

**Presenter's Notes for PowerPoint:
Understanding Validity for Classroom Teachers**

Slide # Title	Presenter Notes
1. Understanding Validity for Teachers	This module provides teachers with basic information they need to understand what validity is and why it is so important. Both reliability and validity are necessary for a teacher to make accurate inferences about student learning based on assessment results. For this reason, you need to study both this module and the Reliability module to understand quality assessment of student learning.
2. Validity & Reliability	<p>This commonly used illustration provides visual metaphors of the concepts of Reliability and Validity.</p> <ul style="list-style-type: none"> • The first target represents a valid and reliable assessment. The bull's eye in the center of the target is shot with accuracy again and again. The correct target is being aimed at and it is being hit consistently. So validity (or the correct target) and reliability (or consistency) are evident. • The second target shows a scenario in which the shots are consistently striking in a particular location – but it isn't on the target. So if this were an assessment it would be reliable but not valid. • The third target shows another scenario. The shots are aimed at the target, but they are striking the target in a variety of locations and some aren't hitting the target at all. There is no consistency or reliability and because of this there is also no validity. <p>So it is important to remember that validity and reliability interact with and influence each other. Both are necessary for quality educational tests.</p>
3. Essential Questions	<p>This module answers the essential questions:</p> <ul style="list-style-type: none"> • What is test validity? • What is content-related validity? • What is criterion-related validity? • What is construct-related validity? • What does a classroom teacher need to know about validity to help ensure the quality of classroom assessment? <p>We'll begin with the question: What is test validity?</p>
4. What is Test Validity?	Validity refers to whether or not an assessment measures what it is supposed to measure. Even if a test is reliable, it may not provide valid results. Let's imagine a bathroom scale that consistently tells you that you weigh 130 pounds. The reliability (consistency) of this scale is very good, but it is not accurate (valid) because you actually weigh 145 pounds! Since teachers, parents, and school districts make decisions about students based on assessment results the validity inferred from assessments results is essential. Also, if a test is valid, it is almost always reliable.
5. Validity is the Most Important Characteristic of a Test	<p>Validity is the most important characteristic of a test. You may remember that a test is or is not reliable but it is the results of a test that are valid or invalid. Because we make important decision about student learning based on assessment results we need to know how to tell if an assessment's results are valid or invalid.</p> <p>Validity evidence answers the questions:</p> <ul style="list-style-type: none"> • Does the test cover what we believe (or are told) that it covers? • To what extent? • Is the assessment being used for an appropriate purpose?
6. What is Test Validity?	<p>Validity is a judgmental inference about the interpretation of students' tests performances.</p> <p>For teachers to be confident that their score based inferences are valid, it is usually necessary to assemble compelling evidence that supports the accuracy of those score based inferences.</p>

7. What is Test Validity?	In terms of student assessment, validity refers to the extent to which a test's results are representative of the actual knowledge and/or skills we want to measure and whether the test results can be used to determine accurate conclusions about those knowledge and/or skills.
8. Test Construct	<p>In order to understand validity evidence, we need to understand what a test construct is.</p> <ul style="list-style-type: none"> • A construct is the “target” or “objective” we want to know about. • Validity addresses how well an assessment technique provides useful information about the construct. • Construct underrepresentation is when the test does not assess the entire construct; the test misses things we should be assessing. • Construct irrelevant variance is when the test is assessing things that are not really part of the construct; we are assessing irrelevant stuff that we don’t want. • Now consider that there can be varying degrees of construct underrepresentation and irrelevant variance. <p>In these illustrations of three different tests, the test content that aligns with the construct is shown in the blue area. The test content in the blue area represents test items that will yield valid results.</p>
9. Validity is a matter of degree.	Because validity is a matter of degree it is appropriate to use relative terms such as high validity, moderate validity, and low validity. So rather than saying “The unit test’s results are valid,” it is more accurate to say “The unit test’s results have a high degree of validity.” To know the degree of validity we need to collect as much evidence as possible that demonstrates that a test’s results accurately represent the students’ knowledge and skills we are trying to measure.
10. Three Categories of Validity Evidence	<p>Traditionally, the ways of accumulating validity evidence are grouped into three categories called construct-related, content-related and criterion-related evidence of validity.</p> <p>To know if a test’s results really do have validity – we need to collect several types of evidence that span all three of the categories.</p> <p>These three categories are Construct-Related, Content-Related, and Criterion-Related Evidence.</p> <ul style="list-style-type: none"> • Construct-Related Validity Evidence is the extent to which an assessment corresponds to other variables, as predicted by some rationale or theory. • Content-Related Validity Evidence is the extent to which the content of the test matches the instructional objectives. • Criterion-Related Validity Evidence is the extent to which scores on the test are in agreement with or predict an external criterion.
11. Activity One What is Test Validity?	Let’s stop now and participate in Activity One where we will address the essential question: What is Test Validity?
12. Content Validity	Content validity refers to the extent a test adequately represents the subject-matter or behavior to be measured. For a test to have optimum content validity it must be based upon a well-defined domain of knowledge or behavior. For teachers, content validity is the most important type of validity for classroom and achievement tests.
13. Where we find well-defined domains:	<p>Where do teachers find the well-defined domain of knowledge or behavior? There are three answers:</p> <ul style="list-style-type: none"> • In the textbooks we use. The learning objectives are usually listed at the beginning of chapters and are addressed in end-of-chapter questions and in terms that are specifically defined. • In the school district’s curriculum guides. The standards and competencies describe what students should know and be able to do. • In state academic curricular standards. Here too, what students should know and be able to do are described.

	Most assessment domains for classroom tests consist of the knowledge and skills included in a teacher's objectives for a certain instructional period. An assessment domain is the content standards being sought by a teacher. For example if a science teacher were wanting students to master a set of six content standards during a semester, the teacher's final exam will sample the skills and knowledge in those six content standards.
14. Bloom's Taxonomy	After we have the assessment domain identified, we need to decide what we actually expect students to know and be able to do. It is good old Bloom's Taxonomy that will help us connect the content to the mental processes students are expected to employ. You may remember that Bloom's Taxonomy represents a continuum of increasing cognitive complexity—from remembering to creating. Through connecting the content to the mental processes expected of students we build content validity into our classroom tests.
15. Focus on the verbs included in learning objectives!	Far too often, educators don't focus on the verbs included in learning objectives. Let's take a look at this 8th grade Kansas History Indicator. Numerous test items can be written about migration in the United States. But to conform to this indicator's real meaning, test items cannot ask students to provide memorized facts about migration to the United States because that would be asking students to demonstrate learning at the Remembering level of Bloom's Taxonomy. Remembering is much less complex than what is intended for the student to know and be able to do. We also cannot ask students to evaluate the "economic, political, and social factors . . ." because that is more complex than what is intended. If they are to be valid, test items for this indicator need to require students to be Analyzing.
16. Often we stop at lowest level.	Because test items that reflect the lowest level of Bloom's Taxonomy are the easiest to write, most teacher-made tests are composed almost entirely of knowledge-level items. As a result, students focus on verbatim memorization rather than on meaningful learning. While the teacher gets some indication of what students know, such tests tell nothing about what students can do with that knowledge. This can also mean that the intended target or objective that calls for students to demonstrate their ability to analyze, evaluate, or create is never actually tested.
17. How do educators determine if assessments have content validity?	<p>How do we determine if assessments have content validity? One way is to use test development procedures focused on assuring that the assessment domain's content is properly reflected in the assessment procedure itself. The higher the stakes associated with the test's use the more effort is devoted to making certain the assessment's content represents the assessment domain.</p> <p>These are some of the activities that might be carried out during the test development process for a high stakes test to assure the new test represents the assessment domain.</p> <ul style="list-style-type: none"> • A panel of national content experts recommends the knowledge and skills that should be measured by the new test. • The proposed content of the new test is systematically contrasted with a list of topics derived from a careful analysis of the content included in the five leading textbooks used in the nation related to the assessment domain. • A group of teachers, each judged to be a "teacher leader" in his or her state provides suggestions regarding the key topics or knowledge and skills that to be measured by the new test. • Several college professors who are international authorities offer recommendations for additions, deletions, and modifications of the knowledge and skills that has been determined by the others. • State and national associations provide reviews of the proposed content to be measured by the new tests.
18. So what can a classroom teacher do?	A classroom teacher cannot go through these elaborate processes to ensure that an assessment has content validity. So what can a classroom teacher do? Make a careful effort to conceptualize an assessment domain and try to see if the test being

	constructed actually contains content that is appropriately representative. One tool that can help a teacher do this is a Table of Specification.
19. Use a Table of Specifications	<p>Even if you are not trying to assess every concept taught, covering all the substantial learning from a unit or quarter or semester can be a time-prohibitive task. Thus, most tests assess a representative sample of the content domains. Teachers who construct the tests are normally responsible for determining a representative sample. To make sure a sample of test questions is sufficient and representative, teachers sometimes create a matrix of standards (or objectives) and the level or type of skill required. This matrix is often called a Table of Specifications. Essentially, a table of specification is a table chart that breaks down the topics that will be on a test and the amount of test questions or percentage of weight each section will have on the final test grade. It provides teachers and their students with a visual of the content that will be tested. As part of the entire teaching process, many education experts advise constructing a table of specification early in the lesson plan building process in order to ensure that the content of lessons and projects match what will ultimately appear on a test. By offering students the opportunity to view a table of specification, teachers offer their students the opportunity to view a certain kind of rubric against which they will be graded. This opportunity allows students to have full knowledge about what they will be tested over and which sections or topics of their study will be tested. According to some educators, the table of specification is just as important for the students as it is for their teachers.</p>
20. Ask another teacher to look over a test's items.	Another simple way that teachers can review their classroom tests for content validity is to ask another teacher to look over a test's items and provide the same kind of judgments that a review panel does for a high stakes national or state-level test. A pair of teachers could do this for one another. The more carefully you review your classroom tests' content coverage, the more likely the test's content coverage will have a higher degree of validity.
21. Evaluating Content Validity in Existing Tests	<i>No notes for this slide.</i>
22. Opportunity to Learn	<p>Before we leave content validity we need to talk about one more thing – Opportunity to Learn. An idea related to content validity is a concern called instructional validity. This depends upon the teacher. The content may be in the state standards and the students' text book and the state standards but not be taught. Teachers sometimes skip items of instruction they don't understand or don't have time to teach. Yet if related items appear on a test, this reduces the validity of the test since the students had no opportunity to learn the knowledge or skill being assessed.</p>
23. Activity Two What is Content-Related Validity?	<p>Let's stop now and participate in Activity Two where we will address the essential question: What is Content-Related Validity?</p>
24. Criterion Related Evidence of Validity	<p>Now let's learn about Criterion related evidence of validity. Basically, Criterion validity demonstrates the degree of accuracy of a test by comparing it with another test, measure or procedure which has been demonstrated to be valid. There are two contexts for using Criterion validity.</p> <ul style="list-style-type: none"> • Predictive validity is when one measure is done in the present one is done at a later date. The later test is known to be valid. This approach allows me to show my current test is valid by comparing it to a future valid test. • Concurrent validity is when both measures are current. This approach allows me to show my test is valid by comparing it with an already valid test. I can do this if I can show my test varies directly with a measure of the same construct or indirectly with a measure of an opposite construct.

25. Predictive Validity	Predictive validity is used for aptitude tests or tests that are used in order to predict how well a student will perform at some later point in time. A comparison must be made between the test and some later behavior that it predicts.
26. Teachers don't need . . .	Teachers don't need to know how to calculate predictive validity – but if they know that a predictor test such as the ACT works well, a teacher can use its results to help make educational decisions about students. For, example, if a student is wanting postsecondary education, but has poor scores on a scholastic aptitude test, a teacher could devise a set of supplemental instructional activities so that the student could try to acquire the needed academic skills before leaving high school. The same type of thing could be done for the preschooler who on an aptitude test shows that they are not yet prepared for kindergarten.
27. Concurrent Validity	Now let's move on to Concurrent validity. Concurrent validity compares scores on a test with current performance on some other test. Unlike predictive validity, where the second test occurs later, concurrent validity requires a second test at about the same time. Concurrent validity for a science test could be investigated by correlating scores for the test with scores from another established science test taken about the same time. Another way to gather concurrent validity evidence is to administer the test to two groups who are known to differ on the content being measured. For example a group of students who have taken a high school chemistry class and another group who has not. One would have support for concurrent validity if the scores for the two groups were very different.
28. Validity Coefficients	<p>The computed statistic in both predictive and concurrent validity is a coefficient “r” which we will call a validity coefficient. The number that indicates high validity can vary depending upon what has been determined to be acceptable. Many commercial assessments such as the ACT consider 0.5 to be an acceptable validity coefficient.</p> <p>To calculate a validity coefficient we use the same process that we used to calculate the correlation coefficient for determining reliability. This is described in Module 2 Reliability.</p> <p>Here you see a Table adapted from <i>Mastering Assessment Booklet, Reliability: What Is It and Is It Necessary?</i> that illustrates simple interpretations of validity coefficients.</p>
29. Activity Three What is Criterion-Related Validity?	Let's stop now and participate in Activity Three where we will address the essential question: What is Criterion-Related Validity?
30. Construct-Related Validity	Construct-related evidence of validity is the most comprehensive of the three varieties of validity evidence. This is because it covers all forms of validity. It subsumes criterion related validity because it uses empirical evidence. It subsumes content-related validity because it also uses such evidence such as quantifiable content ratings from expert review panelists.
31. Three Types of Strategies	<p>Construct-related evidence of validity for educational tests is usually gathered through a series of studies. There three types of strategies that are most commonly used in construct-related evidence studies:</p> <ul style="list-style-type: none"> • Intervention studies • Differential-Population Studies • Related-Measures Studies
32. Intervention Study	<p>This is how to investigate construct-related evidence of validity using an Intervention Study:</p> <p>We hypothesize that students will respond differently to the assessment instrument if they have received some type of instruction or intervention. For example, if we want to see if a mathematics essay examination measures students higher-order math skills, we hypothesize that students will receive significantly higher scores after an</p>

	intensive six-week summer workshop on higher order math skills than the scores they receive on the same assessment prior to attending the workshop.
33. Differential Population Study	<p>This is how to investigate construct-related evidence of validity using an differential population study:</p> <p>We hypothesize that individuals representing distinctly different populations will score differently on the assessment procedure under consideration. For example, if a new oral test of student's bilingual proficiency in English and Spanish had been created, we would locate three groups of students who were</p> <ul style="list-style-type: none"> • Fluent in English but not Spanish • Fluent in Spanish but not English • Fluent in both Spanish and English <p>We predict that the bilingually fluent speakers will outperform their monolingual counterparts.</p>
34. Related Measures Study	<p>In a related measures study, we hypothesize that there will be a relationship between students' scores on the assessment device we're studying and their scores on a related assessment device.</p> <p>For example, if we are studying a new test of students' reading comprehension, we hypothesize that students' scores on this new test will be positively correlated to their scores on an already established and widely used reading comprehension assessment.</p>
35. Convergent Evidence of Validity	When it is hypothesized that two sets of test scores should be related, and evidence is collected to show that positive relationship, this is referred to as Convergent Evidence of validity. Convergent Evidence of validity answers the question: Are test scores related to behaviors and tests that it should be related to?
36. Discriminant Evidence of Validity	In contrast, a study in which two assessments are shown to have a weak relationship –this weak or lower relationship is referred to as discriminant evidence of validity. Discriminant evidence simply means that if your test is assessing what it is supposed to be that test will relate weakly to results of tests designed to measure other constructs. Discriminant evidence answers the question: Are test scores unrelated to behaviors and tests that it should be unrelated to?
37. Many Approaches to the Collection of Construct-Related Validity Evidence	<p>There are many other approaches to the collection of construct-related validity evidence. What is important to remember is that a number of construct-related validation studies are needed before confidence can be placed in score- based inferences about students. Also if the construct being measured lacks clarity, the more uncertain we are about how to go about measuring students' status related to that construct and more construct-related evidence of validity we need to collect. In summary:</p> <ul style="list-style-type: none"> • There is no single measure of construct validity. • Construct validity is based on the accumulation of knowledge about the test and its relationship to other tests and behaviors. • To establish construct validity, we demonstrate that the measure changes in a logical way when other conditions change.
38. Activity Four What is Construct-Related Validity?	Let's stop now and participate in Activity Four where we will address the essential question: What is Construct-Related Validity?
39. Different Meanings Attached to the Term Assessment Validity	At the beginning of this module we told you that Validity is the most important characteristic of a test. After all, if we can't draw valid inferences from a test about students' knowledge, skill, attitudes, or interests – there isn't any reason to give a test in the first place. Because of this, over time there have been different meanings attached to the term assessment validity that may create some confusion. We need to address these because you are going to encounter these concepts and terms and you need to know what they mean.

40. Face Validity???	The first of these terms is Face Validity. A test is said to have face validity if it "looks like" it is going to measure what it is supposed to measure. The problem with Face Validity is that it is not empirical. Face validity means the test "appears it will work," as opposed to saying "it has been shown to work." This can be dangerous! Face validity is often "created" to influence the opinions of participants who are not expert in testing methodologies – for example school district's looking to purchase a test, parents, and politicians.
41. Consequential Validity	Another, more recently introduced term is something called consequential validity. Consequential validity refers to whether or not the USES of test results are valid. Some professionals feel that, in the real world, the consequences that follow from the use of assessments are important indications of validity. For example, if a test 's results are inappropriately used to deny students' progress to the next grade level the test is consequentially invalid because the results are being used in a way that is wrong.
42. Issues We Should Think About	Understandably, as educators, we sometimes see the consequences as more important than the technical validity of the test. Judgments based on assessments we give and use have value implications and social consequences. So we do need to think about these issues, including: <ul style="list-style-type: none"> • What is the intended use of these test scores? • How are the scores really being used? • Does this testing lead to educational benefits? • Are there negative spin-offs?
43. Consequential Validity NOT True Validity Evidence	Keep in mind, while we should be attentive to the consequences of test use, the idea of consequential validity should not be confused with the true purpose of validity. That purpose is to confirm or disconfirm the defensibility of the score-based inferences we make about our students. If we make accurate inferences about students' based on a test, but use those inferences to make terrible decisions, our test will have negative consequences. But it would be the use to which we put our valid score-based inferences that is terrible. The score based inference itself was, in fact, accurate. So Consequential validity is a good way to remind ourselves of the importance of consequences whenever tests are used. But it is NOT a true form of validity evidence.
44. What is Important?	So what is important for a classroom teacher to understand about validity? We are too busy in our classrooms to collect evidence regarding validity. But for our more important tests we need to devote at least some attention to content-related evidence of validity. Giving attention to the content of the curricular domain being assessed is a good first step. That is why we spent describing and working with the Table of Specifications. However, the criterion and construct-related validity evidence require only that you have a reasonable understanding of what they are. If you are asked to help evaluate a high-stakes educational test or serve on a committee developing a district or state assessment, you will want to know about these two types of validity evidence in order to make good decisions and not be intimidated by the process.
45. What is MOST Important to Understand?	What is MOST important for you to understand about validity is that it IS NOT about the test itself. It is about whether or not the score-based inferences you and other are making about your students are accurate or inaccurate.
46. Practical Advice	<i>No notes for this slide.</i>
47. Activity Five: What does a teacher need to know about validity to help ensure the	<i>Let's stop now and participate in Activity Five where we will address the essential question: What does a classroom teacher need to know about validity to help ensure the quality of classroom assessment?</i>

quality of classroom assessment?	
--	--