

[Home](#)

A Primer: Diagnostic, Formative, & Summative Assessment

Richard Swearingen, Heritage University, 2002

Introduction

According to Kellough and Kellough, "Teaching and learning are reciprocal processes that depend on and affect one another. Thus, the assessment component deals with how well the students are learning and how well the teacher is teaching" (1999, p. 417). It is the "assessment component" of teaching and learning that this paper will address. In particular, the distinctions between diagnostic, formative, and summative assessment methods will be discussed.

The Purpose of Assessment

Before addressing the different *types* of assessment, it is instructive to delineate assessment's *purposes*. Kellough et al (p. 418-419) characterizes seven purposes of assessment:

- To assist student learning.
- To identify students' strengths and weaknesses.
- To assess the effectiveness of a particular instructional strategy.
- To assess and improve the effectiveness of curriculum programs.
- To assess and improve teaching effectiveness.
- To provide data that assist in decision making
- To communicate with and involve parents.

Principles That Guide Good Assessment

According to Kellough et al (1999):

Because the welfare and, indeed, the future of so many people depend on the outcomes of assessment, it is impossible to overemphasize its importance. For a learning endeavor to be successful, the learner must have answers to basic questions: Where am I going? Where am I now? How do I get where I am going? How will I know when I get there? Am I on the right track for getting there? These questions are integral to a good program of assessment (pg. 419).

In view of such questions, it is requisite to establish principles that will guide assessment's implementation. Towards such an end, the American Association for Higher Education (AAHE) has established the following nine principles (Pausch & Popp, 1997, *Assessment in Higher Education*, ¶ 1):

- The assessment of student learning begins with educational values.
- Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time.
- Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes.
- Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes.
- Assessment works best when it is ongoing, not episodic.
- Assessment fosters wider improvement when representatives from across the educational community are involved.
- Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about [*sic*].
- Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.
- Through assessment [*sic*] educators meet responsibilities to students.

Diagnostic Assessment

Having defined the purpose and principles that drive assessment, this paper's focus will shift to distinguishing the different types of assessment, beginning with diagnostic assessment.

Although some authors delineate diagnostic assessment as a component of formative assessment, most consider it a distinct form of measurement (Kellough et al, 1999; McMillan, 2000). In practice, the purpose of diagnostic assessment is to ascertain, prior to instruction, each student's strengths, weaknesses, knowledge, and skills. Establishing these permits the instructor to remediate students and adjust the curriculum to meet each pupil's unique needs.

For example, the Heritage College Mathematics Department administers a diagnostic test to all Math 98, 99, and 101 students during the first week of each new semester. Math 98 students take an exam that covers Math 97 material; Math 99 students take one that covers Math 97 and 98 materials; and Math 101 students take one that covers Math 97,

98, and 99 materials. Based on the diagnostic exam results, each student's deficiencies are determined, and each student is subsequently required to complete a computer-based mathematics tutorial program that is tailored to his or her specific difficulties. Because the primary purpose of the diagnostic test is remediation, it is both un-graded and low-stakes.

Formative Assessment

According to the NCTM's *Principles and Standards for School Mathematics* (2000):

Assessment should be more than merely a test at the end of instruction to see how students perform under special conditions; rather, it should be an integral part of instruction that informs and guides teachers as they make instructional decisions. Assessment should not merely be done *to* students; rather, it should also be done *for* students, to guide and enhance their learning (The Assessment Principle, ¶ 1).

Assessment of the type broached in the preceding excerpt is known as formative assessment. Formative assessment should occur regularly throughout the instructional process and, According to the National Center for Fair and Open Testing (NCFOT) (1999), it "occurs when teachers feed information back to the students in ways that enable the student to learn better, or when students can engage in a similar, self-reflective process" (¶ 4). In its purist form, formative tests are not graded and are used as an ongoing diagnostic tool; hence, the instructor employs the results of formative assessment solely to modify and adjust his or her teaching practices to reflect the needs and progress of his or her students. However, formative assessment in its purist form is seldom used (Brookhart, 1999), a fact which led the NCFOT to conclude, "Most teachers do not know well how to engage in such assessment" (¶ 5).

Nevertheless, there is research evidence that indicates the efficacy of formative assessment. For example:

Black and William report that studies of formative assessment show an effect size on standardized tests of between 0.4 and 0.7, larger than most known educational interventions. (The effect size is the ratio of the average improvement in test scores in the innovation to the range of scores of typical groups of pupils on the same tests. . . . Formative assessment is particularly effective for students who have not done well in school, thus narrowing the gap between low and high achievers while raising overall achievement (NCFOT, 1999, ¶ 4).

Hence, if indeed "the primary purpose of assessment is to support high-quality learning" (NCFOT, 1999, ¶ 3), it is incumbent upon individual educators to investigate formative assessment practices and their classroom applications.

Summative Assessment

The final type of assessment that this essay addresses is the summative assessment. Black (1998, as cited by Brookhart, 1999), explaining summative assessment via analogy, stated, "When the cook tastes the soup, that's formative assessment; when the customer tastes the soup, that's summative assessment" (Formative and Summative, ¶ 1). Succinctly, summative assessment is a test, usually given at the end of a term, chapter, semester, year, or the like, the purpose of which is evaluative; in addition, high-stakes tests such as ACT, GRE, SAT, and the WASL are also examples of summative assessments.

Although "quality summative information can . . . shape how teachers organize their course, shape how teachers organize their courses or what schools offer their students" (NCFOT, 1999, ¶ 3), there is evidence that summative assessments, such as standardized exams, can adversely affect students (Phi Delta Kappan, October 1998, as cited by NCFOT, 1999). Nevertheless, especially given the current, artificially stratified, nature of the K-12 system, summative assessment is unavoidable. Hence, it is incumbent upon educators to minimize any adverse effects that such assessment might have on their students.

In order to maximize the efficacy of summative (and formative) assessment, the following factors must be considered: authenticity (Brookhart, 1999; Kellough et al, 1999), variety (Kellough et al, 1999; NREL, 2000), volume (Kellough, 1999; NCFOT, 1999), validity (Brookhart, 1999; Kellough, 1999), and reliability (Kellough, 1999).

Authenticity

Assessment that is aligned with the classroom objectives and that reflects real-world applications is called authentic assessment. In providing examples of authentic assessment, Kellough et al (1999) wrote that:

In English/language arts . . . although it may seem fairly easy to develop a criterion-referenced test, administer it, and grade it, tests often measure language *skills* rather than language use. It is extremely difficult to measure students' communicative competence with a test. . . . An authentic assessment of punctuation, then, would be an assessment of a performance item that involves students in writing and punctuating their own writing. For the authentic assessment of the student's understanding of that which the student has been learning, you would use a performance-based assessment procedure, that is, a procedure that requires students to produce rather than to select a response.

Variety

Another method of insuring quality assessment is to use a variety of assessment techniques. Traditionally, true/false and selected-response test items have been popular

methods of assessing students. However, these are limited in scope and typically test each student's capacity for rote memorization. However, assessment should include all three domains of learning; cognitive, affective, and psychomotor; in addition, assessments of the cognitive domain should reflect, at least partially, its higher levels, such as synthesis and evaluation. Hence, instructors should use a variety of assessment techniques, such as portfolios, cooperative research projects, papers, and performance tests.

Variety, in addition to permitting an instructor to assess each of the domains of learning, is also a method for minimizing assessment bias against at-risk groups. For example, "Decoding the language of a paper-and-pencil test can hinder language-minority students from demonstrating what they know. Teachers will want to use a variety of assessment methods to provide a more complete picture of students' progress and areas of need" (NREL, 2000, Assessment, ¶ 1).

Volume

Unfortunately, teachers often require more summative assessments than are necessary. According to the Academic Board of the University of Queensland (1997):

The quantity of assessment which contributes toward the final result need only be the minimum amount necessary to ensure a valid result. . . . Students resent over-assessment which often occurs across their course because each subject teacher believes his/her workload is reasonable. Large amounts of assessment also take their toll on staff, especially in terms of setting and marking. *It is not surprising that examiners may be tempted to set assessment with more regard for ease of marking than for educational benefit* [emphases added] (Amount, ¶ 1 & 3).

Validity

According to Crooks (1988, as cited by the University of Queensland, 1997), "The validity of assessment refers to 'the extent to which the assessment measures performance on the aspects of the course which are important'" (Amount, ¶ 1). Hence, to use a grossly oversimplified example, if one desires to measure a student's progress in English, one cannot give him or her a mathematics exam. In some sense, validity and authenticity are synonymous; to be valid, a test must explicitly reflect and measure the course's objectives. Simply put, if one desires to measure the speed of one's car, one uses the speedometer, not the fuel gauge.

Reliability

The final issue to discuss is that of reliability. Suppose that one wishes to determine how fast one's car is traveling; a valid assessment tool is the speedometer. However, if when one's car travels at 60 mph the speedometer registers 50 mph, the speedometer is not

reliable. Similarly, a test can be valid, but not reliable. However, instructors can facilitate reliability; for example, using a rubric—and allowing the students to have copies of it—is one method for ensuring reliability when grading a research paper. If the rubric is well designed, several instructors could score the same paper using the rubric and arrive at similar scores. Hence, the hallmark of assessment reliability is the *reproducibility* of assessment results.

Summary

It is evident that assessment—diagnostic, formative, and summative—is a critical component of education. Hence, it is incumbent upon educators to utilize assessment in an effective manner, keeping in mind the purposes of and principles behind it. In particular, it is especially crucial that they investigate and utilize diagnostic and formative assessment, both of which are underused—yet effectual—components of the educational process.

References

Assessment Resource Bank. (2001). Formative Assessment. Retrieved October 29, 2001, from <http://arb.nzcer.org.nz/nzcer3/furform.htm>

Brookhart, S.M. (1999). Wiggins, Grant. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass. Retrieved October 30, 2001, from <http://www.coe.asu.edu/edrev/reviews/rev50.htm>

Kellough, R.D. and Kellough, N.G. (1999). *Secondary School Teaching: A Guide To Methods And Resources; Planning For Competence*. Copyright by Prentice Hill, Upper Saddle River, New Jersey.

McMillan, J.H. (2000). *Classroom Assessment: Principles and Practice for Effective Instruction*. Pearson Technology Group. Retrieved October 30, 2001, from http://www.pearsonptg.com/book_detail/0,3771,020529751X,00.html

The National Center for Fair and Open Testing. (1999). *The Value of Formative Assessment*. *Fair Test Examiner*. Retrieved October 29, 2001, from <http://www.fairtest.org/examarts/winter99/k-forma3.html>

NCTM. (2000). *Principles and Standards for School Mathematics*. Retrieved October 29, 2001, from <http://standards.nctm.org/document/chapter2/assess.htm>

NREL. (2000). *Linking Second-Language Strategies with Content Instruction*. Retrieved June 20, 2001, from http://www.nwrel.org/just_good/8/chapter2.html

Pausch, L.M. and Popp, M.P. (1997). *Assessment of Information Literacy: Lessons from the Higher Education Assessment Movement*. Retrieved October 30, 2001, from <http://www.ala.org/acrl/paperhtm/d30.html>

University of Queensland. (1997). Guidelines on the Amount, Mode and Spread of Assessment. Retrieved October 30, 2001, from http://www.admin.uq.edu.au/AcadBoardOffice/policy/am_mod_spr_html

