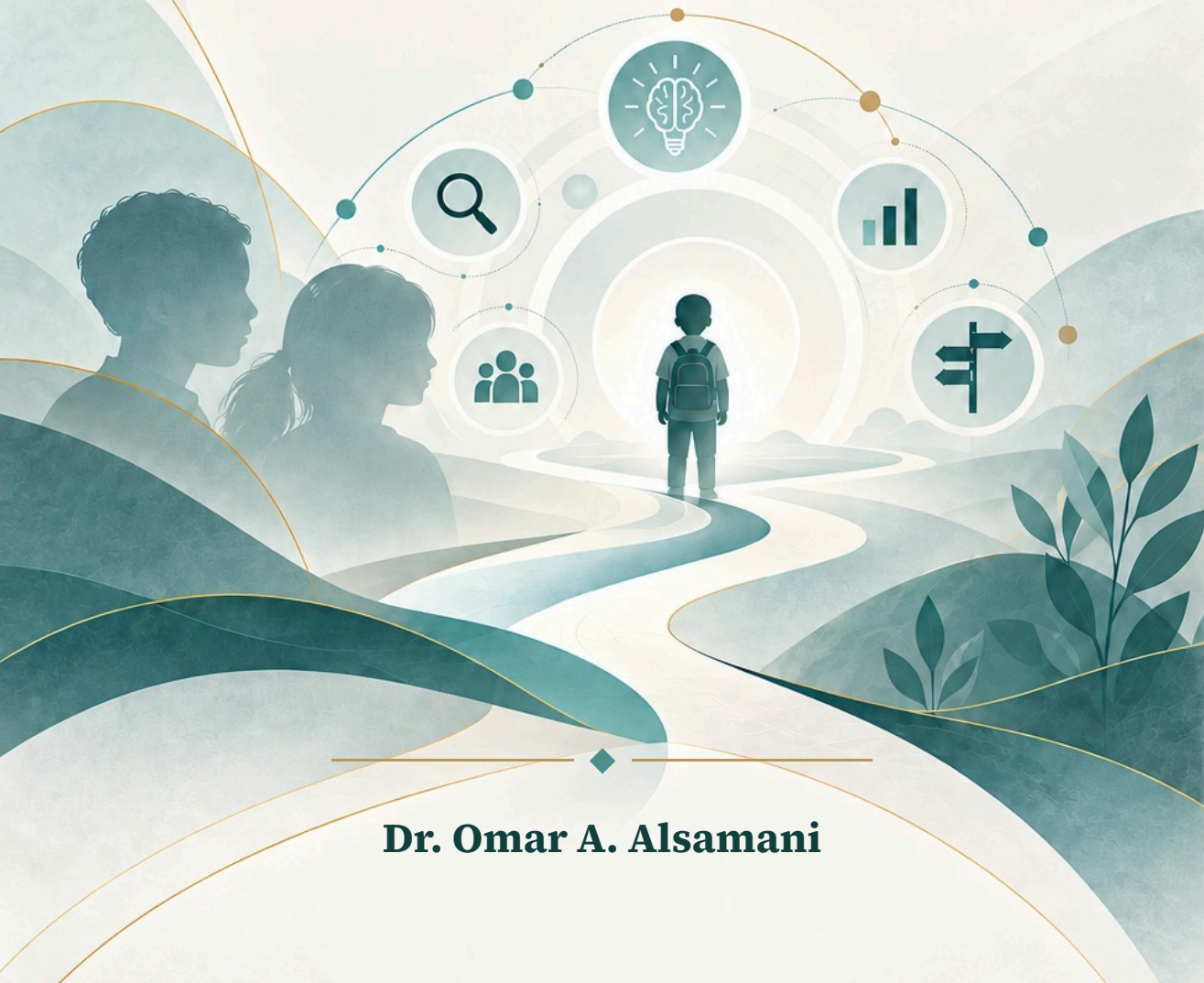


Diagnosis and Identification of Exceptional Learners

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

14

Chapter 14

Fairness in diagnosis

A student can be wronged in diagnosis without anyone intending it: a hard-of-hearing child comes close to an autism spectrum disorder diagnosis because no one checked his hearing; another student is nearly diagnosed with intellectual disability because he was tested in a language he has not mastered; and the youngest child in the class has activity that is perfectly expected for his age read as a disorder. In this chapter we move from the individual case to a broader question: how can we tell whether the diagnostic system as a whole is fair, and whether its fairness reaches beyond the single student?

Scene from an end-of-year report

A special education supervisor in an education directorate reviewed the referral records of the schools in her charge. Three things caught her eye. Boys were referred for behavior roughly four times as often as girls. Students who were not native speakers of Arabic made up twice as large a share of learning disabilities referrals as of the school population. And the schools in remote villages had not referred a single student to the gifted program in three years.

She asked herself: do these numbers reflect real differences among students, or do they reflect the way we ourselves work?

The fair answer is that both may be true at once. Our aim in this chapter is to show how to tell the two apart.

Chapter objectives

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

- 1 Calculate the risk index and risk ratio from a school's data, and explain the limits of the risk ratio when numbers are small.
- 2 Identify, in a given scenario, the point at which bias entered the diagnostic pathway, and propose a step to limit it.
- 3 Calculate the representation index and the minimum fair representation in a gifted program, and propose two steps to raise the representation of an underrepresented group.
- 4 Distinguish, in the description of a student learning in a second language, between conversational fluency and academic proficiency, and specify the evidence needed before referral.
- 5 Critique the interpretation of a score pattern with the Culture-Language Interpretive Matrix in light of studies that have examined its validity.
- 6 Name the cognitive bias at work in a given assessment scenario, and propose a strategy to reduce it.
- 7 Judge whether the evidence is sufficient for a fair decision in a new case, and justify that judgment.

Introduction

A student can suffer injustice in diagnosis without anyone meaning it to happen, when his hearing goes unchecked, when he is tested in a language he has not mastered, or when activity expected for his age is read as a disorder. Yet fairness is not tested in the individual case alone. Patterns of referral and diagnosis across the schools of an entire education directorate, like those the supervisor noticed in the opening scene, may reveal that certain groups are referred more often, or less often, than others. Judging the fairness of a diagnostic system requires distinguishing disparities in rates from actual bias, knowing the points through which bias enters, and knowing the tools that expose it.

We open the chapter by defining what fairness in diagnosis means (Section 14.1), then turn to disproportionality, the ways it is measured and what happens when numbers are small (Section 14.2), and the points through which bias enters, along with its sources within the instrument (Section 14.3). We then address fairness in gifted identification, the representation index and the fair minimum (Section 14.4), followed by language and culture, the Culture-Language Interpretive Matrix and what studies say about its validity (Section 14.5), and assessor biases (Section 14.6). We offer a fairness checklist (Section 14.7) and close with three cases and a school audit (Section 14.8).

14.1 What is fairness in diagnosis?

Fairness in diagnosis does not mean that every group should appear in special education or gifted programs in equal proportions. Some differences are real: attention-deficit/hyperactivity disorder, for example, genuinely is more prevalent among boys. Fairness means, rather, that every student has the same chance of being identified if he needs services, and that no student is misdiagnosed because of something unrelated to need: his language, his sex, his family's poverty, the location of his school, or the expectations of the person assessing him. The "Standards for Educational and Psychological Testing"

devote their third chapter to fairness in testing, defining it as the extent to which score interpretations are equally valid for all individuals and groups to whom the instrument is administered. They approach it from four angles: fair treatment of examinees during administration, freedom of measurement from bias, access to the construct the test measures without obstacles unrelated to it, such as the language of the instructions or the format of presentation, and the validity of the interpretation of an individual's own score (AERA et al., 2014).

Diagnostic error runs in two directions, and each carries a cost:

Overdiagnosis

A student who does not need services is diagnosed: he may be separated from his peers, expectations of him drop, he carries a label that does not fit him, and resources are diverted from those who need them.

Underdiagnosis

A student who needs services goes unidentified: he struggles for years without understanding or support, is blamed for what lies beyond his control, and loses time when intervention would have done the most good.

14.2 Disproportionality and how it is measured

Disproportionality is present when the share of a group in a diagnostic category or program differs markedly from its expected share. It is measured with two simple indicators:

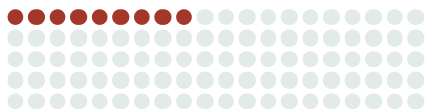
- **Risk index:** the proportion of a group's members who have been diagnosed or referred, out of the group's total number.
- **Risk ratio:** the risk index of one group divided by the risk index of a comparison group. A value of 1 indicates proportionate representation; a value above 1 indicates overrepresentation, and a value below 1 underrepresentation. U.S. regulations use the risk ratio to determine

significant disproportionality at the state and district levels (U.S. Department of Education, 2024).

Worked example: behavior referrals in one school

- 1 The school has 400 boys and 380 girls. Thirty-six boys and 8 girls were referred for behavior.
- 2 Risk index for boys: $36 \div 400 = 9\%$; for girls: $8 \div 380 \approx 2.1\%$.
- 3 Risk ratio for boys compared with girls: $9 \div 2.1 \approx 4.3$.
- 4 In other words, a boy is more than four times as likely as a girl to be referred for behavior.
- 5 **Caution:** the figure of 4.3 rests on only 8 girls. Had two more girls been referred (10 girls), the risk ratio would be about 3.4; had two fewer been referred (6 girls), it would be about 5.7. A change of two referrals moves the ratio anywhere between “three times” and “nearly six times.”

Boys: 9 in every 100



Risk index 9%

Girls: 2 in every 100



Risk index about 2%

Figure 14.1: The risk index made visible. Each grid has one hundred dots, each dot representing a student; the shaded dots are those referred for behavior. The risk ratio here is about 4.3.

△ The number raises a question; it does not answer it

A high risk ratio does not, on its own, prove bias. Behavior problems may genuinely be more common among boys. The right questions are these: is the difference larger than studies of real differences would lead us to expect? Are boys referred for behaviors that are overlooked in girls? And are internalizing problems in girls, such as anxiety, withdrawal and sadness, missed because they do not disrupt the classroom the way overt behavior does? The literature holds a genuine debate on this point: one large study found that some groups were less likely to be diagnosed than expected once other factors were controlled (Morgan et al., 2015), while other researchers criticized it on the grounds that the method of control itself may conceal bias (Skiba et al., 2016). The lesson is that the numbers mark the beginning of an inquiry and an invitation to complete it.

When numbers are small

The risk ratio is a fraction whose denominator is the risk index of the comparison group. When only a few members of that group have been referred, the denominator becomes fragile, and a change of one or two cases is enough to make the result jump, as the caution in the previous example showed. When the number is zero, the division cannot be performed at all. For this reason, a risk ratio in a small school, or in a rare category, should be read only alongside the raw counts on which it rests. More useful than a single year's ratio is to pool data across several years and look at the trend over time.

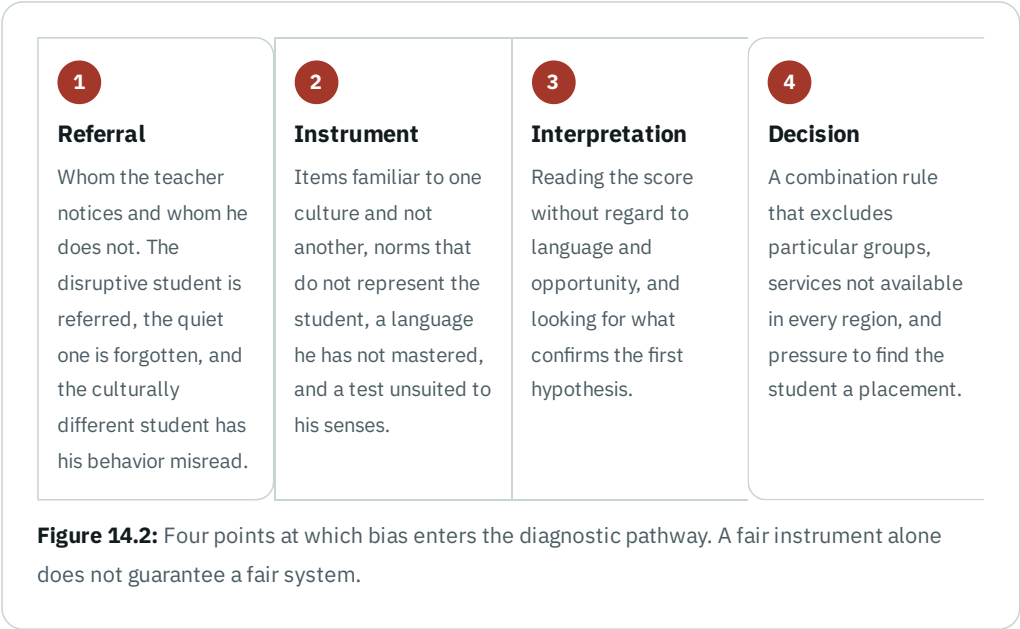
✓ **For the specialist: how do U.S. regulations deal with the instability of the ratio?**

Section §300.647 of the implementing regulations of the Individuals with Disabilities Education Act (IDEA) governs how “significant disproportionality” is calculated, and it addresses the problem of small numbers with three tools. **The first** is a minimum cell size, meaning the number of group members diagnosed or referred, together with a minimum group size (**Minimum n-size**), meaning the group’s total membership. The ratio is not calculated unless both thresholds are met, and the regulation regards a threshold of no more than 10 for the cell and 30 for the group as presumptively reasonable. **The second** is the alternate risk ratio: when the comparison group in the district falls below these thresholds, the group’s risk index is compared with the risk index of all other students across the state. **The third** is that a state may decline to identify significant disproportionality unless the ratio exceeds its threshold for several consecutive years, up to three (U.S. Department of Education, 2024). We know of no published Saudi text that sets out a method or a threshold for monitoring disproportionality, but the same logic holds for any education directorate reviewing its own data.

14.3

Where does bias enter?

Many people assume that bias in diagnosis comes from the test. But the test is only one of four points along the way, and bias can enter at any of them:



Sources of bias in the instrument

Table 14.1: Sources of bias in the instrument and ways of detecting it.

Type	What it means	Example
Content bias	Items that assume knowledge or experience not equally available to all students	A question about snow or a train put to a child who has never seen either, or about vocabulary specific to one region
Norm bias	A norming sample that does not represent the student	Comparing a Saudi student with American norms, or a rural student with urban-only norms
Language bias	A test that measures ability through a language the student has not mastered	A student whose first language is not Arabic is tested on vocabulary and verbal comprehension in Arabic, and his verbal ability score drops because of his weak Arabic

Type	What it means	Example
Predictive bias	The same score predicts different performance in two groups. It is examined statistically by comparing the regression lines in the two groups: the slope of each and its intercept with the axis	A test that predicts the success of urban students in a program better than it predicts the success of village students
Situational bias	Testing conditions affect one group more than others	An unfamiliar examiner in a formal setting with a child unused to dealing with strangers
Differential item functioning and measurement noninvariance	An item on which two groups perform differently even though they are equal in the trait being measured, or an instrument that does not measure the same construct in the same way in the two groups	An item on a verbal reasoning scale that is harder for students from one region because it contains a word common in another dialect, even though students from the two regions with equal total scores should perform equally on it

The last row of the table differs from the others: rather than describing a type of bias, it describes **two standard methods for detecting it**. The first is **differential item functioning** (**Differential item functioning, DIF**). The idea is to compare students from two groups who are *equal in ability*, rather than comparing the two groups as wholes. If an item is harder for one group even though its members are equal in ability, then for that group the item is measuring something else as well, such as cultural knowledge or a local word. A difference between the two groups’ mean scores on an item does not by itself indicate bias, since it may reflect a real difference in the trait. The second is **measurement invariance** (**Measurement invariance**), which asks at the level of the whole instrument: is the factor structure the same in both groups? Are the items related to the factors with the same strength? Does the same score

signify the same level of the trait? If invariance does not hold, comparing the two groups' means is not warranted, nor is applying a single cutoff score to both (AERA et al., 2014). The practical question for the practitioner is whether the manual of the instrument he uses reports the results of these analyses for the groups with whom he uses it, such as boys and girls, different regions, and speakers of languages other than Arabic. If it does not, fairness remains an untested assumption.

14.4 Fairness in gifted identification

The examples in the preceding sections all concern overrepresentation, where a group appears in special education more often than it should. In gifted programs, the best-known problem runs the other way: **underrepresentation**. Students from low-income families, from remote schools, and those learning in a second language are usually fewer in these programs than their share of the student body. Nor does the matter stop at the test. Groups differ in the proportions who reach advanced levels of performance in the first place, differences known as **excellence gaps** (**Excellence gaps**), and these gaps reflect, in large part, differences in opportunity more than in potential (Plucker & Peters, 2016). A program that admits only those who have already reached advanced performance inherits these gaps, then widens them by serving those who were already ahead.

The representation index and the fair minimum

The risk index and risk ratio apply here as well, but the gifted education literature often relies on two indicators that are easier to read:

- **Representation index (Representation index):** the group's share among program students divided by its share among students at the same grade level. A result of 1 means proportionate representation; anything below it means underrepresentation.

- **Equity formula (Equity formula)**, proposed by Donna Ford: it accepts that the two percentages may differ somewhat, since perfect correspondence is unrealistic, but it places a ceiling on the difference. The minimum acceptable representation of a group in the program is its share among students minus 20% of that share, that is, **its share among students $\times 0.8$** . Anything below that counts as underrepresentation that calls for review (Ford, 2013).

The 20% margin is a practical convention. It spares the program from chasing small differences while obliging it to account for large ones.

Worked example: a gifted program in an education directorate

- 1 Grades three through six in an education directorate enroll 4,000 students, 1,200 of them (30%) in village schools. The gifted program has 200 students, 20 of them (10%) from village schools.
- 2 Representation index for village students: $10 \div 30 \approx 0.33$. Their representation in the program is thus about one third of their share of the student body.
- 3 Fair minimum under Ford's formula: $30\% \times 0.8 = 24\%$, that is, at least 48 of the 200 students. The shortfall is 28 students.
- 4 In risk ratio terms: the risk index for village students is $20 \div 1,200 \approx 1.7\%$, for urban students $180 \div 2,800 \approx 6.4\%$, and the risk ratio ≈ 0.26 . A village student, in other words, is about one quarter as likely as an urban student to be identified as gifted.
- 5 **The next question** is "Where are they falling out?" before "How many should we add?" Did nomination or registration reach them? Were they given the test? Did they pass it and then fail to enroll because the program was too far away? Each point calls for a different remedy.

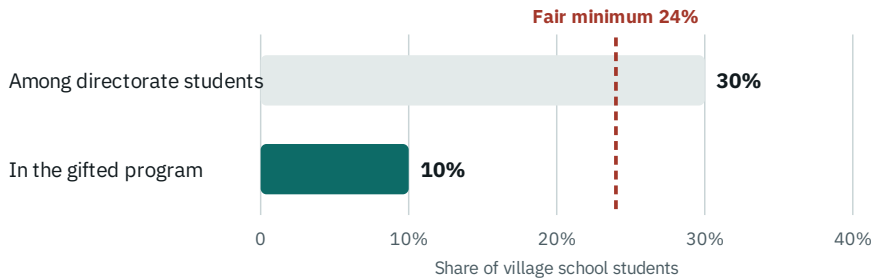


Figure 14.3: Share of village school students among the directorate’s students and in the gifted program, in the worked example. The dashed line marks the fair minimum under Ford’s formula ($30\% \times 0.8 = 24\%$). Anything below it, like the second bar, is underrepresentation that calls for review.

Where does the shortfall come from?

Studies indicate that the shortfall accumulates at points close to those in Figure 14.2, with one more point added after the decision: the service itself.

- Nomination:** when the door to a program is teacher nomination or a family request, who walks through it reflects the teacher’s judgment and the family’s awareness as much as the student’s ability. A large U.S. study found that the likelihood of a student being identified as gifted varied with the background of his teacher, even when students’ test achievement was equal (Grissom & Redding, 2016). In Saudi Arabia, Alfaiz examined the accuracy of parent and teacher nominations in identifying gifted students (Alfaiz, 2026). When a large U.S. school district moved from nomination to **universal screening (Universal screening)**, administering an initial instrument to every student in the grade, the representation of students from low-income families and from minority groups rose markedly, because many of them had never been nominated in the first place (Card & Giuliano, 2016).
- Instrument and norms:** comparing a student in a less advantaged school with a national sample measures his opportunities as much as his ability.

Hence the proposal of **local norms** (**Local norms**), which compare the student with peers in his own school or in schools with similar circumstances; they have been found to increase the representation of underrepresented groups (Peters et al., 2019).

- **Decision rule:** a single high cutoff score on a single instrument, or a requirement to meet every criterion at once, narrows the door for students with a clear strength in one area and a weakness in another.
- **Service:** a program offered only in urban centers, or one that requires transportation the family does not have, loses students who have already been identified.

No single measure solves the problem. Peters, reviewing what these measures have achieved, argues that the effect of each on representation is limited, and that much of the gap forms years before identification, in opportunities to learn themselves, so that better identification alone cannot close it (Peters, 2022). Fair identification is necessary, then, but not sufficient.

📍 **In the Saudi context: fairness issues**

Four issues in the Saudi system bear on fairness in diagnosis and identification. For each, we present what has been published, and we state plainly what we could not find in print.

1. **Voluntary registration and fees.** The Mawhiba Multiple Cognitive Abilities Test, the instrument of the National Program for Gifted Identification, is administered to students whose families register them through the National Center for Assessment website, with a parental contribution of 200 riyals from which supported students are exempt; a student may sit the test once at each level (King Abdulaziz and His Companions Foundation for Giftedness and Creativity, n.d. [Arabic]). In the absence of universal screening, the first gate here is the family's decision: a student whose family does not know about the program, or does not see it as a priority, never reaches the test, however able he may be. The exemption eases the barrier of cost, but the barrier of awareness remains. We found no published data on registration rates by region, school or family income.
2. **Different admission rules in boys' and girls' programs.** The guide to the School Program for the Care of Gifted Girls admits the "top 5% of the outcomes" of the national project (Ministry of Education, 1437 AH [Arabic]). The gifted classes guide, by contrast, nominates students classified in the project and fills the remaining places using criteria of which at least two must be met, one of them a cognitive ability test with a score of no less than 120 (Ministry of Education, n.d. [Arabic]). The "top 5%" is calculated from test-takers alone, so it depends on which girls registered and which did not. And because boys and girls are educated separately, those who nominate and assess in girls' schools are mostly women, and those who nominate in boys' schools mostly

men, so any difference in nomination standards between the two groups turns into a sex difference in representation. We found no published data comparing referral rates to special education, or identification rates in gifted programs, across the two systems. On the other hand, survey studies have been published that offer a basis for comparison, among them a study of 5,495 boys and girls in Al-Zulfi Governorate that estimated the prevalence of dyslexia at about 6.5% among boys and 3.8% among girls (Aldakhil et al., 2023). Because figures of this kind do not pass through referral, they can serve as a benchmark against which actual referral rates in schools are compared.

3. **Regional variation.** The national test standardizes the instrument across regions, but we do not know how everything around it is distributed. The Special Education Organizational Guide requires that assessment instruments be administered and interpreted by “qualified specialists” (Ministry of Education, 1436 AH [Arabic]), yet we found no published data on their availability in cities and villages, nor on the distribution of gifted classes and programs. The peer-reviewed studies published on national test scores come from particular regions, such as a study of students in Al-Ahsa (Almulhim et al., 2026), and so do not allow regions to be compared.
4. **Learners of Arabic as a second language and non-Saudi students.** The Regulations exclude from learning disabilities difficulties attributable to “learning conditions or family care” (Ministry of Education [Al-Ma'arif], 1422 AH [Arabic]). This exclusion is what protects a student learning in a second language from misdiagnosis, but the published documents do not explain how it is verified. Nor did we find a published text specifying the eligibility of non-Saudi students for the National Program for Gifted Identification, or any data on their participation in it.

A question for reflection: if your education directorate wanted to know whether the Mawhiba test reaches students in all its schools equally, which three pieces of data would you request first, and from which body?

14.5 Language and culture

The language of conversation and the language of learning

One of the most common sources of misdiagnosis is confusing second language learning with disability. Jim Cummins distinguished between two levels of language proficiency: **basic interpersonal communicative skills** (**Basic Interpersonal Communicative Skills, BICS**), the language of everyday talk and the playground, in which a second language learner reaches fluency close to that of peers in **about two years**; and **cognitive academic language proficiency** (**Cognitive Academic Language Proficiency, CALP**), the language of textbooks, abstract concepts and lengthy instructions, which takes roughly **five to seven years** to reach the level of native speakers (Cummins, 1981).

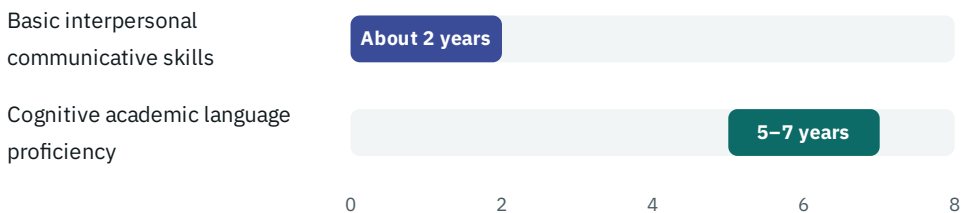


Figure 14.4: Approximate time, in years, that a second language learner needs to approach native-speaking peers, after Cummins (Cummins, 1981). Conversational fluency comes in about two years; academic proficiency between the fifth and seventh year. A student who speaks Arabic fluently at recess after two years may need several more years before he can learn in it as effectively as his classmates.

The danger is that a teacher sees a student chatting fluently in Arabic with classmates, then stumbling in reading and science, and concludes that he has a learning disability, when in fact he has not yet attained academic language proficiency. Saudi schools enroll many students from expatriate families whose first language is not Arabic, and these students most closely resemble those on whom the framework was built.

These figures have limits that readers should know before using them. **First**, they are drawn from data on immigrant children learning English in Canada, and they do not carry over unchanged to another language or another education system. **Second**, it is often suggested that an Arab student who speaks a colloquial dialect encounters in Modern Standard Arabic something akin to a second academic language, since it departs from his dialect in vocabulary and structure. The analogy is reasonable, but it is an inference awaiting research, and it does not justify applying Cummins's timeframes to that student. **Third**, the distinction between the two levels has itself drawn theoretical criticism: among other things, that it reduces language proficiency to two poles, and that it can be used to explain a student's difficulties as a deficit in him rather than in his instruction. Cummins has addressed these criticisms, presenting the distinction as a tool for understanding the difference between fluency and learning, with approximate timeframes (Cummins, 2008). The practical conclusion nonetheless stands: **fluency in conversation is not evidence of command of the language of learning.**

—

How can we tell whether the low scores of a linguistically diverse student reflect his ability or his language? Flanagan and colleagues (Flanagan et al., 2013) proposed a method called the **Culture-Language Interpretive Matrix** (C-LIM). The idea is to classify subtests along two dimensions: how much **cultural content** they carry, and how much **language** they demand. The student's scores are then placed in the cells of the matrix, and the scores in each cell are compared with an **expected range** for a student whose linguistic

and cultural difference from the norming sample is similar to his. A highly different student, such as a recent arrival, is expected to show steeper drops in the high-loading cells than a student who has spent years in schools taught in that language. If scores decline steadily as cultural and linguistic loading increases, *and fall within the range expected for such a student*, language and culture likely explain the decline. If they fall below the expected range, however, language alone is not a sufficient explanation, and another cause should be sought. The criterion, then, is the size of the graded decline relative to students with similar exposure to the language, not merely its presence (Flanagan et al., 2013).

	Low linguistic demand	Moderate	High
Low cultural loading	Figural reasoning 98	Rapid visual attention 94	Auditory verbal memory 89
Moderate	Block design 95	Picture story sequencing 87	Verbal reasoning 81
High	Pictorial information 86	Social situations comprehension 79	Vocabulary 74

Figure 14.5: A simplified hypothetical example of the Culture-Language Interpretive Matrix. The numbers are standard scores with a mean of 100. Scores decline steadily from the top cell to the bottom, that is, as cultural and linguistic loading increases, from 98 to 74, and cell shading deepens in the same direction. This pattern is *consistent with* an effect of language and culture, but it does not prove one on its own (see the next paragraph). The task names here are generic and their classification is illustrative, independent of how particular subtests are classified in published tables.

What studies say about the validity of the matrix

The matrix spread through practice faster than it was examined. And when it was examined, the results were troubling. Kranzler and colleagues administered a cognitive battery to students learning English as a second language and did not find the graded decline the matrix predicts with the regularity it assumes (Kranzler et al., 2010). Styck and Watkins studied the files of students referred for evaluation, some learning English as a second language and some native speakers, and found that the matrix patterns distinguished poorly between the two groups (Styck & Watkins, 2013). This is to be expected once we recall that a native speaker with a language difficulty will also show declining scores as linguistic demands increase. The graded pattern may mean “second language,” or it may mean “genuine language weakness,” and the matrix alone cannot tell them apart.

Add to this that its classifications are built on English-language tests, and we know of no published classification of Arabic tests by cultural and linguistic loading. The matrix, then, is **a tool for organizing one's thinking before judgment**. Its core idea remains valid in any language: **look at the pattern of scores as well as their level**. But the pattern is a hypothesis that needs independent evidence to confirm or refute it, of which three kinds matter most: the student's proficiency in his first language, measured directly rather than through the family's report; his level in the second academic language, compared with peers who have had the same length of exposure; and his response to direct, focused instruction monitored over several weeks. Among the tools for the last kind of evidence is dynamic assessment, which measures how much the student benefits from brief instruction provided during the assessment itself (see Chapter 5), complemented by asking the family about the student's language and educational history (Rhodes et al., 2005).

🔗 Common misconceptions

- **“A nonverbal test is fair to all students.”** A nonverbal test reduces the influence of language, but it does not remove the influence of experience and schooling, and it does not measure the verbal and quantitative abilities a student needs in most academic programs. For this reason it is not sufficient on its own for identifying gifted students among second language learners; it is more useful to compare the student with others who share similar opportunities, using verbal, quantitative and nonverbal measures together (Lohman, 2005).
- **“Translating a test makes it valid for the student.”** Translation can change the difficulty and meaning of items, and it does not carry over the norms. A translated instrument needs evidence of measurement invariance before its scores can be compared with the original.
- **“Equal proportions mean fairness, and unequal ones mean injustice.”** The standard is equal chances of identification and equally valid interpretation, not identical proportions. A difference raises a question to be investigated.

14.6 Assessor biases

Assessors are human, and the mind sometimes falls back on mental shortcuts in judgment that lead to predictable errors. These errors have long been studied in medicine, where many diagnostic errors were found to stem from the way clinicians think more than from gaps in their knowledge (Croskerry, 2003). The most prominent in our field are the following:

Table 14.2: Common assessor biases and how to reduce them.

Bias	How it shows	Example	What reduces it
Anchoring	The first idea governs everything that follows	A teacher notices that a student does not turn when his name is called, assumes autism from the outset, and never asks for a hearing check	Writing at least three alternative hypotheses before the assessment
Confirmation bias	Seeking what supports the hypothesis and overlooking what contradicts it	Dismissing a parent’s rating because it differs from what the team expected, instead of asking why the two differ	Deliberately searching for evidence that contradicts the hypothesis
Diagnostic overshadowing	Attributing every symptom to a known diagnosis	Anxiety in a student with intellectual disability is explained away as “part of his disability”	Asking: does the known diagnosis actually explain this symptom?
Halo effect, and its negative counterpart, the horn effect	One salient trait colors judgment of everything else: a positive one raises it, a negative one lowers it	Nominating well-behaved high achievers for giftedness while overlooking a gifted girl who asks endless questions and keeps untidy notebooks	Observation tools with specific items, more than one rater, and judging work without knowing whose it is
Availability bias	Judging by whatever comes most readily to mind	A teacher who attended a course on attention-deficit/hyperactivity disorder begins to see it in every highly active student	Returning to the written criteria instead of relying on impressions

 For the teacher: five minutes before you refer or nominate

You are the first point in Figure 14.2, which is why bias either begins with you or ends with you. Before you sign a referral or nomination form, ask yourself:

- What specific behavior did I see? How often? In which lesson? Write it down as it happened, before any interpretation.
- If another student in the class had done the same thing, would I refer him? And if this student were a girl rather than a boy, or the reverse, or came from a different family?
- Who in my class has never once crossed my mind? The quiet student, the newcomer, the one who has not yet mastered Arabic may need referral or nomination more than those who caught my attention.
- Did I give the student a chance to show what he can do, through an open-ended task or in a form other than writing, before judging him?

14.7 A fairness checklist

These are questions for the team to ask before any diagnostic decision, arranged according to the points in Figure 14.2:



Referral

Did the student receive sound instruction and adequate support before referral? And was this documented?



Referral

Was the student's behavior compared with that of peers of the same age and background?



Instrument

Were hearing and vision checked?

☐ **Instrument**

Was the instrument normed on a sample resembling the student in language and environment? And when was it normed?

☐ **Instrument**

Was it administered in the student's stronger language or through his means of communication? Did he need a nonverbal instrument?

☐ **Instrument**

Does the instrument's manual present evidence that its items are free of differential functioning, and of measurement invariance, for the group to which the student belongs?

☐ **Interpretation**

Was the pattern of scores considered alongside their level? Were confidence intervals taken into account?

☐ **Interpretation**

Did the team write alternative hypotheses and look for evidence against its first hypothesis?

☐ **Interpretation**

Was information gathered from home and school, and was any conflict between them understood?

☐ **Decision**

Is the decision rule made public, and does it avoid excluding any particular group without justification?

☐ **Decision**

At the program level: are representation rates reviewed every year using raw counts and compared with the fair minimum?

☐ **Decision**

Did the family take part in the decision and understand it?

14.8

Three cases and a school audit

Case one: Zain, age 10

Most likely a second language effect

The story: Zain arrived with his family two and a half years ago; his first language is Urdu. He speaks Arabic fluently with his classmates but is struggling in reading, science and social studies. His teacher referred him on suspicion of a learning disability.

Table 14.3: Evidence in Zain's case.

Source	Result	What it means
Length of time learning Arabic	Two and a half years	Conversational language mastered; academic language most likely not yet fully developed
Direct measure of academic Arabic	He read a passage from his grade's science textbook, answered comprehension questions and defined subject vocabulary. His performance fell clearly below that of his Arabic-speaking classmates and close to that of two classmates who had been learning Arabic for the same length of time	His difficulties are within what is expected for someone with his exposure to Arabic
Ability test in Arabic	Scores declining steadily with linguistic and cultural loading, as in the pattern of Figure 14.5	A pattern consistent with a language effect, but one that cannot by itself distinguish a second language from a language weakness

Source	Result	What it means
Family interview with an interpreter	He reads Urdu well and was a high achiever at his previous school	Important information, but it rests on the family's report alone
Reading sample in Urdu, with help from a teacher at the school who speaks it fluently	He read a grade-level passage accurately and fluently and summarized what he read	Direct, though unstandardized, evidence of sound reading in his first language
Dynamic assessment in Arabic reading	Rapid improvement after two sessions of training in decoding strategies	An initial indicator of learning potential that needs further corroboration

Decision: Zain is not diagnosed with a learning disability at this point. A genuine learning disability usually shows itself in both languages, and the direct evidence from the reading sample, beyond what the family reports, indicates that he reads soundly in Urdu. He will receive eight weeks of focused support in academic Arabic, including explicit teaching of subject vocabulary, with weekly measurement of his reading progress compared against that of peers who have been learning Arabic for the same length of time. If he grows as they do, the language explanation is confirmed. If his growth stalls despite the support, his case will be revisited through a more comprehensive evaluation, and the possibility of a learning disability remains open.

What evidence could change the decision? If, after the eight weeks of focused support, his growth in Arabic reading remained clearly slower than that of peers with the same length of exposure, or if a closer assessment of his Urdu reading revealed difficulties with

accuracy or fluency, the possibility of a learning disability would return to the fore, and the comprehensive evaluation would be completed.

Case two: Waad, age 9 Horn effect

The story: On a gifted nomination scale, Waad's teacher rated her academic ability low, writing: "Her handwriting is very poor, and her notebook is a mess." The program coordinator, however, asked two other teachers to rate Waad's answers on a problem-solving task after they had been retyped on a computer, without telling them whose work it was. Both placed her among the top 10% of girls in the class.

Decision: Waad was referred for gifted assessment. Her case illustrates the negative face of the halo effect, known as the **horn effect** (**Horn effect**): one conspicuous negative trait, here her handwriting, colored the judgment of her ability as a whole. Rating work without knowing whose it is offers a simple way to curb it.

An assessment of her fine motor skills and handwriting was also requested, and for a reason that goes beyond the handwriting itself. The wide gap between her problem-solving ability and her written performance may be a sign of **twice-exceptionality**: giftedness combined with a writing difficulty (dysgraphia). A student who carries both is exposed to two errors at once: her difficulty may mask her giftedness, so that she is never nominated, as nearly happened here, and her giftedness may mask her difficulty, so that she is never supported. Both sides are therefore assessed together, rather than stopping at her admission to the program (see Chapter 13).

What evidence could change the decision? If the individual gifted assessment returned average results despite the two teachers' ratings, her performance on a single task would be treated as a relative strength that needs further evidence before nomination. And if the assessment of her fine motor skills and handwriting revealed a clear deficit, the question of twice-exceptionality would become part of the decision itself.

Case three: Ohoud, age 14, ninth grade

First, open the gate for her

The story: Ohoud attends a small intermediate school in a remote village, with eleven girls in her class. The nearest gifted classes for girls are in a city about an hour and a half away by car. She has never been registered for the national test at any level, and no teacher has ever nominated her. As part of the review the supervisor launched in the opening scene, her education directorate ran open-ended problem-solving activities with girls in the village schools for four weeks. During that time, the mathematics teacher noticed that Ohoud was solving problems that had not yet been taught, and proposing more than one method for some of them.

Table 14.4: Evidence in Ohoud's case.

Source	Result	What it means
National test registration record	Not registered at the first or second level.	She never reached the first gate at all, and the reason
	Her mother said she	lies in the family's
	had never heard of	awareness of the test, quite
	the test	apart from the student's ability

Source	Result	What it means
School grades	Top of her class in mathematics and science since seventh grade	A limited indicator: ranking first among eleven girls says little about where she would stand in a larger group
Teacher's observation after the open-ended activities	Original solutions to unfamiliar problems, and questions that go beyond the lesson	Behavior that would not have surfaced in a classroom that offers no such tasks
Brief group ability test administered by the directorate to all ninth-grade girls in the sector's village schools	In the top 3% of 240 girls in these schools	A clear strength relative to those who share her opportunities; this is a screening indicator that comes before any eligibility decision

Decision: No judgment is made yet as to whether Ohoud is gifted or not; the first task is to open the gate she never reached. With the school's help, she will be registered for the Mawhiba test this year, and her eligibility for a fee exemption will be checked. She is at the third level, where she is entitled to only one attempt, so that attempt should not be lost to missing information or to a lack of transport to the test center. In the meantime, she will receive mathematics enrichment within her own school, because her need exists whatever the test result. If she qualifies, the team will work out with her family how she can reach the service, since identification without access to services changes nothing.

What the case teaches: Ohoud was wronged because no one had measured her in the first place; neither a test nor an assessor ever reached her. This kind of injustice appears in no individual assessment report. Only a system-level review, like the one the supervisor began, can bring it to light.

What evidence could change the decision? If her Mawhiba test result fell below the nomination threshold after a registration carried out under suitable conditions, the mathematics enrichment in her school would continue, and her need for a gifted program would be reconsidered. And if it turned out that her family could not reach the service location, the decision would shift to finding an alternative that reaches her in her village.

☐ **School audit: back to the supervisor**

The supervisor in the opening scene has three numbers in front of her.

Here are the steps we suggest:

1. **Behavior referrals:** Calculate the risk ratio for each school, record the raw counts next to it, and pool three years of data for small schools before drawing conclusions. Then review a sample of referrals: which behaviors are cited? Are girls referred for the same behaviors? And look for girls with internalizing problems who were never referred.
2. **Learning disabilities referrals of non-native speakers of Arabic:** Review how long these students have been learning Arabic; whether their proficiency in their first language was measured directly, alongside asking the family; whether their academic Arabic was measured against peers with the same length of exposure; and whether the referral was preceded by language support whose effect was monitored.
3. **Village schools missing from gifted programs:** Calculate the representation index and the fair minimum, then trace where students drop out: Were they registered for the national test? Did screening or nomination reach them? Were local norms used? Were they given opportunities to display their abilities? Ohoud's case is an example of what such tracing can uncover.

Conclusion

The central argument of this chapter is that fairness means an equal chance of being identified for those who need it, freedom from misdiagnosis for those who do not, and valid score interpretation for every group, which runs deeper than equal percentages. It follows that a risk ratio should be treated as a

question that calls for verification, read alongside raw counts and across several years, and that bias should be sought at its four points of entry: referral, instrument, interpretation and decision. It follows, too, that conversational language should not be confused with academic language, that the pattern of scores across cultural and linguistic loading should be read as a hypothesis requiring independent evidence, and that assessor biases should be curbed through alternative hypotheses, multiple raters and a return to the criteria.

Some questions in this chapter nonetheless remain open. Cummins's time frames are estimates drawn from a different context; the Culture-Language Interpretive Matrix discriminates poorly between second language learning and language impairment; and part of the representation gap in gifted programs arises before identification itself, so identification tools alone cannot close it. In the Saudi context, optional registration for the national test makes the family's decision the first gate, and no data are published that would allow comparison across regions or between boys' and girls' programs, so system-level monitoring of fairness remains limited by whatever data are available.

14.9 Summary, key terms and activities

Chapter summary

- Fairness means an equal chance of being identified for those who need it, freedom from misdiagnosis for those who do not, and valid score interpretation for every group; it runs deeper than equal percentages.
- The risk ratio is a simple tool for detecting disproportionality, but it raises a question rather than proving bias, and it becomes unstable when numbers are small, so it is read alongside raw counts and across several years.

- Bias enters at four points: referral, instrument, interpretation and decision. Instrument bias is detected statistically through analyses of differential item functioning and measurement invariance.
- In gifted programs, underrepresentation is the dominant problem. The representation index and the fair minimum detect it, and universal screening and local norms reduce it, but part of the gap arises before identification.
- In the Saudi context, optional registration for the national test makes the family's decision the first gate, and no data are published that would allow comparison across regions or between boys' and girls' programs.
- Conversational language is acquired much faster than academic language, and confusing the two is a common source of misdiagnosis. Cummins's time frames are estimates drawn from a different context.
- The pattern of scores across cultural and linguistic loading is more telling than their level, but the Culture-Language Interpretive Matrix discriminates poorly between second language learning and language impairment, so the pattern it shows is a hypothesis that needs independent evidence.
- Assessor biases are to be expected; they are reduced by alternative hypotheses, rating work without knowing whose it is, multiple raters and a return to the criteria.

Glossary of terms

Disproportionality	A clear difference between a group’s share of a category or program and its expected share.
Risk index	The number of students in a group who have been diagnosed, divided by the group’s total number.
Risk ratio	The risk index of one group divided by the risk index of a comparison group.
Representation index	A group’s share among program students divided by its share among all students; a value of 1 indicates proportionate representation.
Equity formula	A rule that sets the minimum acceptable representation of a group in a gifted program at its share among students multiplied by 0.8.
Excellence gaps	Differences between groups in the proportions of students who reach advanced levels of performance.
Universal screening	Administering an initial instrument to all students in a grade or school level, rather than waiting for nomination or a family request.
Differential item functioning Differential item functioning (DIF)	A difference in two groups’ performance on an item among individuals of equal standing on the ability being measured, indicating that the item measures something else besides that ability.

Measurement invariance	The property of a scale measuring the same construct in the same way in two groups, so that their scores can be compared and a single cut-off score applied to both.
Basic Interpersonal Communicative Skills (BICS)	The language of everyday conversation, in which a second language learner reaches fluency close to that of peers in about two years.
Cognitive Academic Language Proficiency (CALP)	The language of learning and abstract concepts, which takes roughly five to seven years to develop.
Culture-Language Interpretive Matrix (C-LIM)	A method for interpreting the pattern of scores according to their cultural loading and linguistic demands, by comparing them with the range expected for students whose exposure to the language resembles the student's own.
Confirmation bias	Seeking evidence that supports a hypothesis while overlooking evidence that contradicts it.
Anchoring	Allowing the first idea that occurs to the assessor to govern the interpretation of everything that follows.
Diagnostic overshadowing	Attributing a new symptom to a diagnosis the student already has, so that the possibility of another problem is overlooked.
Halo effect	The influence of a single salient trait, positive or negative, on judgments across many unrelated areas.

Horn effect	The negative face of the halo effect: a salient negative trait that lowers judgments of areas unrelated to it.
Internalizing problems	Emotional problems the student directs inward, such as anxiety, withdrawal and sadness; they are easily overlooked because they do not disrupt the classroom.

Review questions

1. A school has 300 students who are non-native speakers of Arabic and 900 who are native speakers. Twenty-one students in the first group and 27 in the second have been diagnosed with a learning disability. Calculate the risk ratio for the non-native speakers and interpret it. Then suppose another school has 30 non-native speakers of Arabic, two of whom have been diagnosed: what should you be wary of before comparing its ratio with that of the first school? (Objective 1)
2. Why does a high risk ratio not, on its own, prove bias? What further questions does it raise? (Objective 1)
3. A gifted program has 150 students, nine of whom are learning Arabic as a second language; these learners make up 15% of students at that school level. Calculate the representation index and the fair minimum using Ford’s formula, determine the shortfall if any, and propose two measures to close it, each addressing a different point of entry. (Objective 3)
4. For each of the following situations, identify the point at which bias enters and propose a measure to reduce it: (a) A teacher refers only students who disrupt the class. (b) A program requires a score of 130 on a single verbal test for admission. (c) A team interpreted the low scores of a recently arrived student as low ability without asking about his language. (Objective 2)

5. A study found that an item on a reasoning scale was harder, on average, for girls than for boys. Is this enough to conclude that the item is biased? What comparison do you need before reaching a judgment? (Objective 2)
6. After two years, a student speaks Arabic fluently but struggles with reading. What three pieces of evidence do you need before referral? And why is his fluency not sufficient evidence that he has mastered the language of instruction? (Objective 4)
7. The Culture-Language Interpretive Matrix showed a pattern of gradual decline for a referred student. One team member said: "The pattern proves the problem is linguistic, so there is no need for further assessment." Evaluate this view in light of studies that have examined the validity of the matrix, and identify two independent pieces of evidence the team needs. (Objective 5)
8. Choose one of the three cases in this chapter, or a case you have encountered in your own work, and name a cognitive bias that could have led to a wrong decision in it; then propose a strategy to reduce it. (Objective 6)
9. Design one procedure for your school that reduces the halo or horn effect in gifted nomination, and explain how you would know it had worked. (Objectives 3 and 6)
10. **Case for application:** Adam is eleven years old and in fifth grade. His first language is Somali, and he enrolled in a Saudi school four years ago. He was referred on suspicion of intellectual disability after obtaining a full-scale score of 68 (confidence interval 63–74) on an Arabic intelligence scale, with a verbal score of 62 and a nonverbal score of 84. His teacher says he speaks Arabic fluently but does not understand long instructions and struggles in every subject. His family says he helps out in the family shop, counts money and gives change quickly, and interprets for his parents at official appointments. His adaptive behavior has not been assessed, and his hearing has not been checked in three years. (a) At which of the four points might bias enter his case? (b) For eligibility under intellectual disability, the Special Education Organizational Guide requires an IQ of no more than 75 on a Wechsler scale or its equivalent, along with deficits in at least two

areas of adaptive behavior (Ministry of Education, 1436 AH [Arabic]). Is the available evidence sufficient to judge that both conditions are met? Explain. (c) What evidence is missing? Rank it by priority. (d) What is your decision now: diagnosis, no diagnosis, or deferring judgment? Justify your decision, and specify what is needed to finalize it. (Objectives 2, 4 and 7)

Brief answers to the computational questions: Question 1: The risk index for non-native speakers of Arabic is $21 \div 300 = 7\%$, and for native speakers $27 \div 900 = 3\%$, so the risk ratio is $7 \div 3 \approx 2.33$; that is, a non-native speaker is more than twice as likely as a native speaker to be diagnosed with a learning disability. In the second school the risk index is $2 \div 30 \approx 6.7\%$, close to the first, but it rests on two cases: a change of a single case moves it between roughly 3.3% and 10%, so it can be compared only alongside the raw counts and several years of data. Question 3: Their share in the program is $9 \div 150 = 6\%$, the representation index is $6 \div 15 = 0.4$, and the fair minimum is $15\% \times 0.8 = 12\%$, or 18 students, a shortfall of nine students.

Table 14.5: Alignment of chapter objectives with review questions and learning activities.

Objective	Assessed by
1. Calculate the risk index and risk ratio and interpret their limits	Questions 1 and 2; Activity 1
2. Identify the point at which bias enters and propose ways to reduce it	Questions 4 and 5; part (a) of Question 10; Activity 2
3. Calculate the representation index and the fair minimum and propose measures	Questions 3 and 9; Activity 1
4. Distinguish conversational fluency from academic proficiency	Question 6; part (c) of Question 10
5. Critique interpretations of the Culture-Language Interpretive Matrix	Question 7

Objective	Assessed by
6. Name the cognitive bias and propose ways to reduce it	Questions 8 and 9; Activity 3
7. Judge whether the evidence suffices for a fair decision	Question 10 (case for application)

Learning activities

Any activity that uses students' data or work is carried out with the consent of the family and the school administration, without any information that could identify the student, and under the supervision of a qualified specialist. Its results are not used in any decision concerning the student and are reported only through official channels.

- 1. Calculating from school data:** With the administration's permission, and using aggregate counts only, without names, obtain the numbers of referrals to special education, or of students enrolled in a gifted program, in one school or a group of schools, broken down by gender, grade or school type. Calculate the risk indices and ratios, or the representation index and the fair minimum; flag every cell with fewer than ten students; and write three questions the numbers raise. If no data are available to you, apply the steps to the figures in the opening scene and the two worked examples. *Product: a table of calculations and three questions on a single page. Time: about two hours.*
- 2. Item analysis:** Choose ten items from an Arabized scale whose test booklet you have access to, without reproducing them in your report, and identify those that may carry cultural or dialectal content unfamiliar to a student in your region. Then search the scale's manual: does it report an analysis of differential item functioning or of measurement invariance across groups? *Product: a table of the items, numbered and classified, and a paragraph on the fairness evidence in the manual. Time: about an hour and a half.*

- 3. Rating without knowing whose work it is:** Collect five pieces of student work, rewrite them in a uniform hand, and ask a colleague to rate them before and after their authors are revealed; then compare. If no student work is available to you, write two answers of equal content to a single task yourself, one in neat handwriting and the other in poor handwriting, and ask colleagues to rate them. *Product: a table showing both ratings for each piece of work, and a paragraph interpreting the differences. Time: about an hour.*

Chapter references

- King Abdulaziz and His Companions Foundation for Giftedness and Creativity. (n.d.). *[National Program for Gifted Identification]* [In Arabic]. Retrieved September 29, 2026, from <https://www.mawhiba.sa/en/services/the-national-program-for-gifted-identification/>
- Ministry of Education. (n.d.). *[Guide to gifted classes]* [In Arabic]. General Administration for the Gifted. Retrieved September 29, 2026, from <https://www.moe.gov.sa/ar/aboutus/aboutministry/RPRLibrary/دليل%20فصول%20الموهوبين.pdf>
- Ministry of Education. (1436 AH). *[Special education organizational guide]* [In Arabic] (1st ed., academic year 1436–1437 AH). <https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20التنظيمي%20للتربية%20الخاصة.pdf>
- Ministry of Education. (1437 AH). *Organizational and procedural guide for the school program for gifted girls* [In Arabic]. General Directorate for Gifted Girls. [https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20\(1\).pdf](https://www.moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/الدليل%20الإجرائي%20والتنظيمي%20لبرنامج%20رعاية%20الموهوبات%20المدرسي%20(1).pdf)
- Ministry of Education [Al-Ma'arif]. (1422 AH). *[Regulations for Special Education Institutes and Programs]* [In Arabic]. <https://moe.gov.sa/en/aboutus/aboutministry/RPRLibrary/القواعد%20التنظيمية%20لمعاهد%20وبرامج%20التربية%20الخاصة%201422%20هـ.pdf>
- Aldakhil, A. F., Ebrahim, M. T., & Gadelrab, H. F. (2023). Diagnostic survey of dyslexia and accompanying behavioral indicators in primary school students in Saudi Arabia. *Research in Developmental Disabilities*, 134, Article 104424. <https://doi.org/10.1016/j.ridd.2023.104424>

- Alfaiz, F. S. (2026). The accuracy of parent and teacher nominations in gifted identification in Saudi Arabia. *Cogent Education*, 13(1), Article 2654922. <https://doi.org/10.1080/2331186X.2026.2654922>
- Almulhim, N., Aljughaiman, A. M., & Alfaiz, F. S. (2026). Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis. *Journal of Intelligence*, 14(8), Article 166. <https://doi.org/10.3390/jintelligence14080166>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences*, 113(48), 13678–13683.
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, 78(8), 775–780.
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 2(2), 132–149.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 2. Literacy* (2nd ed., pp. 71–83). Springer.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Wiley.
- Ford, D. Y. (2013). *Recruiting and retaining culturally different students in gifted education*. Prufrock Press.
- Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, 2(1).
- Kranzler, J. H., Flores, C. G., & Coady, M. (2010). Examination of the cross-battery approach for the cognitive assessment of children and youth from diverse linguistic and cultural backgrounds. *School Psychology Review*, 39(3), 431–446.
- Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, 49(2), 111–138.

- Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, 44(5), 278–292.
- Peters, S. J. (2022). The challenges of achieving equity within public school gifted and talented programs. *Gifted Child Quarterly*, 66(2), 82–94.
- Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open*, 5(2).
- Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.
- Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. Guilford.
- Skiba, R. J., Artiles, A. J., Kozleski, E. B., Losen, D. J., & Harry, B. (2016). Risks and consequences of oversimplifying educational inequities: A response to Morgan et al. (2015). *Educational Researcher*, 45(3), 221–225.
- Styck, K. M., & Watkins, M. W. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition among referred students. *School Psychology Review*, 42(4), 367–382.
- U.S. Department of Education. (2024). *Significant disproportionality*, 34 C.F.R. §§ 300.646–300.647. Retrieved September 29, 2026, from <https://www.ecfr.gov/current/title-34/subtitle-B/chapter-III/part-300>

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.

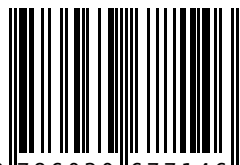
The book brings together two fields that the literature usually treats separately: disabilities and disorders, and giftedness and twice-exceptionality. It reads international standards and the American experience in light of published Saudi regulations, and describes the status of the Arabic version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