

# Module 2C: Assessment Classifications and Scoring Models

by Dr. Tim Brophy

Hello, and welcome back to Passport to Great Teaching, Creative Assessment. My name is Tim Brophy, and it's time for module 2C, in which we're going to discuss assessment classifications and scoring models.

So let's talk first about a class of assessments called "norm-referenced." Now, these are the ones that are based on a set of assumptions that permit us to compare one individual's performance to others who have completed the same assessment. You're familiar with these. The SAT is one. The GRE is one. Examinations generally fall into this category as well.

So these kinds of assessments consist of what we call "dichotomous items." Now, those are ones that have one clear and correct answer, such as those that we put on tests, quizzes, exams. And we'll talk more about that in another module.

So generally, these norm-referenced tests involve the test taker earning a certain number of points for each correct answer. Then the scorer, who is you as the professor, totals the number of points that are earned for the correct answer. This creates a score. And then we assume that that individual's score represents their knowledge of the subject matter that we're testing.

So the score is then placed on some predetermined scale that ranges from low to high-- for example, 0 to 100. The assumption is that the higher the score, the more knowledge the student possesses, or the individual who takes the test possesses. And based on that assumption, then, we can compare these scores among individuals.

There's another category of assessments that do not fall into that model at all. And these are criterion-referenced assessments. These are designed to compare a student's performance to a particular standard or a criterion.

So test takers are generally given some kind of task or an assignment. Their response, whether it's a performance or a behavior or some kind of final product, we assess that for the degree to which it meets certain levels of quality that are outlined in the rubric itself.

So measurements, then, are largely done through expert judgment by individuals qualified to review the response. And that's generally the professor or graduate assistants or other instructors or lecturers. And this could be another disciplinary expert, as well, outside of the university. In some cases, that's what happens.

So the resulting measurements, though-- what makes this different is that it's not intended to be used to compare achievement among those who complete the assessment. Instead, the degree to which an individual meets the requirements and their criteria for the task is the goal here.

So we remember the norm-referenced assessments are specifically designed to compare performance between individuals on the same test, whereas a criterion-referenced assessment is designed to assess the individual's achievement of a specific set of criteria at a certain level of quality that we've established-- not at all to compare one to the other. So we measure these using rubrics. So again, we will cover this later in modules.

So there are basically four rubric models for criterion-referenced assessments. We all are pretty familiar with the scales we use and how we construct those for the norm-referenced assessments.

And in module 3, we'll talk all about writing items for these tests. But let's focus now on building rubrics for these criterion-referenced assessments.

The first one is what I call the bilevel, single criterion model. Then there's the multilevel, single criterion model. There's the bilevel, multiple criteria model and the multilevel, multiple criteria model. So I'm going to describe these briefly as we go through the rest of this session.

The bilevel, single criterion one is the easiest to implement and the most practical for performance assessment situations, because in this model, one criterion is assessed at two levels of achieving-- either demonstrated, usually, or not demonstrated.

There are lots of ways to describe those two levels of achievement. But that makes it very easy in a live-performance situation-- if a student's presenting something or communicating, reading out loud, or whatever they're doing. You can assess it as achieved or not-achieved pretty quickly.

The multilevel, single criterion one is where you first determine what the criterion is you're going to be assessing. Then you determine various levels of performance that are expected and how you can notate this on whatever data collection form that you're using in the most sensible manner.

All right. So this type of guide really helps the teacher record meaningful descriptive data. So depending on how many levels of achievement you have for that one criteria that you're looking for, you could then begin to distinguish levels of achievement in a more granular way.

The bilevel, multiple criterion assessment is one where we designed two or more criteria as either being demonstrated or not demonstrated. So we have two criteria, or three, or four. Each one of those is going to be assessed as being achieved or not achieved.

And in this particular model, it becomes a little more challenging to complete in a live-performance situation like I described earlier. However, it is also appropriate for those products that are assessed after an assessment is completed.

The multilevel, multiple criterion approach is where we have two or more criteria simultaneously judged at more than two levels of achievement. Now, there are two designs here-- analytic and holistic. And we'll get into those more fully in module 3.

But this kind really is effective for our product assessments, this particular model-- particularly if it's a complex, creative, or written task, or a task or a complex performance task, or if the task includes performance as well as something that's created or something written. So this is actually very commonly used on our campus, because most of the tasks that our UF faculty create are these complex tasks that I just described.

So let's think for just a minute about what you do in your courses. Select one of the assessments that you currently administer. Just describe it briefly to yourself, and then answer these questions.

Would you classify this as a norm-referenced or a criterion-referenced assessment? All right. What characteristics of the assessment support your classification?

And think about what scoring approach best describes the one you use for this assessment? Is it a predetermined scale? Or is it one of the criterion-referenced models that I've described here? Thank you. And I'll see you in module 2D.

