

Assessment Literacy for Educators in a Hurry

by W. James Popham

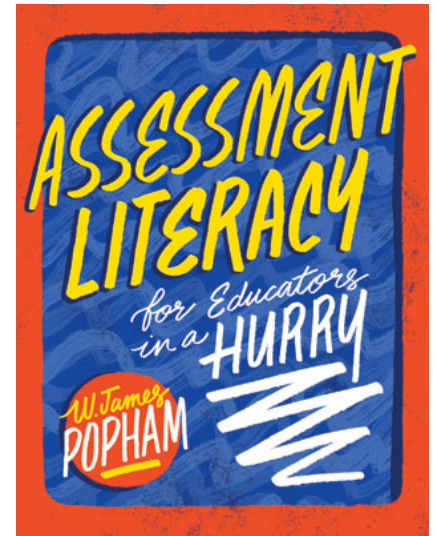
An ASCD Study Guide for Assessment Literacy for Educators in a Hurry

This Study Guide is designed to deepen your comprehension of *Assessment Literacy for Educators in a Hurry*, a book by W. James Popham published by ASCD in August 2018.

You can use the guide before or after you have read the book, or as you finish each chapter. The study questions provided are not meant to cover all aspects of the book but, rather, to address specific ideas that might warrant further reflection.

Most of the questions contained in this study guide are ones you can contemplate on your own, but you might consider pairing with a colleague or forming a study group with others who have read (or are reading) *Assessment Literacy for Educators in a Hurry*. And although most of the guide's questions are intended for consideration by an individual reader, a group activity suggestion is also included for each of the book's chapters.

All the questions and suggested activities were provided by the book's author.



Preface

1. The author explains that *Assessment Literacy for Educators in a Hurry* will focus on a small number of assessment-related understandings "that will typically reduce the number and the magnitude of measurement-based mistakes within the world of education." Do you think this claim is realistic, or is it simply a ploy to get readers to take this book more seriously? Can a book about testing truly have a meaningful impact on how students are *taught* and how well they *learn*?
2. Beyond teachers and school administrators, who do you think would benefit from developing assessment literacy? Make a list of other stakeholder groups, and rank these groups in priority order, beginning with the one that most needs to know the basics of educational measurement.

Chapter 1. Assessment Literacy: The What, the Why, and the How

1. Do you agree with author that increasing assessment literacy would be the most cost-effective way to improve our schools? Why or why not?
2. The author provides his definition of assessment literacy on page 2. Try generating your own version of this definition using synonyms and equivalent phrases. Then try generating *another* version. The more ways you can explain assessment literacy, the better you will have internalized its meaning.

3. The author contends that many most educators "know squat" about educational assessment's fundamentals. Consider your own experience as an educator working among educators. What evidence have you seen to support this conclusion? Why do you think testing is a relatively mysterious realm to so many educators?
4. The author identifies three major categories of mistakes made by assessment-unknowledgeable educators: (1) *using the wrong tests*, (2) *misusing the results of the right tests*, and (3) *failing to employ instructionally useful tests*. First, think through the nature of each of these three decision errors, then rank them in order of their harmfulness. If you are tackling this study guide with a colleague, compare your sets of rankings and discuss why you ordered them as you did.
5. *Interim tests*, sometimes referred to as "benchmark tests," are often recommended as an important element in a balanced assessment system. Does your school use interim tests? Do you and your colleagues find them helpful? If so, how?

Group Activity: *A Mine-Field of Assessment-Spawned Mistakes*

Working in two randomly composed, essentially equal groups, create a list of poor instructional decisions that might result from a lack of assessment literacy, consider both errors of commission and omission. Rank these mistakes from most to least educationally harmful, and focus on the top three. After 15 minutes, have each team presents its list, along with explanations of the mistakes' adverse impact. Reunite as a large group to discuss these errors and their effects.

Chapter 2. Validity: The Overt in Search of the Covert

1. Review the Validity Understanding and see if you can use synonyms and equivalent terms to rephrase it in two different ways.
2. Imagine you are standing in line at the grocery checkout. While the cashier is looking up the product code for an exotic organic squash, the person behind you taps you on the shoulder and asks you how, in the field of educational testing, overt responses are indicative of a student's covert status. You have 20 seconds to respond. What do you say?
3. Why does the author make such a big deal about there being no such thing as "a valid test" or "an invalid test"? What's the danger of seeing validity as an attribute of a test rather than as a judgment of the test's interpretive accuracy consonant with the test's intended purpose? How might you explain this concept during a parent conference?
4. Imagine you are chairing a district-level assessment committee whose mission is to enhance the assessment literacy of district teachers and administrators. What kinds of evidence could your committee assemble to demonstrate the accuracy of test-based interpretations of students' skills and knowledge?
5. What are the kinds of evidence that should be examined in order to determine if a particular test will fulfill its stated educational mission? Consider the different types of testing purposes (*comparative, instructional, and evaluative*). Does the kind of validity evidence required vary according to the test's purpose?

Group Activity: Dueling Assessment Arguments

Choose a kind of test (*comparative, instructional, or evaluative*) and a setting for that test's administration. Then, working in two teams, identify the key features of a validity argument that could be used to support that test's suitability. After 15 minutes, have each team describe their validation plan. Conclude with a general discussion of the strengths and weaknesses of the two plans.

Chapter 3. Reliability: Assessment's Righteous Rascal

1. Imagine you need to explain to a state legislator (who knows nothing about educational testing) why validity and reliability are important and how the two concepts are related. Carefully think through the necessary elements of your explanation, then try delivering it out loud.
2. When the topic of reliability is raised, most educators then to think about *test-retest reliability*—the stability of students' responses over time. Why do you think this is so? What occasions, if any, have you had to think about or consider *alternate-form* and *internal consistency* reliability evidence?
3. Why is it such a big deal that a test's reliability evidence mesh with that test's stated measurement mission?
4. The Reliability Understanding states that reliability evidence should be reported not only for *groups* of test-takers but also for *individual* test-takers. Why is this dual focus so important?
5. Imagine you've been called upon to explain the three kinds of reliability at the next PTA meeting. First, why might this be a topic of concern to your school's parents? Second, how would you explain *test-retest reliability*, *alternate-form reliability*, and *internal consistency reliability* to this audience? What examples and illustrations could you use to clarify these concepts?

Group Activity: Turn-and-Talk Reliability Exemplars

Break into small groups of three and review the three different kinds of reliability evidence. Then work individually for 5 to 10 minutes to come up with specific examples of each kind. Take turns presenting your examples, discussing any disagreements or points of confusion. The goal of this activity is help everyone become more comfortable differentiating among the three categories of reliability evidence.

Chapter 4. Fairness in Testing: It's About Time

1. In Chapter 4, the author emphasizes that unfairness in a test decreases that test's reliability. Why is this so? How would you explain the relationship between test fairness and reliability to a colleague?
2. The 2014 edition of the *Joint Standards* elevates test fairness to a position equal to that of validity and reliability. Why do you think it took so long for the fairness of educational testing to be considered so significant?
3. A test's fairness can be evaluated based on *judgmental evidence* and *empirical evidence*. Which of the two evidence types—if either—do you regard as more persuasive? Why so? How might the value of these evidence types vary with different test purposes and testing circumstances?

4. Suppose you were asked to explain to a group of colleagues how a bias-review committee typically reviews, evaluates, and then deletes or revises test items found to exhibit assessment bias. What are the necessary steps and key procedural activities?
5. The author recommends the following question be asked when reviewing test items for potential bias: "Might this item offend or unfairly penalize any groups of students because of personal characteristics such as gender, ethnicity, religion, or race?" How might the phrasing of the item-review directions reviewers receive affect their per-item judgments? Are there any changes you would make to the phrasing of the author's recommended question? If so, what changes and why?

Group Activity: *True Depictions or False Distortions of Assessment Fairness Concepts*

Break into two or more subgroups to caucus privately and come up with at least two true/false statements that accurately or inaccurately describe an aspect of *accessibility*, *universal design*, and *accommodation*. Then take turns reading these truly depicted or falsely distorted statements aloud. A spokesperson for the other subgroup(s) then indicates whether the statements uttered are depicted (D) or distorted (also D). Because the response will always be "D," engage in follow-up discussion of why each response is either true or false.

Chapter 5. Score Reports: Information That Supports Action

1. Say you're an educator who wants excellent score reports that furnish actionable insight accompany every standardized test. What steps might you take to help bring this about?
2. Educators need score reports that are easily interpreted and in accord with the test's intended use. What specific procedures might be employed to ascertain the degree to which a test's reports support that test's purpose?
3. Why is it such a big deal that so many score reports don't furnish information in a way that readily informs instructional next steps?
4. The author contends that Americans put great evaluative stock in the results of standardized tests even though the vast majority of the tests we use to evaluate schools or teachers provide no evidence supporting this use. Why do you believe this situation continues to exist?
5. In general, a test's results are more useful when they are provided with reasonable specificity, but is it possible for a score report to be too specific? What are the pros and cons of score reports that furnish individual results on an item-by-item basis?

Group Activity: *Positives and Negatives of Score-Reporting Procedures*

Break into two or more subgroups to discuss the three popular ways of reporting test scores, described in Chapter 5: *percentiles*, *scale scores*, and *performance-level categories*. For each of the three reporting mechanisms, identify at least two positive consequences of using it and two negative consequences. Then rank, in priority order, the most positive and the most negative features of each reporting scheme. Back in the large group, discuss each reporting method, one at a time, with the subgroups

sharing their rankings and thoughts. Conclude the activity with a total-group look-back discussion of the virtues and vices of all three reporting schemes.

Chapter 6. Formative Assessment: Ends-Means Magic

1. The author shares his trusted definition of formative assessment on page 89: "a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics." By this definition, does a teacher's immediate adjustment in instructional activities in response to students' puzzled facial expressions or confused questions count as "formative assessment"? Why or why not? What are the benefits of executing a planned process as compared to a spontaneous reaction?
2. If a beginning teacher asked you to explain how the formative assessment process exemplifies traditional ends-means thinking, what would you say?
3. The author contends that the formative assessment process is being seriously underused in the United States. Assuming he is correct, how do you explain formative assessment's underuse? What factors might preclude the use of formative assessment in your classroom, your school, and in general? Are these factors fixable?
4. The author mentions that outside the United States, formative assessment is celebrated more for its value in fostering self-regulated learning than for its ability to improve teachers' instructional effectiveness. Why do you think this disparity exists? Should U.S. educators push harder to build students' self-regulation skills? If you use formative assessment now, how might you adjust your application of it to pursue this end?
5. *Learning progressions* are championed as a valuable tool for planning the formative assessment process. What's your experience designing learning progressions, and what's your outlook on their utility? What advice would you offer a new teacher about the optimal number of learning progression building blocks when promoting students' mastery of curricular aims?

Group Activity: *Dueling Promulgation Ploys*

Break into two or more small teams and caucus for a while about what circumstances, tactics, or pitches would most effectively encourage educators to try formative assessment. Reconvene and have each team describe its three highest-priority ploys.

Chapter 7. Measuring Affect: Getting a Fix on Behavioral Changes

1. How important do you think it is to pursue affective educational objectives as well as cognitive ones? Are what students know and can do more important than what students feel, believe, and value? Much more important? Of equal importance? Why?
2. Do you agree or disagree with Figure 7.1's depiction of the predictive potency of students' current affective status? Why?

3. Which of the three varieties of student affect (*attitudes, interests, and values*) do you think has the most influence on students' current and future behaviors? Why?
4. What experience do you have assessing student affect? Which of the three affective variables (*attitudes, interests, and values*) do you think is most difficult to assess? Why?
5. Imagine a parent has just asked you why it's so important, when assessment affective, to draw inferences only about groups of students and never about an individual student. What would you incorporate in your explanation? (*Warning: This is a more difficult explanatory task than it might first appear to be!*)

Group Activity: Affective Inventory Item-Writing

Working in one large group or smaller subgroups, identify a single affective target (e.g., students' confidence as learners, students' belief that they are safe in school). Now agree on the age of the children to be assessed (e.g., high school) and the setting in which the students will be assessed (e.g., a middle-class suburban school district). Review the format of the self-report Likert-like inventory and refer to the example on page 116. Have each member of the group write two items that might be included in a self-report affective inventory appropriate for the designated objective, population, and setting. Reconvene to share, review, and improve these items.

Chapter 8. Wrapping Up, Reaching Out

1. Imagine you're stuck in the center seat on a cross-country flight. To your complete surprise, the person to your right strikes up a conversation about educational assessment and asks you to identify the six—no more, no less—assessment-related understandings that, if truly comprehended, would render someone assessment literate. What would you say? Practice providing a lucid recounting of the "splendid six" until you feel that you "own" them.
2. What's your reaction to the proposed establishment of assessment translators? Do you think this is a practical possibility or simply a pie-in-the-sky yearning from a desperate author?
3. Do you think it's possible for an educator to become assessment literate without also wanting to enhance the assessment literacy of others? Why or why not?
4. Who do you think would be good audiences for Chapter 8's op-ed essays? Does any essay seem particularly well-suited for a specific audience? If so, explain. How might you revise one or more of these essays to better target a specific audience you have in mind?
5. As another way of internalizing the book's six assessment-related understandings, take a moment to review them all and then rank them in priority order from most important to least important. If you can, reach out to someone else who has read this book and see if you can get that to do the same prioritizing. To what degree are your six rankings similar or different?

Group Activity: *Making a Battle Plan for the Assessment Literacy Army*

Working in one large group or several smaller ones, come up with a realistic strategy for promoting increased assessment literacy among members of one or more target groups with a stake in education. In addition to sharing the four op-ed essays in Chapter 8, what else could you do to advance the cause of assessment literacy? Conceptualize an overall strategy for enhancing the assessment literacy of one specific group of stakeholders, including as many *specific* and *practically implementable* tactics as you can. At the close of your group's analysis, pause to consider whether there are any elements of your overall strategy and its subsumed tactics that your group might actually implement. If there are, get busy!

Assessment Literacy for Educators in a Hurry was written by W. James Popham. This 145-page, 7" × 9" book (Stock #119009; ISBN-13: 978-1-4166-2648-0) is available from ASCD. Copyright © 2018 by ASCD. To order a copy, call ASCD at 1-800-933-2723 (in Virginia 1-703-578-9600) and press 2 for the Service Center. Or buy the book from ASCD's Online Store.