

ASSESSMENT TERMINOLOGY
Division of Student Affairs
University of North Carolina at Chapel Hill

Note: Listed below is a glossary of terms and concepts commonly used in assessment and research literatures. Some of these terms may appear to be very simple while others appear to be more complex. This list is not meant to be comprehensive of all terminology used in these fields but is meant to be representative of those concepts and terms commonly used and perhaps most relevant to your department's assessment process. We hope this document will provide you with a common language for interacting with colleagues across campus. The numbers following each definition refers to the citation source listed at the end of the document in the reference section.

Assessment: an effort to gather, analyze and interpret evidence that describes institutional, departmental, divisional, or program effectiveness. (21)

Academic program review: the periodic peer evaluation of the effectiveness of an educational degree program usually encompassing student learning, faculty research, scholarship, and service, and assessment resources. (6)

Accountability: use of results for program continuance/discontinuance; the public reporting of student, program, or institutional data to justify decisions or policies; using results to determining funding. (6)

Assessment plan: a document that outlines the student learning outcomes (for academic programs) or unit outcomes (for support units), the direct and indirect assessment methods used to demonstrate the attainment of each outcome, a brief explanation of the assessment methods, an indication of which outcome(s) is/are addressed by each method, the intervals at which evidence is collected and reviewed, and the individual(s) responsible for the collection/review of evidence. (2)

Behavioral observations: measuring the frequency, duration, typology, etc. of student actions, usually in a natural setting with non-interactive methods, for example, formal or informal observations of a classroom. Observations are most often made by an individual and can be augmented by audio or videotape. (2)

Benchmark: a criterion-referenced objective; "Performance data that are used for comparative purposes. A program can use its own data as a baseline benchmark against which to compare future performance. It can also use data from another program as a benchmark. In the latter case, the other program often is chosen because it is exemplary and its data are used as a target to strive for, rather than as a baseline." (p. 35, 10)

Closing the loop: using assessment results for program change and improvement. (6)

Commercial, norm-referenced, standardized exams: group administered, mostly or entirely multiple-choice, "objective" tests in one or more curricular areas. Scores are based on comparison with a reference or norm group. Typically must be purchased from a private vendor. (2)

Construct: an abstract or general idea inferred or derived from specific instances (2)

Construct validity: the degree to which the degree to which a test measures what it claims, or purports, to be measuring (13)

Content analysis: a qualitative methodology in which researchers analyze written material in a way that which will produce reliable evidence about a large sample. To do this, researchers establish a set of categories and then count the number of instances that fall into each category. The crucial requirement is that the categories are sufficiently precise to enable different coders to arrive at the same results when the same body of material is examined. (18)

Content validity: a test has content validity if the item measures what it is intended to measure. (12)

Continuous Quality Improvement (CQI): taken from the business world, CQI is a methodology intent on meeting or exceeding customer requirements by continuous improvement and innovation in products, processes, and services. Marchese (cited in Palomba & Banta, 1999) notes that "the quality movement brings with it a sense of collective responsibility for learning, a habit of listening to the people we serve, a preference for data, an ethic of continuous improvement, a determination to develop fully the talent of every learner, and an acknowledgement that we are professionally accountable to one another and to those we serve for results." (p.11, 15)

Convergent validity: general agreement among ratings, gathered independently of one another, where measures should be theoretically related. (2)

Course-embedded assessment: collecting assessment data information within the classroom because of the opportunity it provides to use already in-place assignments and coursework for assessment purposes. This involves taking a second look at materials generated in the classroom so that, in addition to providing a basis for grading students, these materials allow faculty to evaluate their approaches to instruction and course design. (15)

Criterion-referenced: criterion-referenced tests determine what test-takers can do and what they know, not how they compare to others. Criterion-referenced tests report on how well students are doing relative to a predetermined performance level on a specified set of educational goals or outcomes included in the curriculum. (2)

Culture of assessment: an environment in which continuous improvement through assessment is expected and valued. (6)

Culture of evidence: an environment in which the use of research and/or assessment results to guide policy decisions is expected and valued. (14)

Direct measures: direct measures of student learning require student to display their knowledge and skills as they respond to the instrument itself. Objective tests, essays, presentations, and classroom assignments all meet this criterion. (15)

Embedded assessment: including questions from assessment instruments or selecting questions from existing tests of existing courses; paucity of number of questions can affect reliability. (24)

Evaluation: any effort to use this evidence to improve effectiveness. (21)

Exit and other interviews: asking individuals to share their perceptions of their own attitudes and/or behaviors or those of others, evaluating student reports of their attitudes and/or behaviors in a face-to-face-dialogue. (2)

External assessment: use of criteria (rubric) or an instrument developed by an individual or organization external to the one being assessed. Usually summative, quantitative, and often high-stakes (see below). Example: GRE exams. (2)

External validity: refers to the extent to which the results of a study are generalizable or transferable to other settings. Generalizability is the extent to which assessment findings and conclusions from a study conducted on a sample population can be applied to the population at large. Transferability is the ability to apply the findings in one context to another similar context. (2)

Face validity: the extent to which an instrument superficially appears, to an examinee, to measure the construct being investigated. (8)

Focus groups: typically conducted with 7-12 individuals who share certain characteristics that are related to a particular topic related a research or evaluation question. Group discussions are conducted by a **trained** moderator with participants (several times, if possible) to identify trends/patterns in perceptions. Moderator's purpose is to provide direction and set the tone for the group discussion, encourage active participation from all group members, and manage time. Moderator must not allow own biases to enter, verbally or nonverbally. Careful and systematic analysis of the discussions provides information that can be used to evaluate and/or improve the desired outcome. (2)

Formative assessment: the gathering of information about student learning during the progression of a course or program and usually repeatedly-to improve the learning of those students. Assessment feedback is short term in duration. Example: reading the first lab reports of a class to assess whether some or all students in the group need a lesson on how to make them succinct and informative. (2)

“High stakes” use of assessment: the decision to use the results of assessment to set a hurdle that needs to be cleared for completing a program of study, receiving certification, or moving to the next level. Most often the assessment so used is externally developed, based on set standards, carried out in a secure testing situation, and administered at a single point in time. Examples: at the secondary school level, statewide exams required for graduation; in postgraduate education, the bar exam. (2)

Indirect measures: indirect methods such as surveys and interviews ask students to reflect on their learning rather than to demonstrate it. (15)

Institutional effectiveness: broadly refers to the extent to which an institution meets its unique mission. Typically includes functions associated with assessment and institutional research. (11)

Inter-rater reliability: the degree to which different raters/observers give consistent estimates of the same phenomenon. (2)

Interviewing: asking questions in a systematic way as to determine another's perspective. (16)

Internal validity: refers to (1) the rigor with which the study was conducted (e.g., the study's design, the care taken to conduct measurements, and decisions concerning what was and wasn't measured) and (2) the extent to which the designers of a study have taken into account alternative explanations for any causal relationships they explore. (2)

Learning: the process by which a relatively lasting change in potential behavior occurs as a result of practice or experience. (19)

Learning outcomes: refers to the specific knowledge or skills that students actually develop through their college experience. (6)

Longitudinal study: a study that investigates development, learning, or other types of change in individuals over time. (6)

Measurement: refers to the two over-arching but not mutually exclusive methods used to conduct assessments: quantitative and qualitative. (21)

Meta-analysis: a systematic way of compiling results across studies in order to clarify the findings in an area of research. (7)

Norms: Statistics that describe the test performance of specified groups, such as pupils of various ages or grades in the standardization group for a test. Norms are often assumed to be representative of some larger population, as of pupils in the country as a whole. (2)

Norm-reference: a norm-referenced test is designed to highlight achievement differences between and among studies to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers. (2)

Observer effect: the degree to which the assessment results are affected by the presence of an observer. (2)

Objective measures: tests or measures that can be scored without need for subjective judgment (e.g., multiple-choice format). (19)

Objectives: refers to the specific knowledge, skills, or attitudes that students are expected to achieve through their college experience; expected or intended student outcomes. (6)

Ongoing assessment: tracks student learning by performance on tasks that are part of the natural Lesson Plan process. Ongoing assessment, producing samples of work clustered by the outcome demonstrated, provides accumulated data for summative analysis. (6)

Open-ended: assessment questions that are designed to permit spontaneous and unguided responses. (2)

Operational (--ize): defining a term or object so that it can be measured. Generally states the operations or procedures used that distinguish it from others. (2)

Outcome measure: instruments used for gathering information on student learning and development. (6)

Outcomes: refers to the specific knowledge, skills, or developmental attributes that students actually develop through their college experience; assessment results. (6)

Performance assessment: assessment technique involving the gathering of data through systematic observation of a behavior or process and evaluating that data based on a clearly articulated set of performance criteria to serve as the basis for evaluative judgments. (23)

Portfolios: collections of multiple student work samples usually compiled over time and rated using rubrics. The design of a portfolio is dependent upon how the scoring results are going to be used. (2)

Portfolio assessment: a portfolio becomes a portfolio assessment when (1) the assessment purpose is defined; (2) criteria are made clear for determining what is contained in the portfolio, by whom, and when; and (3) criteria for assessing either the collection or individual pieces of work are identified and used to make judgments about performance. Portfolios can be designed to assess student progress, effort, and/or achievement, and encourage students to reflect on their learning. (6)

Predictive validity: the ability of a measure to predict some future behavior. (3)

Program review: many different perceptions and models exist of program review. In higher education, a program review could consist of anything from a judgment by experts of the worth of the curriculum based primarily on observations to a comparison between student performance indicators and objectives and actual student performance on either standardized or evaluator-constructed measures. (25)

Qualitative assessment: Collects data that does not lend itself to quantitative methods but rather to interpretive criteria. (2)

Quantitative assessment: Collects data that can be analyzed using quantitative methods. (2)

Questionnaire: a list of questions concerning a particular topic administered to a group of individuals to obtain information concerning their preferences, beliefs, interests, and behavior. (4)

Reliability: is the extent to which an experiment, test or any measuring procedure yields the same result on repeated trials. (2)

Research: is about studying, developing or testing a theory by gathering data in a systematic way. (21)

Rubrics: a rubric is a set of categories that define and describe the important components of the work being completed, critiqued or assessed. Each category contains a graduation of levels of completion or competence with a score assigned to each level and a clear description of what criteria need to be met to attain the score at each level. (2)

Self-assessment: the process of evaluating one's own learning. The process often includes the ability to judge one's own achievements and performances, understanding how the product or performance was achieved, understanding why one followed the process he or she did, and understanding what might be done to improve the process, product or performance. (5)

Self-report: student evaluation of their own learning or abilities. (6)

Standard: A pre-determined criterion or expectation of a level of student learning; a passing score. (9)

Standardization: a consistent set of procedures for designing, administering, and scoring a measure. The purpose of standardization is to assure that all students are evaluated under the same conditions so that their scores/ratings have the same meaning and are not influenced by differing conditions. Standardized procedures are very important when scores/ratings will be used to compare individuals or groups. (6)

Statistically significant: The probability that the association between the factor and the outcome or the difference between the means of two samples for example is due to chance is less than a specified level (by convention, $p < 0.05$).

Student outcome: the achieved results or the actual consequences of what a student has demonstrated or accomplished; maybe academic and occupational, as well as the intellectual, personal, civic development, attitudes, values, and beliefs that students attain as a result of postsecondary education. (6)

Summative assessment: assessment that is done at the conclusion of a course or some larger instructional period (e.g., at the end of the program). The purpose is to determine success or to what extent the program/project/course met its goals. (2)

Triangulate (triangulation): the use of a combination of assessment methods in a study. An example of triangulation would be an assessment that incorporated surveys, interviews, and observations. (2)

Validity: validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. Validity has three components: relevance - the option measures your educational objective as directly as possible; accuracy - the option measures your educational objective as precisely as possible; utility - the option provides formative and summative results with clear implications for educational program evaluation and improvement. (2)

Written surveys/questionnaires: asking individuals to share their perceptions about the study target-e.g. their own or others skills/attitudes/behavior, or program/course qualities and attributes. (2)

References

1. Angelo, T.A., & Cross, P.A. (1993). *Classroom assessment techniques*. San Francisco: Jossey-Bass.
2. Assessment Services of Northern Illinois University. (2005). Assessment terms. Retrieved from http://www.niu.edu/assessment/_resourc/asterms.shtml on February 22, 2005.
3. Bordens, K.S., & Abbott, B.B. (1988). *Student workbook and study guide to accompany research design and methods: A process approach*. Mountain View, CA: Mayfield.
4. Borden, V.H., & Zak-Owens, J. (2001). *Measuring quality: Choosing among surveys and other college and university quality assessments*. Tallahassee, FL: Association for Institutional Research; Washington, DC: American Council on Education.
5. Brown, S.A., & Glasner, A. (Eds.). (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. New York: Taylor & Francis.
6. Center for Assessment and Research Studies of James Madison University. (2005). Dictionary of Student Outcome Assessment. Retrieved from <http://people.jmu.edu/yangsx/> on February, 22, 2005.
7. Cooper, H., & Hedges, L.V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
8. Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.
9. Erwin, T.D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. San Francisco: Jossey-Bass.
10. Hatry, H., van Houten, T., Plantz, M., & Greenway, M.T. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
11. Komives, S.R., & Woodard, D.B., Jr. (1996). *Student services: A handbook for the profession (3rd ed.)*. San Francisco: Jossey-Bass.
12. McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
13. Mehrens, W.A., & Lehman, I.J. (1991). *Measurement and evaluation in education and psychology (4th ed.)*. Fort Worth, TX: Holt, Rinehart and Winston.
14. O'Banion, T. R. (1997). *A Learning College for the 21st Century*. Phoenix, AZ: Oryz Press.

15. Palomba, C. & Banta, T. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. Washington, DC: Higher Education.
16. Patton, M.Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.
17. Pendergrass, P.W. (1986). Oakland City College—Project QUE. In W.H. Bergquist & J.L. Armstrong (Eds.). *Planning effectively for educational quality*. Washington, DC: Jossey-Bass
18. Silverman, D. (2000). *Qualitative research: A practical handbook*. Thousand Oaks, CA: Sage.
19. Stratton, P., & Hayes, N. (1993). *A student's dictionary of psychology*. New York: Hodder and Stroughton.
20. Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
21. Upcraft, M. & Schuh, J. (1996). *Assessment in student affairs: A guide for practitioners*. NY: Jossey-Bass, Inc.
22. Waller, N. G., & Meehl, P.E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.
23. Wheeler, P., & Haertel, G. D. (1993). *Resource handbook on performance assessment and measurement: A tool for students, practitioners, and policymakers*. Berkeley, CA: The Owl Press.
24. Wilson, M. & Sloane, K. (2000). From principles to practice: an embedded assessment system. *Applied Measurement in Education* 13, (2), 181-208.
25. Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (1997). *Program evaluation: Alternative approaches and practical guidelines (2nd ed.)*. White Plains, NY: Longman.

Compiled by: Sabourin Ward & Wolf Johnson