

Glossary of Assessment Terms
IUPUI
Advanced Practices Committee

The IUPUI Glossary of Assessment Terms is designed to facilitate understanding by providing a *common language* for discussions. The terms provided should not be considered an exhaustive list. The glossary provides *broad and general* descriptions designed to define terms across academic disciplines. It is understandable that a specific discipline (e.g., Psychology, History) may define terms in a different manner.

- Assessment
- Benchmarking
- Bloom's Taxonomy
- Criterion Referenced Assessment
- Curriculum Mapping
- Descriptive Rubric
- Effectiveness Studies
 - Cost Benefit Analysis
 - Economic Modeling
 - Efficiency
 - Productivity
- Evaluation or Program Evaluation
 - Formative Evaluation
 - Summative Evaluation
 - Action Designs
 - Action Research
 - Theory-Driven Evaluation
 - Logic Modeling
 - Kirkpatrick Four Level Model
- Evaluation Methods
- Measure
 - Direct Measure
 - Authentic
 - Embedded
 - Portfolios Assessment
 - Indirect Measure
 - Focus Groups
 - Interviews
 - Questionnaires
 - Surveys
- The Higher Learning Commission (HLC)
- Institutional Research
- Norm-referenced Assessment
- North Central Association of Colleges And Schools
- Performance Measurement
- Program Review (Academic)
- Psychometric Properties
 - Correlation
 - Measurement
 - Reliability
 - Statistical Significance
 - Triangulation
 - Validity
 - Construct Validity
 - External Validity
 - Internal Validity
- Research Approaches
 - Applied Research
 - Basic Research
 - Mixed Method Research Designs
 - Pre-Post Design
 - Qualitative Research
 - Quantitative Research
- Rubric
- Scholarship of Teaching and Learning
- Student Learning Outcomes
 - Learning Goal
 - Learning Objectives
 - Learning Outcomes
- Value Added

Assessment: is the systematic collection, review, and use of information about educational programs undertaken for the purposes of improving student learning and development (Palomba & Banta, 1999). The purpose of assessment is to provide information about the student learning and development that occurs as a result of a program. A “program” may be any activity, project, function, or policy that has an identifiable purpose or set of objectives.

Benchmarking: is a method used by organizations to compare their performance, processes, or practices with peer organizations or those in other sectors. This method can focus on performance, in which case one identifies the most important indicators of success and then compares one's own performance with that of other organizations. The focus can also be a particular process, such as billing or information technology (Mathison, 2005).

Bloom's Taxonomy: a classification of educational goals and objectives created by a group of educators led by Benjamin Bloom. They identified three areas of learning objectives (domains): cognitive, affective and psychomotor. The cognitive domain is broken into six areas from less to more complex: knowledge, comprehension, application, analysis, synthesis and evaluation. The taxonomy may be used as a starting point to help one develop learning objectives.

Criterion Referenced Assessment: an assessment where an individual's performance is compared to a specific learning objective or performance standard and not to the performance of other students. Criterion referenced assessment tells us how well students are performing on specific goals or standards rather than just telling how their performance compares to a norm group of students nationally or locally (CRESST, 2011).

Curriculum Mapping: The process of aligning courses with program/major level goals and objectives, often done systematically with faculty involvement. Curriculum mapping is a process for recording what content and skills are actually taught in a classroom, school, or program.

Descriptive Rubric: A rubric with brief descriptions of the performance that merits each possible rating. They help to make faculty expectations explicit and are useful when there is more than one evaluator.

Effectiveness Studies: studies that make a determination of the effectiveness of a program, intervention, and/or policy in terms of predetermined criteria (e.g., productivity, student learning, monies, etc):

Cost Benefit Analysis: is a technique for deciding whether to make a change. As its name suggests, it compares the values of all benefits from the action under consideration and the cost associated with it. The cost-benefit ratio is determined by dividing the projected benefits of the program by the projected costs (OECD, 2011).

Cost Effectiveness: is the defining element of a method for comparing both the costs and results of different alternatives for addressing particular goals. Criteria for measuring effectiveness must be similar among alternatives for a cost-effectiveness comparison. Effectiveness estimates are based on the usual experimental, quasi-experimental, or statistical designs (Levin as cited in Mathison, 2005).

Economic Modeling: provides academic administrators with a logical framework for analyzing costs associated with the processes involved in the delivery of education. The specific costs associated with activities such as teaching, research, and service may be determined for a school as a whole or for specific responsibility centers (e.g., programs and services within the school) (Courneyer, Powers, Johnson, & Bennett, 2000).

Efficiency: the ratio or proportionality between the value of the human end achieved (benefits or satisfactions) and the value of the scarce resources expended to achieve it (opportunity costs) (Johnson, 2005).

Productivity: the value of output (goods and services) produced per unit of input (productive resources) used. Thus an increase in productivity means producing more goods and services with the same amount of resources, or producing the same goods and services with fewer resources, or some combination of these two possibilities (Johnson, 2005).

Evaluation or Program Evaluation: the collection of methods, skills, and sensitivities necessary to determine whether a program is needed and likely to be used, whether it is intense enough to meet an unmet need, whether the service is offered as planned, whether the program is provided at a reasonable cost without undesirable side effects (Posavac & Carey, 2002). Another conception of evaluation is that it investigates and judges the quality or worth of a program, project, or other entity rather than student learning. Under this definition, evaluation is a broader concept than assessment. While assessment focuses on how well student learning goals are achieved, evaluation addresses how well all the major goals of a program are achieved (Suskie, 2009).

Formative Evaluation: evaluations are intended - by the evaluator - as a basis for improvement (Scriven, 1996). Formative assessment is the process of gathering and evaluating information about student learning during the progression of a course or program of study (Lehman College, 2011). Formative evaluation is typically conducted during the development or improvement of a program or product and it is conducted, often more than once, for in-house staff of the program with the intent to improve. The reports normally remain in-house; but serious formative evaluation may be done by an internal or an external evaluator or preferably, a combination; of course, many program staff are, in an informal sense, constantly doing formative evaluation (Scriven, 1991).

Summative Evaluation: Summative evaluation provides information on the product's efficacy (its ability to do what it was designed to do). For example, did the learners learn what they were supposed to learn after using the instructional module. In a sense, it lets the learner know "how they did," but more importantly, by looking at how the learner's did, it helps you know whether the product teaches what it is supposed to teach.

Action Research: Action research is an effective inquiry-based method for promoting collaboration between researchers and program practitioners in the analysis and subsequent improvement of academic program outcomes and processes. Action research provides a constructive framework for ensuring that critical information is used by key stakeholders to implement data-driven interventions for continuous academic improvement (Hansen & Borden, 2006).

Action Designs: are systematic procedures used by teachers, or other individuals in an educational setting, to gather quantitative, qualitative (or both) data about, and subsequently improve, the ways their particular setting operates, how they teach, and how well their students learn (Creswell, 2008).

Theory-Driven Evaluation: is a contextual or holistic assessment of a program based on the conceptual framework of program theory. The purpose of theory-driven evaluation is to provide information on not only the performance or merit of a program but on how and why the program achieves such a result (Mathison, 2005).

Logic Modeling: logic models are well suited for defining and evaluating how inputs / investments (e.g., time, money, resources) into a program effort result in outputs (e.g. documents and workshops), that can support desired outcomes (e.g., learning and increased productivity) (W. K. Kellogg Foundation, 2004).

Kirkpatrick Four Level Model: An approach to the evaluation of training in organizations. This evaluation model delineates four levels of training outcomes: reaction, learning, behavior, and results (Owen, 2005).

Evaluation Methods (Direct & Indirect):

Measures: Measurement refers to the process by which the attributes or dimensions of some physical object (e.g., student) are determined. In the context of assessment of student learning or development measurement can involve a combination of qualitative and quantitative information to determine levels or qualities of student learning and development. The word measure is also intended may address the type or level of program activities conducted (process), the direct products and services delivered by a program (outputs), and/or the results of those products and services (outcomes).

Direct Measures Direct measures require students to demonstrate their knowledge and skills. They provide tangible, visible and self-explanatory evidence of what students have and have not learned as a result of a course, program, or activity (Suskie, 2004, 2009; Palomba and Banta, 1999).

Authentic: based on examining genuine or real examples of students' work. Work that closely reflects goals and objectives for learning. Authentic assessment reveals something about the standards that are at the heart of a subject; asking students to use judgment and innovation as they "do" and explore the subject. (Wiggins, 1989, 1990 as in Palomba & Banta, 1999).

Embedded: program, general education, or institutional assessments that are embedded into course work. In other words, they are course assessments that do double duty, providing information not only on what students have learned in the course but also on their progress in achieving program or organizational goals. Because embedded assessment instruments are typically designed by faculty and staff, they match up well with local learning goals. They therefore yield information that faculty and staff value and are likely used to improve teaching and learning (Suskie, 2009).

Portfolios Assessment: a type of performance assessment in which students' work is systematically collected and reviewed for evidence of student learning. In addition to examples of their work, most portfolios include reflective statements prepared by students. Portfolios are assessed for evidence of student achievement with respect to established student learning outcomes and standards (Palomba & Banta, 1999).

Indirect Measures: Assessments that measure opinions or thoughts about students' or alumni's own knowledge, skills, attitudes, learning experiences, perception of services received or employers' opinions. While these types of measures are important and necessary they do not measure students' performance directly. They supplement direct measures of learning by providing information about how and why learning is occurring (Hansen, 2011).

Focus Groups: a group selected for its relevance to an evaluation that is engaged by a trained facilitator in a series of discussions designed for sharing insights, ideas, and observations on a topic of concern to the evaluation (National Science Foundation, 2010).

Interviews: occur when researchers ask one or more participants general, open-ended questions and records their answers (Creswell, 2008).

Questionnaires: are forms used in a survey design that participants in a study complete and return to the researcher. Participants mark answers to questions and may supply basic, personal, or demographic information about themselves (Creswell, 2008).

Surveys: A survey is a method of collecting information from people about their characteristics, behaviors, attitudes, or perceptions. Surveys most often take the form of questionnaires or structured interviews (Palomba & Banta, 1999). General definition: an attempt to estimate the opinions, characteristics, or behaviors of a particular population by investigation of a representative sample.

Institutional Research: provides fundamental support for a campus, school, and program planning and evaluation activities by: developing for academic deans and other campus administrators a series of management reports and analyses that integrate information from a variety of institutional and external data sources. (Indiana University, 2011).

The Higher Learning Commission (HLC) is an independent corporation and one of two commission members of the North Central Association of Colleges and Schools (NCA), which is one of six regional institutional accreditors in the United States. The Higher Learning Commission accredits degree-granting post-secondary educational institutions in the North Central region. The Commission accredits more than 1,000 colleges and universities in nineteen states. The states are Arkansas, Arizona, Colorado, Iowa, Illinois, *Indiana*, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, Ohio, Oklahoma, New Mexico, South Dakota, Wisconsin, West Virginia, and Wyoming.

Norm Referenced Assessment: An assessment where student performance is compared to a larger group. Usually the larger group or “norm group” is a national sample representing a wide diverse cross-section of students. Students, schools, districts, and even states are compared or rank-ordered in relation to the norm group (CREST, 2011).

North Central Association of Colleges And Schools: The purpose of the Association shall be to require its Commission members to have accrediting processes that foster quality, encourage academic excellence, and improve teaching and learning. The Association shall also encourage and support cooperative relationships among schools, and colleges and universities that hold membership in the Association.

Performance Measurement: The ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals. It is typically conducted by program or agency management. Performance measures may address the type or level of program activities conducted (process), the direct products and services delivered by a program (outputs), or the results of those products and services (outcomes). A “program” may be any activity, project, function, or policy that has an identifiable purpose or set of objectives (Government Accountability Office, 2011).

Program Review (Academic): The periodic peer evaluation of the effectiveness of an educational degree program usually encompassing student learning, faculty research, scholarship, and service, and assessment resources. While program (or peer) review and evaluation have similar meanings, program review is a term used almost exclusively in higher education, while program evaluation tends to be used in the P-12 education, business, and not-for-profit sectors. Academic program review at IUPUI involves the following processes:

1. A *planning committee establishes the purposes* for review and includes the school dean, the department chair, the executive vice chancellor, the director of program review, and the director of the graduate school if graduate programs are involved.
2. A *Self Study* is developed by the unit using our guidelines (see ‘IUPUI Program Review’ at <http://www.planning.iupui.edu/assessment/>) and customized data reports that we provide.

3. The **review team visit** (composed of internal and external stakeholders and experts) typically encompasses 2 ½ days and includes meetings with the dean, the chair, faculty, staff, students, and alumni.
4. The final written report is disseminated to the chancellor, executive vice chancellor, dean, and department chair.
5. The chair works with colleagues to **draft a response** to the report and sends this report to the director of program review within six months.
6. A **follow-up meeting** is scheduled that includes the executive vice chancellor, the director of program review, the director of the graduate office
7. The **department chair is invited to PRAC** to report on progress in the unit since the review as well as to comment on the quality of the review process itself.

Psychometric Properties

Correlation: is a measure of the degree or strength of relationship between two or more variables. It does not prove causation because we may not know which variable came first or whether alternative explanations for the presumed effect exist. (Munoz as cited in Mathison, 2005).

Measurement: may be defined as the set of rules for transforming behaviors into categories or numbers. Constructing an instrument to measure a social science variable involves several steps, including conceptualizing the behaviors that operationally define the variable, drafting items that indicate the behaviors administering draft items to try out samples, refining the instrument based on item analysis, and performing reliability and validity studies (Petrosko as cited in Mathison, 2005).

Reliability: there are several definitions of *reliability* for measures of individual differences, each stressing one of its many facets. It may be defined as the accuracy or precision of an instrument in measuring the "true" standing of persons to whom a particular test was administered. Another definition states that reliability is an estimate of the stability, dependability, or predictability of a measure. Finally, reliability is defined as the proportion of observed score variance that is attributable to true score variance (Thomas as cited in Mathison, 2005).

Reliability (simplified): As applied to an assessment tool, it refers to the extent to which the tool can be counted on to produce consistent results over time.

Types of Reliability

Test-retest: A reliability estimate based on assessing a group of people twice and correlating the two scores.

Parallel forms: A reliability estimate based on correlating scores collected using two versions of the procedure.

Inter-rater: How well two or more raters agree when decisions are based on subjective judgments

Internal Consistency: A reliability estimate based on how highly parts of a test correlate with each other.

Coefficient Alpha: An internal consistency reliability estimate based on correlations among all items on a test

Split-half: An internal consistency reliability estimate based on correlating two scores, each calculated on half of a test

Statistical Significance: is a mathematical procedure for determining whether a null hypothesis can be rejected at a given alpha level. Tests of statistical significance play a large role in quantitative research designs but are frequently misinterpreted. The most common misinterpretation of the test of significance is to confuse statistical significance with the practical significance of the research results. (Munoz as cited in Mathison, 2005).

Triangulation: is the process of corroborating evidence from different individuals (e.g., a principal and a student), types of data (e.g., observational field notes and interviews), or methods of data collection (e.g., documents and interviews) or descriptions and themes in qualitative research (Creswell, 2008).

Validity: means that researchers can draw meaningful and justifiable inferences from scores about a sample or population (Creswell, 2002). The extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate (CRESST, 2011). Types of validity include construct, external, and internal.

Validity (simplified): As applied to an assessment tool, it refers to a judgment concerning the extent to which the assessment tool measures what it purports to measure. The validity of a tool can never be proved absolutely; it can only be supported by an accumulation of evidence from several categories

Types of Validity

Construct: Examined by testing predictions based on the theory (or construct) underlying the procedure.

Criterion-related: How well the results predict a phenomenon of interest, and it is based on correlating assessment results with this criterion

Face: Subjective evaluation of the measurement procedure. This evaluation may be made by test takers or by experts for improving what is being assessed.

Formative: How well an assessment procedure provides information that is useful for improving what is being assessed.

Sampling: How well the procedures components, such as test items, reflect the full range of what is being assessed

Research (Approaches): are procedures for collecting, analyzing, and reporting research in quantitative and qualitative formats (Creswell, 2008).

Applied Research: applied research is an original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective (OECD, 2011).

Basic Research: basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view (OECD, 2011).

Note: Applied research contrasts with basic research, which has the purpose of addressing fundamental questions with wide generalizability; for example, testing a hypothesis derived from a theory in economics. Both applied and basic research can use any of the social science research methods, such as the survey, experimental, and qualitative methods. Differences between the research roles do not relate to methods of inquiry; they relate to the purpose of the investigation. Applied researchers focus on concrete and practical problems; basic researchers focus on problems that are more abstract and less likely to have immediate application (Petrosko as cited in Mathison, 2005).

Mixed Method Research Designs: are procedures for collecting both quantitative and qualitative data in a single study, and for analyzing and reporting this data based on a priority and sequence of the information (Creswell, 2008).

Pre-Post Design: is a method for assessing the impact of an intervention by comparing scores on variables before and after an intervention occurs. The simplest type of this design involves one group – for example, program participants in a summative evaluation. Validity of the design is enhanced by adding a control group whose members do not experience the intervention and by randomly assigning persons to treatment and control conditions. The more valid the design, the greater confidence of the evaluator in making decisions about the efficacy of an intervention (Petrosko as cited in Mathison, 2005).

Qualitative Research: is an inquiry approach useful for exploring and understanding a central phenomenon. To learn about this phenomenon, the inquirer asks participants broad, general questions, collects the detailed views of the participants in the form of words or images, and analyzes the information for description and themes. From this the researcher interprets the meaning of the information drawing on personal reflections and past research. The final structure of the report is flexible, and it displays the researcher's biases and thoughts (Creswell, 2002).

Quantitative Research: is an inquiry approach useful for describing trends and explaining the relationship among variables found in the literature. To conduct this inquiry the investigator specifies narrow questions, locates or develops instruments to gather data to answer the questions, and analyzes numbers from the instruments, using statistics. From the results of these analyses, the researcher interprets the data using prior predictions and research studies. The final report, presented in a standard format, displays researcher objectivity and lack of bias (Creswell, 2002).

Rubric: A set of categories that define and describe the important components of the work being completed, critiqued, and assessed. Each category contains a gradation 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.

Scholarship of Teaching and Learning: is first and foremost a commitment to the improvement of student learning, made possible through individual and collective knowledge-building; is rigorous and thoughtful investigation of student learning, with the results made available for public review and use beyond a local setting (Cambridge,1999).

Student Learning Outcomes: specify what students will know, be able to do, or be able to demonstrate when they have completed or participated in academic program(s) leading to certification or a degree. Outcomes are often expressed as knowledge, skills, attitudes, behaviors, or values. A multiple methods approach is recommended to assess student learning outcomes indirectly and directly. Direct measures of student learning require students to demonstrate their knowledge and skills. They provide tangible, visible and self-explanatory evidence of what students have and have not learned as a result of a course, program, or activity (Suskie, 2009; Palomba & Banta, 1999).

Learning Goal: A broad statement of desired outcomes – what we hope students will know and be able to do as a result of completing the program/course. They should highlight the primary focus and aim of the program. They are not directly measurable; rather, they are evaluated directly or indirectly by measuring specific objectives related to the goal.

Learning Objective: Sometimes referred to as intended learning outcomes, student learning outcomes (SLOs) or outcomes statements. Learning objectives are clear, brief statements used to describe specific measurable actions or tasks that learners will be able to perform at the conclusion of instructional activities. Learning objectives focus on student performance. Action verbs that are specific, such as list, describe, report, compare, demonstrate, and analyze, should state the behaviors students will be expected to perform. Verbs that are general and open to many interpretations such as understand, comprehend, know, appreciate should be avoided.

Learning Outcomes: The learning results—the end results—the knowledge, skills, attitudes and habits of mind that students have or have not taken with them as a result of the students' experience in the course(s) or program.

Value Added: the increase in learning that occurs during a course, program, or undergraduate education. Can either focus on the individual student (how much better a student can write, for example, at the end than at the beginning) or on a cohort of students (whether senior papers demonstrate more sophisticated writing skills in the aggregate than freshmen papers). Requires a baseline measurement for comparison (Leskes, 2002).

REFERENCES

- Cambridge, B. (1999). The scholarship of teaching and learning. *American Association for Higher Education (AAHE) Bulletin*, 52(4):7.
- Cournoyer, B.R., Powers, G., Johnson, J. & Bennett, R. (2000). Economic Modeling in Social Work Education. *Indiana University School of Social Work*. Retrieved September 13, 2011 from:
<http://journals.iupui.edu/index.php/advancesinsocialwork/article/viewArticle/21>.
- Creswell, J. W. (Ed.). (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd Ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Government Accountability Office. (2005). Glossary: performance measurement and evaluation: Definitions and relationships. Retrieved on October 10, 2011 from: <http://www.gao.gov/new.items/d05739sp.pdf>.
- Hansen, M. (2011). Direct and Indirect Measures of Student Learning. Minutes: Program Review and Assessment Committee (March 11, 20011). IUPUI.
- Hansen, M. J. and Borden, V. M. H. (2006), Using action research to support academic program improvement. *New Directions for Institutional Research*, 2006: 47–62. doi: 10.1002/ir.179
- Indiana University, Office of Research Administration. (n.d.). *Human Subjects Office*. Retrieved April 22, 2011, from <http://researchadmin.iu.edu/HumanSubjects/definitions.html>.
- Johnson, P. M. (2005). A glossary of political economy terms. Retrieved April 29, 2011 from, Auburn University, Department of Political Science Web site: <http://www.auburn.edu/~johnspm/gloss/efficiency>.
- Lehman College, Office of Assessment Glossary accessed : <http://www.lehman.cuny.edu/research/assessment/glossary.php>
- Leskes, A. (2002). Beyond confusion: An assessment glossary [Electronic version]. *Association of American Colleges and Universities (AACU): Peer Review, Winter / Spring, 2002*. Retrieved May 6, 2011 from:
http://assessment.uconn.edu/docs/resources/Andrea_Leskes_Assessment_Glossary.pdf.
- Levin, H. M. (2005). Cost Effectiveness. In *Encyclopedia of evaluation* (p. 90.). Thousand Oaks, CA: Sage Publications.
- Munoz, M. A. (2005). Correlation. In *Encyclopedia of evaluation* (p. 86.). Thousand Oaks, CA: Sage Publications.
- Munoz, M. A. (2005). Statistical Significance. In *Encyclopedia of evaluation* (p. 390.). Thousand Oaks, CA: Sage Publications.
- Owen, J. M. (2005). Kirkpatrick Four-Level Evaluation Model. In *Encyclopedia of evaluation* (p. 221-226). Thousand Oaks, CA: Sage Publications.
- Palomba, C. A. & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco, CA: Jossey-Bass.
- Petrosko, J. M. (2005). Measurement. In *Encyclopedia of evaluation* (p. 247.). Thousand Oaks, CA: Sage Publications.

- Petrosko, J. M. (2005). Basic and Applied Research. In *Encyclopedia of evaluation* (p. 18.). Thousand Oaks, CA: Sage Publications.
- Petrosko, J. M. (2005). Pre-Post Design. In *Encyclopedia of evaluation* (p. 247.). Thousand Oaks, CA: Sage Publications.
- Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage Publications.
- National Center for Research on Evaluation, Standards, and Student Testing (CRESST). (n.d.). *Assessment glossary*. Retrieved May 6, 2011 from: <http://www.cse.ucla.edu/products/glossary.php>.
- National Science Foundation. (2010). *The 2010 user-friendly handbook for project evaluation*. United States. Retrieved April 22, 2011 from: <http://caise.insci.org/uploads/docs/TheUserFriendlyGuide.pdf>
- Organization for Economic Co-operation and Development (OECD). (2011). *Glossary of statistical terms*. Retrieved May 10, 2011 from: <http://stats.oecd.org/glossary/>.
- Suskie, L. (2004). *Assessing student learning: A common sense guide*. Bolton, MA: Anker Publishing Company.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Scriven, M. (1991). Evaluation thesaurus (4th ed.). Newbury Park, CA: Sage. Thomas, C. L. (2005). Reliability. In *Encyclopedia of evaluation* (p. 90.). Thousand Oaks, CA: Sage Publications.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *American Journal of Evaluation*, 17 (2), 151-161.
- Thomas, C. L. (2005). Cost Effectiveness. In *Encyclopedia of evaluation* (p. 90.). Thousand Oaks, CA: Sage Publications.
- Walvoord, B.E. (2004). *Assessment: Clear and simple*. San Francisco: Jossey-Bass.
- W. K. Kellogg Foundation. (2004, January). *Using logic model to bring together planning, evaluation, and action: Logic model developmental guide*. Retrieved April 29, 2001 from http://opas.ous.edu/Committees/Resources/Publications/WKKF_LogicModel.pdf.