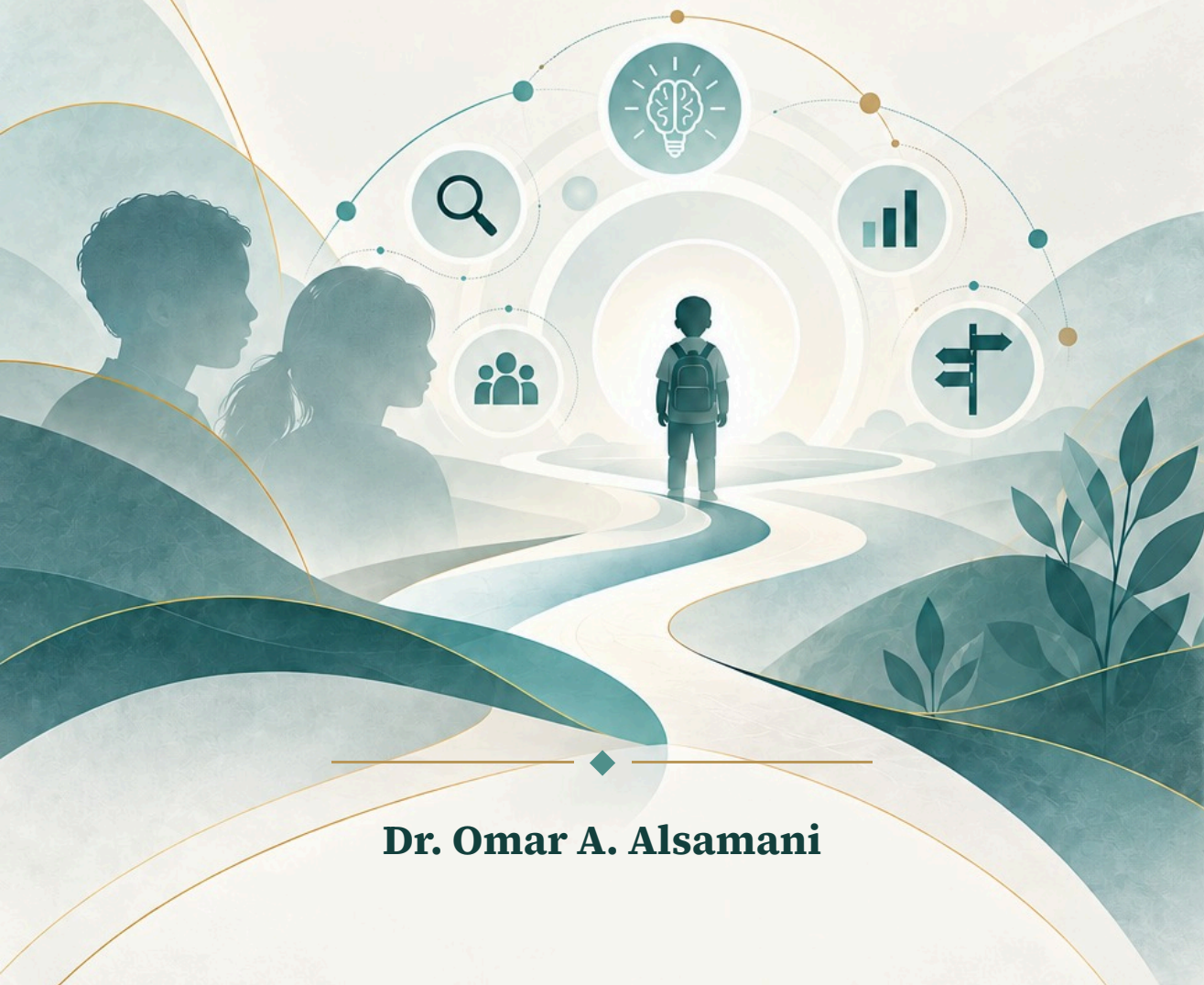


Diagnosis and Identification of Exceptional Learners

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

4

Chapter 4 Informal methods

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

Scene from a second-grade classroom

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

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

Chapter objectives

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

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

Introduction

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

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

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

4.1 Why informal methods?

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

Table 4.1: Formal standardized methods and informal methods.

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

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

4.2 Structured observation

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

Step one: the operational definition

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

Non-operational definition

“Mishal is aggressive with his classmates.”

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

Operational definition

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

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

Step two: choosing a recording method

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

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

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

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

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

Interval recording: three methods, different results

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

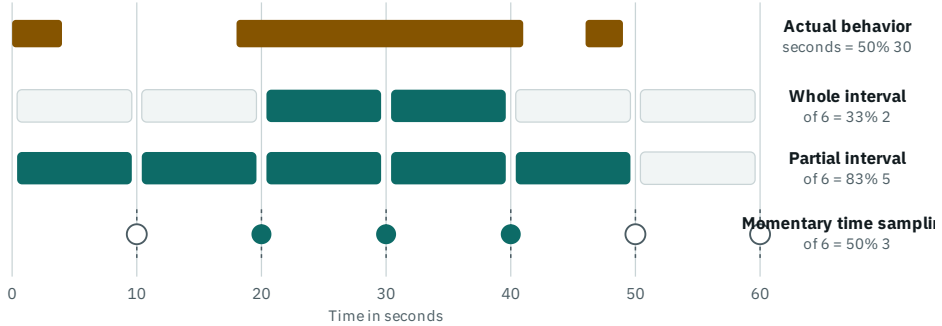


Figure 4.1: One minute divided into six intervals. The top bar shows when the student was actually out of his seat. In **whole-interval** recording, an interval is scored only if the behavior lasts throughout it, so the method tends to underestimate the behavior. In **partial-interval** recording, an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

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

Step three: checking interobserver agreement

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

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

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

Worked example

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

⚠ For the specialist: when percentage agreement misleads

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

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

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

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

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

Calculating kappa

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

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

4.3 The interview

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

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

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

 For the teacher: a short family interview guide

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

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

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

4.4 Curriculum-based measurement

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

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

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

The best-known measures

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

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

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

Scene from practice: reading fluency

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

Graphing data and making decisions

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

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

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



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

⚠ For the specialist: what supports the decision rules?

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

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

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

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

What does the graph tell us about diagnosis?

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

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

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

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

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

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

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

4.5 Error analysis

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

Hypothetical worksheet from a third-grade student

¹ $\begin{array}{r} 52 \\ - 37 \\ \hline 25 \end{array}$	¹ $\begin{array}{r} 63 \\ - 28 \\ \hline 45 \end{array}$	¹ $\begin{array}{r} 84 \\ - 21 \\ \hline 63 \end{array}$
مدرسة	مدرسه	
الشمس	الشمس → الشمس	
	اشمس	
سأل	سئل	

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

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

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

Oral reading errors: from “meaning” to decoding

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

❓ Misconception: “As long as the error didn't change the meaning, it's fine”

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

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

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

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

Error type	Example (written → read)	What it may indicate
Short-vowel error	“كَتَبَ” → “كُتِبَ” (kataba → kutiba)	Weakness in reading short-vowel marks, or reading the word's outline without its vowels
Adding or omitting a long vowel	“سَفَرٌ” → “سَافِرٌ” (safar → sāfara)	Inattention to long-vowel letters, and confusion between morphological patterns
Substituting a letter for one that looks similar	“بِنْتُ” → “بَيْتٌ” (bint, “girl” → bayt, “house”)	Weakness in discriminating diacritical dots and in analyzing the word into its letters
Substituting a letter for one that sounds similar	“صَارَ” → “سَارَ” (šāra → sāra)	Weakness in phonological discrimination between similar sounds
Substituting one form for another from the same root	“الطَّلَابُ” → “الطَّالِبُ” (al-ṭullāb, “the students” → al-ṭālīb, “the student”)	Reading the root and guessing the structure, as in confusing the singular with the broken plural

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

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

4.6 Portfolios and rubrics

The portfolio

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

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

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

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

Rubrics

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

Table 4.6: A brief analytic rubric for written expression.

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

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

Limits of portfolios and performance assessment in high-stakes decisions

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

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

4.7

Limits of informal methods

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

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

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

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

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

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

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

4.8 Forms for use

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

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

Form A: Oral reading fluency record sheet

Curriculum-based measurement in reading

Student or code: _____ Grade: _____

Date: _____ Examiner: _____

☐ Progress monitoring, ☐ Partially
 Purpose: Screening Baseline week: Passages: Vowelized Unvowelized vowelized

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

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

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

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

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

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

Form B: Interval recording form

Ten minutes, six intervals per minute

Student or code: _____ Date: _____

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

Observer: _____ ☐ First observer ☐ Second observer

Target behavior: _____

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

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

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

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

Behavior intervals: _____ of _____ = _____ %

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

Agreements: _____ Disagreements: _____

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

Kappa (optional): _____

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

Form C: Antecedent–behavior–consequence recording form

Descriptive recording in natural settings

Student or code: _____ Observer: _____

Period: from _____ to _____

Target behavior and its operational definition: _____

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

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

After several days of recording:

Most frequent antecedent: _____

Most frequent consequence: _____

Tentative hypothesis about the function of the behavior: _____

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

Form D: Evidence-supported referral form

Organized according to the RIOT/ICEL framework

Student: _____ Grade: _____
Age: _____ Referring teacher: _____
Date of referral: _____

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

Student's strengths: _____

What was tried before referral:

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

Baseline (median): _____ Goal: _____

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

☐ Progress-monitoring graph attached ☐ Passages or worksheets are at a single level and have been tried out
☐ Intervention was implemented as planned (documentation available)

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

Attendance and punctuality this year: _____

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

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

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

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

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

Referring teacher Special education teacher School principal

Conclusion

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

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

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

4.9

Summary, key terms, and activities**Chapter summary**

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

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

Glossary of terms

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

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

Review questions

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

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

Brief answers to the calculation questions: Question 3: Percentage agreement = $22 \div 30 \times 100 \approx 73\%$, which is below 80% and therefore not acceptable until the definition has been reviewed. The agreement expected by chance = $(0.50 \times 0.43) + (0.50 \times 0.57) =$ approximately 0.50, so kappa = $(0.73 - 0.50) \div (1 - 0.50) \approx 0.47$, that is, “moderate” agreement by the common classification.

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

Case for application

Bandar, 13 years old, first grade of middle school

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

Table 4.11: Bandar's informal data.

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

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

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

Learning activities

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

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

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

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

Chapter references

- Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). <https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf>
- Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. <https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf>
- Abu-Hamour, B. (2014a). A pilot study for standardizing curriculum-based measurement oral reading fluency (CBM ORF) in Arabic. *Journal of the International Association of Special Education*, 15(1), 16–26.
- Abu-Hamour, B. (2014b). Using Arabic word identification fluency to monitor first-grade reading progress. *Dyslexia*, 20(2), 167–174. <https://doi.org/10.1002/dys.1472>
- Almulla, A. A., & Mahama, I. (2026). Elementary teachers' perspectives on RTI programmes in supporting students with learning disabilities in Saudi Arabia. *Humanities and Social Sciences Communications*, 13(1), Article 974. <https://doi.org/10.1057/s41599-026-07376-9>
- Altheiabi, A. S., Bagadood, N. H., & Sulaimani, M. F. (2025). Perceptions of learning difficulty teachers regarding the barriers to applying the response to the intervention model in Saudi Arabia. *Australian Journal of Learning Difficulties*, 30(1), 75–92. <https://doi.org/10.1080/19404158.2025.2473425>
- Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, 51(1), 1–18. <https://doi.org/10.1016/j.jsp.2012.09.004>
- Ashlock, R. B. (2010). *Error patterns in computation: Using error patterns to help each student learn* (10th ed.). Allyn & Bacon.
- Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*, 18(3), 296–307.

- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Christ, T. J. (2008). Best practices in problem analysis. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 159–176). National Association of School Psychologists.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly*, 5(1), 9–30.
- Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms* (Technical Report No. 1702). Behavioral Research and Teaching, University of Oregon.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement* (2nd ed.). Guilford.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5–16. <https://doi.org/10.1111/j.1745-3992.1994.tb00443.x>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Mahfouz, M. A. S., & Mohamed, A. H. H. (2023). Using word reading fluency curriculum-based measurements to monitor students' reading progress in grade 2. *Education Sciences*, 13(2), Article 217. <https://doi.org/10.3390/educsci13020217>
- Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60–63.
- Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted Lebanese students: Is DISCOVER effective? *Journal for the Education of the Gifted*, 33(2), 275–295. <https://doi.org/10.1177/016235320903300206>

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

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.

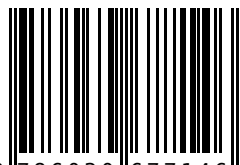
The book brings together two fields that the literature usually treats separately: disabilities and disorders, and giftedness and twice-exceptionality. It reads international standards and the American experience in light of published Saudi regulations, and describes the status of the Arabic version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