



LEARNING POINT

Start with Purpose When Choosing Assessments

Assessments are built for a specific purpose or purposes. When evidence shows that a test actually measures what it was built to measure, then we say that the test has validity for that intended use. If, on the other hand, a test is used to measure something that it wasn't designed to provide information on, and no evidence has been collected to support that additional use, then we would say that there is no support for the validity of using the test for that purpose. Test builders need to provide evidence that a test is valid for each specific purpose and use. If we want our test to measure, say, both proficiency and growth, evidence must be presented that shows the test provides information related to both of those goals.

It is important to note that a test, itself, is not valid or invalid. Rather it is the use of the results of an assessment for a specific purpose that can be valid or invalid, depending on whether there is information to support the intended use. For example, suppose we have a test designed to diagnose student understanding of fractions. We can imagine that this test would have items about numerators, denominators, finding com-

mon denominators, reducing, and basic operations with fractions. This test might provide excellent diagnostic information regarding which aspects of fractions and their uses students have mastered and which they are still working to develop.

Suppose we like this test so much for the useful diagnostic information it gives that we start using it to determine whether students are proficient in fourth-grade mathematics. Now we have created a problem because we have over-reached in our interpretation of what this test can tell us. The fourth-grade mathematics curriculum is more than just fractions so this assessment would be a very incomplete sampling of the fourth-grade mathematics curriculum and thus not a very good assessment of whether students have mastered all of the fourth-grade mathematics content. For this broader purpose, we would

say the test is invalid, since the evidence we could readily collect would show that fractions are only a small part of what fourth-graders are expected to know in mathematics. That doesn't diminish its utility and validity for its designed purpose—to diagnose understanding (or lack thereof) of fractions.

The above example illustrates an important point about test construction. The diagnostic assessment would be built so that it assesses a relatively narrow bit of content, but to do so in some depth. A broader assessment for checking mastery of an entire body of content (e.g., fourth grade mathematics) would pull from the entire domain of content but won't necessarily go deeply into all areas due to time constraints. We don't have enough time, or will, to ask students to go deeply into all bits of content for fourth grade mathematics on a summative assessment.

When there is evidence to support the use of a particular test for a specific purpose, we say that the test is valid for that purpose.

Ed Roeber, PhD, Michigan Assessment Consortium

As another example, suppose we would like to build an assessment to predict how well students will do on the upcoming summative assessment and inform how to improve their performance on that assessment. Perie, Marion, and Gong (2009) point out that these two purposes might not be compatible and could reduce the validity of using the test for one or both purposes. They point out

“A confounding variable on any predictive test is that if it provides good feedback on how to improve a student’s learning, then its predictive ability is likely to decrease. That is, if the test predicts that a student is on track to perform at the basic level, and then appropriate interventions are used to bring the student to proficient, the statistical analysis of the test’s predictive validity should under predict student performance over time.”

As you can see, it is important to be very clear about what we want from each and every test that we administer to students. We need to ensure that reasons we administer a test match what that test was built to do, and that there are data to support those uses. If we use assessments for things that they weren’t designed for, we run the risk of drawing poor conclusions based on the invalid testing data. This will in turn diminish the efficiency of our educational process and our effectiveness as educators.

Perie, M., Marion, S. and Gong, B. (2009), *Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments*. *Educational Measurement: Issues and Practice*, 28: 5–13.

What primary purpose does educational assessment serve?

- Assessment is a process of gathering information for the purpose of making judgments about a current state of affairs.
- In educational assessment, the information collected is designed to help teachers, administrators, policy makers, and the public infer what students know and how well they know it, presumably for the purpose of enhancing future outcomes.
- Some of these outcomes are more immediate such as the use of assessment in the classroom to improve student learning, and others are more delayed such as the use of assessment for program evaluation.

What functions and purposes is assessment supposed to serve?

- Educational assessment typically occurs in multiple contexts:
 - Small-scale: individual classrooms
 - Intermediate-scale: districts
 - Large-scale: states, nations, international
- Within and across contexts it can be used by different stakeholders to accomplish differing purposes:
 - Assist learning (formative)
 - Measure individual (or group) achievement (interim/summative)
 - Evaluate programs (interim/summative)

Source: Jim Pellegrino, Ph.D., Learning Sciences Research Institute, University of Illinois at Chicago. Assessment Learning Network presentation, 12.16.16

To learn more

Criteria for High Quality Assessment

Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). Stanford, CA: Stanford Center for Opportunity Policy in Education.

<http://tinyurl.com/criteria-assessment>

Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments

Perie, M., Marion, S. and Gong, B. (2009), *Educational Measurement: Issues and Practice*, 28: 5–13.

<http://ncaase.com/docs/PerieMarionGong2009.pdf>

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.