



# LEARNING POINT

## What fundamental understandings are necessary for assessment literacy?

Assessment expert W. James Popham has described six practical and action-oriented understandings that form the basis of assessment literacy. Internalizing these “high-priority” understandings can equip educators to make sound assessment and instructional decisions and improve the quality of education their students receive.

This Learning Point summarizes Popham’s “six understandings of fundamental assessment concepts and procedures”: validity, reliability, fairness, score reporting, formative assessment, and affective assessment.

### Validity

Validity, the degree to which an evidence-based argument supports the accuracy of a test’s interpretations for a proposed use of the test’s results, is the necessary precursor to all educational assessment.

There is no such thing as a “valid test”; rather validity describes the quality of the argument made about a test, built on ev-

**“It is my firm conviction that if you comprehend and internalize the six assessment-related understandings presented in this book, your assessment literacy level will be sufficient to render your on-the-job decisions much sounder and more defensible. The beneficiaries of this enlightened decision making will be those you educate.”**

—W. James Popham

idence that the test will capture the kind of data it purports to capture and support the interpretations it is intended to support. Validity relies on our confidence—supported by evidence—in the accuracy of score-based inferences about test-takers and the test’s utility in supporting the test’s intended use.

Without validity, educational testing would have no point, no purpose, and no legitimate application.

“Unlike baseball, where it takes three strikes to get a batter out,” Popham writes, “educational testing is a one-strike contest. If the test loses out on validity, it’s not just out—the whole game is over.”

### Reliability

Assessment reliability, the consistency with which a test measures whatever it measures, is represented by three conceptually different kinds of evidence, and it should be reported for both test-taker groups and individual test-takers.

Dr. Popham describes in his book three of the more important types of reliability:

- Test-retest reliability (also referred to as “test stability”): the consistency between the score a student would achieve during one test occasion with a score he or she

would achieve on a second test occasion.

- Alternate-form reliability: the consistency with which alternate versions of a test measure student performance.

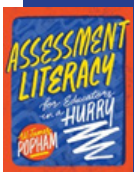
- Internal consistency reliability: the homogeneity of a test’s items, or the degree to which a test’s items are measuring the same skills or knowledge in the same way throughout the test.

Evidence must be supplied for both solo and group test takers. Reliability is one necessary condition for valid test-based interpretations.

### Fairness

Fairness describes the degree to which a test contains measures that are free of bias. Fairness—along with validity and reliability—forms the third element in a “big three” trio of concerns that must be considered when developing or evaluating educational tests. Educators seek two types of evidence to document test fairness (when practicable):

- Empirical analyses of the per-item performances of all subgroups taking a test is a powerful measure of fairness, but it comes with several constraints that make it infeasible for use in individual school districts, classrooms, grade-level teams, or departments.



This Learning Point was adapted with permission from *Assessment Literacy for Educators in a Hurry*, (ASCD, 2018) by W. James Popham. Learn more about the book and the author at [www.ascd.org/publications/books/119009.aspx](http://www.ascd.org/publications/books/119009.aspx).

■ Judgmental scrutiny for bias in a large-scale assessment involves item-by-item bias review by a committee that includes representatives of any subgroups thought capable of being adversely affected by a test containing biased items. For classroom tests, judgmental scrutiny can happen among a colleague or two. It involves scrutinizing test items, asking whether bias might be present, and taking corrective steps to ensure the quality of a test, consequently increasing educators' ability to make better educational decisions.

Accessibility, universal design, and accommodation are additional concepts that test developers consider when promoting fairness in educational testing for all students—including those with special needs.

## Score Reporting

Because inferences about students are based on test-takers' score reports, users must demand that results be easily interpreted in accord with the test's intended use.

Well-designed score reports can help educators make good decisions about students. Test results need to be reported in a way that engenders actionability related to one of several possible missions: making comparisons among test-takers, improving ongoing instruction and learning, or evaluating the quality of instruction.

"We educators must clamor for score reports that help us interpret students' test performances," Popham writes. "A demand for such actionable score reports will engender the building of better, purpose-supportive tests."

## Formative Assessment

Although currently underused, formative assessment is a remarkable, research-ratified process in which teachers and students use classroom assessments to make needed adjustments that can dramatically improve students' learning. Formative assessment is a process that features the instructional use of classroom assessments. Its enormous effectiveness—backed by extensive supportive research evidence—results in part from its consistency with a classic "ends-means" goal attainment strategy:

1. Identify the desired end.
2. Select the means and implement it.

## To learn more

### Standards for Educational and Psychological Testing

<https://www.apa.org/science/programs/testing/standards>  
Titles below are available from online booksellers.

### Assessing Affective Characteristics in the Schools

Lorin W. Anderson and Sid F. Bourke (2nd ed. 2000)

### Using Formative Assessment to Enhance Learning, Achievement, and Academic Self-Regulation

Heidi L. Andrade, & Margaret Heritage (2018)

### Formative Assessment: Making it Happen in the Classroom

Margaret Heritage (2010)

### Classroom Assessment: What Teachers Need to Know

W. James Popham (8th ed. 2017a.)

### An Introduction to Student-Involved Assessment for Learning

Rick Stiggins and Jan Chapuis (7th ed. 2017)

3. Evaluate the effect and, if necessary, select an alternate means.

Popham contends that teachers and students should use the formative assessment process far more frequently to adjust what they are doing in order to achieve important curricular goals.

## Affective Assessment

Because affective dispositions acquired in school can have a profound influence on students' success, both during and following school, students' affect should be assessed regularly using anonymously completed self-reported inventories.

Educators design lessons to promote students' mastery of behavioral objectives across three categories: cognitive (knowledge or skills); affective (attitudes, interests, or values); and psychomotor (small- or large-muscle physical skills). Educators traditionally have focused most on promoting and assessing knowledge and skills. However, because of the impact that students' current dispositions have on behavior, educators could improve outcomes by addressing and assessing student affect.

"Modest effort on the part of a teacher in students' affect can sometimes change the lives of students," Popham writes. However, he adds one caution: "If you

decide to move personally into the realm of students' affective modification, be sure to do so with at least rudimentary knowledge about the assessment and influence of students' attitudes, interest, and values. Affect is too important to mess up.

## Conclusion

These six understandings, if shared among all educators at the classroom and leadership levels, could equip them with assessment and testing practices that:

- are worth the valuable time they take to administer
- accurately measure what students have learned
- fairly reflect teacher and school effectiveness
- provide instructionally useful data that will help students learn faster and better

*Popham himself sums it up best: "It is my firm conviction that if you comprehend and internalize the six assessment-related understandings presented in this book, your assessment literacy level will be sufficient to render your on-the-job decisions much sounder and more defensible. The beneficiaries of this enlightened decision making will be those you educate."*

The Michigan Assessment Consortium's Assessment Learning Network (ALN) is a professional learning community consisting of members from MI's professional education organizations; the goal of the ALN is to increase the assessment literacy of all of Michigan's professional educators.