

Measurement, Assessment, and Evaluation in Education

We measure distance, we assess learning, and we evaluate results in terms of some set of criteria.

Dr. Bob Kizlik

Updated June 3, 2014

Throughout my years of teaching undergraduate courses, and to some extent, graduate courses, I was continuously reminded each semester that many of my students who had taken the requisite course in "educational tests and measurements" or a course with a similar title as part of their professional preparation, often had confusing ideas about fundamental differences in terms such as measurement, assessment and evaluation as they are used in education. When I asked the question, "what is the difference between assessment and evaluation," I usually got a lot of blank stares. Yet, it seems that understanding the differences between measurement, assessment, and evaluation is fundamental to the knowledge base of professional teachers and effective teaching. Such understanding is also, or at the very least should be a core component of the curricula implemented in universities and colleges required in the education of future teachers. Understanding the properties, purposes, similarities and differences between educational measurement, assessment and evaluation is a fundamental component of the knowledge base of professional teachers.

In many places on the ADPRIMA website the phrase, "Anything not understood in more than one way is not understood at all" appears after some explanation or body of information. That phrase is, in my opinion, a fundamental idea of what should be a cornerstone of all teacher education. Students often struggle with describing or explaining what it means to "understand" something that they say they understand. I believe that in courses on the subject of educational tests and measurements it is often that case that "understanding" is inferred from responses on multiple-choice tests or solving statistical problems. A semester later, when questioned about very fundamental ideas in statistics, measurement, assessment and evaluation, the students in my courses seemingly forgot most, if not all of what they "learned."

Measurement, assessment, and evaluation mean very different things, and yet most of my students were unable to adequately explain the differences. So, in keeping with the ADPRIMA approach to explaining things in as straightforward and meaningful a way as possible, here is what I think are useful descriptions of these three fundamental terms. These are personal opinions, but they have worked for me for many years. They have operational utility, and therefore may also be useful for your purposes.

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. One exception seems to be in the use of the word measure in determining the IQ of a person. The phrase, "this test measures IQ" is commonly used. Measuring such things as attitudes or preferences also applies. However, when we measure, we generally use some standard instrument to determine how big, tall, heavy, voluminous, hot, cold, fast, or straight something actually is. Standard instruments refer to physical devices such as rulers, scales, thermometers, pressure gauges, etc. We measure to obtain information about what is. Such information may or may not be useful, depending on the accuracy of the instruments we use, and our skill at using them. There are few such instruments in the social sciences that approach the validity and reliability of say a 12" ruler. We measure how big a classroom is in terms of square feet, we measure the temperature of the room by using a thermometer, and we use an Ohm meter to determine the voltage, amperage, and resistance in a circuit. In all of these examples, we are not assessing anything; we are simply collecting information relative to some established rule or standard. Assessment is therefore quite different from measurement, and has uses that suggest very different purposes. When used in a learning objective, the definition provided on the ADPRIMA for the behavioral verb measure is: *To apply a standard scale or measuring device to an object, series of objects, events, or conditions, according to practices accepted by those who are skilled in the use of the device or scale.* An important point in the definition is that the person be skilled in the use

of the device or scale. For example, a person who has in his or her possession a working Ohm meter, but does not know how to use it properly, could apply it to an electrical circuit but the obtained results would mean little or nothing in terms of useful information.

Assessment is a process by which information is obtained relative to some known objective or goal. Assessment is a broad term that includes testing. A test is a special form of assessment. Tests are assessments made under contrived circumstances especially so that they may be administered. In other words, all tests are assessments, but not all assessments are tests. We test at the end of a lesson or unit. We assess progress at the end of a school year through testing, and we assess verbal and quantitative skills through such instruments as the SAT and GRE. Whether implicit or explicit, assessment is most usefully connected to some goal or objective for which the assessment is designed. A test or assessment yields information relative to an objective or goal. In that sense, we test or assess to determine whether or not an objective or goal has been obtained. Assessment of skill attainment is rather straightforward. Either the skill exists at some acceptable level or it doesn't. Skills are readily demonstrable. Assessment of understanding is much more difficult and complex. Skills can be practiced; understandings cannot. We can assess a person's knowledge in a variety of ways, but there is always a leap, an inference that we make about what a person does in relation to what it signifies about what he knows. In the section on this site on behavioral verbs, to assess means to stipulate the conditions by which the behavior specified in an objective may be ascertained. Such stipulations are usually in the form of written descriptions.

Evaluation is perhaps the most complex and least understood of the terms. Inherent in the idea of evaluation is "value." When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures, and so on. When we evaluate, we are saying that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made. For example, I often ask my students if they wanted to determine the temperature of the classroom they would need to get a thermometer and take several readings at different spots, and perhaps average the readings. That is simple measuring. The average temperature tells us nothing about whether or not it is appropriate for learning. In order to do that, students would have to be polled in some reliable and valid way. That polling process is what evaluation is all about. A classroom average temperature of 75 degrees is simply information. It is the context of the temperature for a particular purpose that provides the criteria for evaluation. A temperature of 75 degrees may not be very good for some students, while for others, it is ideal for learning. We evaluate every day. Teachers, in particular, are constantly evaluating students, and such evaluations are usually done in the context of comparisons between what was intended (learning, progress, behavior) and what was obtained. When used in a learning objective, the definition provided on the ADPRIMA site for the behavioral verb evaluate is: To classify objects, situations, people, conditions, etc., according to defined criteria of quality. Indication of quality must be given in the defined criteria of each class category. Evaluation differs from general classification only in this respect.

To sum up, we measure distance, we assess learning, and we evaluate results in terms of some set of criteria. These three terms are certainly share some common attributes, but it is useful to think of them as separate but connected ideas and processes.