical upshot: authentic assessment enters the giftedness decision as one piece of evidence among others, not as the sole gateway. It is strengthened by training scorers on the rubric with worked examples, by having each product scored by two independent scorers whose agreement is calculated, 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):



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 (Allothman 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>

Allothman, 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-spel.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., & Ghaemmahami, 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.

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.

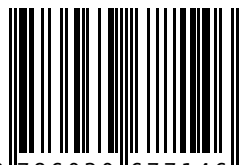
The book brings together two fields that the literature usually treats separately: disabilities and disorders, and giftedness and twice-exceptionality. It reads international standards and the American experience in light of published Saudi regulations, and describes the status of the Arabic version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