Assessment Terminology:  
A Glossary of Useful Terms

The following is a list of terms that are used to describe current educational assessment practices. This guide is intended not to promote the use of jargon, but to establish a clear and common understanding of what these terms mean.

**Achievement Test** A standardized test designed to efficiently measure the amount of knowledge and/or skill a person has acquired, usually as a result of classroom instruction. Such testing produces a statistical profile used as a measurement to evaluate student learning in comparison with a standard or norm.

**Action Research** School and classroom-based studies initiated and conducted by teachers and other school staff. Action research involves teachers, aides, principals, and other school staff as researchers who systematically reflect on their teaching or other work and collect data that will answer their questions.

**Alternative Assessment** Many educators prefer the description "assessment alternatives" to describe alternatives to traditional, standardized, norm- or criterion-referenced traditional paper and pencil testing. An alternative assessment might require students to answer an open-ended question, work out a solution to a problem, perform a demonstration of a skill, or in some way produce work rather than select an answer from choices on a sheet of paper.

**Analytic Scoring** A type of rubric scoring that separates the whole into categories of criteria that are examined one at a time. Student writing, for example, might be scored on the basis of grammar, organization, and clarity of ideas. Useful as a diagnostic tool. (See Rubric.)

**Assessment** The evaluation or estimation of the nature, quality, or ability of someone or something. In an educational context, the process of observing learning; describing, collecting, recording, scoring, and interpreting information about a student's or one's own learning. Traditionally, student assessments are used to determine placement, promotion, graduation, or retention. In the context of school reform, assessment is an essential tool for evaluating the effectiveness of changes in the teaching-learning process.

**Assessment Literacy** Knowledge about the basic principals of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment.

**Authentic Assessment** Evaluating by asking for the behavior the learning is intended to produce; ideally mirroring and measuring student performance in a "real-world" context. Tasks used in authentic assessment are meaningful and valuable, and are part of the learning process. The concept of model, practice, feedback in which students know what excellent performance is and are guided to practice an entire concept rather than bits and pieces in preparation for eventual understanding. A variety of techniques can be employed in authentic assessment. The goal of authentic assessment is to gather evidence that students can use knowledge effectively and be able to critique their own efforts.

**Benchmark** Student performance standards (the level(s) of student competence in a content area.) An actual measurement of group performance against an established standard at defined points along the path toward the standard. Subsequent measurements of group performance use the benchmarks to measure progress toward achievement. A standard or point of reference against which things may be compared or assessed, "his painting sets the benchmark of quality"

**Cohort** A group whose progress is followed by means of measurements at different points in time.
**Competency Test** A test intended to establish that a student has met established minimum standards of skills and knowledge and is thus eligible for promotion, graduation, certification, or other official acknowledgement of achievement.

**Criterion Referenced Tests** A test in which the results can be used to determine a student's progress toward mastery of a content area. Performance is compared to an expected level of mastery in a content area rather than to other students' scores. The scores have meaning in terms of what the student knows or can do, rather than how the test-taker compares to a reference or norm group. Criterion referenced tests can have norms, but comparison to a norm is not the purpose of the assessment.

**Curriculum Alignment** The degree to which a curriculum's scope and sequence matches a testing program's evaluation measures.

**Curriculum-embedded or Learning-embedded Assessment** Assessment that occurs simultaneously with learning such as projects, portfolios and "exhibitions." Occurs in the classroom setting, and, if properly designed, students should not be able to tell whether they are being taught or assessed. Tasks or tests are developed from the curriculum or instructional materials.

**Dimension** Aspects or categories in which performance in a domain or subject area will be judged. Separate descriptors or scoring methods may apply to each dimension of the student's performance assessment.

**Essay Test** A test that requires students to answer questions in writing. Responses can be brief or extensive. Tests for recall, ability to apply knowledge of a subject to questions about the subject, rather than ability to choose the least incorrect answer from a menu of options.

**Evaluation** Both qualitative and quantitative descriptions of progress towards and attainment of project goals. Using collected information (assessments) to make informed decisions about continued instruction, programs, activities.

**Formative Assessment** Assessment occurring during the process of a unit or a course.

**High Stakes Testing** Any testing program whose results have important consequences for students, teachers, schools, and/or districts. Such stakes may include promotion, certification, graduation, or denial/approval of services and opportunity.

**Holistic Method** In assessment, assigning a single score based on an overall assessment of performance rather than by scoring or analyzing dimensions individually. The product is considered to be more than the sum of its parts and so the quality of a final product or performance is evaluated rather than the process or dimension of performance.

**Item Analysis** Analyzing each item on a test to determine the proportions of students selecting each answer. Can be used to evaluate student strengths and weaknesses; may point to problems with the test's validity and to possible bias.

**Inter-rater reliability** Inter-rater reliability is the extent to which two or more individuals (coders or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system.

**Journals** Students' personal records and reactions to various aspects of learning and developing ideas. A reflective process often found to consolidate and enhance learning.

**Mean** One of several ways of representing a group with a single, typical score. It is figured by adding up all the individual scores in a group and dividing them by the number of people in the group. Can be affected by extremely low or high scores.

**Measurement** Quantitative description of student learning and qualitative description of student attitude.
**Median** The point on a scale that divides a group into two equal subgroups. The median is not affected by low or high scores as is the mean. (See Norm.)

**Metacognition** The knowledge of one's own thinking processes and strategies, and the ability to consciously reflect and act on the knowledge of cognition to modify those processes and strategies.

**Multidimensional Assessment** Assessment that gathers information about a broad spectrum of abilities and skills (as in Howard Gardner's theory of Multiple Intelligences).

**Multiple Choice Tests** A test in which students are presented with a question or an incomplete sentence or idea. The students are expected to choose the correct or best answer/completion from a menu of alternatives.

**Norm** A distribution of scores obtained from a norm group. The norm is the midpoint (or median) of scores or performance of the students in that group. Fifty percent will score above and fifty percent below the norm.

**Norm Group** A random group of students selected by a test developer to take a test to provide a range of scores and establish the percentiles of performance for use in establishing scoring standards.

**Norm Referenced Tests** A test in which a student or a group's performance is compared to that of a norm group. The student or group scores will not fall evenly on either side of the median established by the original test takers. The results are relative to the performance of an external group and are designed to be compared with the norm group providing a performance standard. Often used to measure and compare students, schools, districts, and states on the basis of norm-established scales of achievement.

**Objective Test** A test for which the scoring procedure is completely specified enabling agreement among different scorers. A correct-answer test.

**On-Demand Assessment** An assessment process that takes place as a scheduled event outside the normal routine. An attempt to summarize what students have learned that is not embedded in classroom activity.

**Outcome** An operationally defined educational goal, usually a culminating activity, product, or performance that can be measured.

**Percentile** A ranking scale ranging from a low of 1 to a high of 99 with 50 as the median score. A percentile rank indicates the percentage of a reference or norm group obtaining scores equal to or less than the test-taker's score. A percentile score does not refer to the percentage of questions answered correctly, it indicates the test-taker's standing relative to the norm group standard.

**Performance-Based Assessment** Performance-based assessment is a test of the ability to apply knowledge in a real-life setting. Assessment of the performance is done using a rubric, or analytic scoring guide to aid in objectivity.

**Performance Criteria** The standards by which student performance is evaluated. Performance criteria help assessors maintain objectivity and provide students with important information about expectations, giving them a target or goal to strive for.

**Portfolio** A systematic and organized collection of a student's work that exhibits to others the direct evidence of a student's efforts, achievements, and progress over a period of time. The collection should involve the student in selection of its contents, and should include information about the performance criteria, the rubric or criteria for judging merit, and evidence of student self-reflection or evaluation.

**Portfolio Assessment** Portfolios may be assessed in a variety of ways. Each piece may be individually scored, or the portfolio might be assessed merely for the presence of required pieces, or a holistic scoring process might be used and an evaluation made on the basis of an overall impression of the student's collected work.
**Process** A generalizable method of doing something, involving steps or operations which are usually ordered and/or interdependent. Process can be evaluated as part of an assessment, as in the example of evaluating a student's performance during pre-writing exercises leading up to the final production of an essay or paper.

**Product** The tangible and stable result of a performance or task. An assessment is made of student performance based on evaluation of the product of a demonstration of learning.

**Profile** A graphic compilation of the performance of an individual on a series of assessments.

**Project** A complex assignment involving more than one type of activity and production. Projects can take a variety of forms, some examples are a mural construction, a shared service project, or other collaborative or individual effort.

**Quartile** The breakdown of an aggregate of percentile rankings into four categories: the 0-25th percentile, 26-50th percentile, etc.

**Rating Scale** A scale based on descriptive words or phrases that indicate performance levels. Qualities of a performance are described (e.g., advanced, intermediate, novice) in order to designate a level of achievement. The scale may be used with rubrics or descriptions of each level of performance.

**Reliability** The measure of consistency for an assessment instrument. The instrument should yield similar results over time with similar populations in similar circumstances.

**Rubric** In general a rubric is a scoring guide used in subjective assessments. A rubric also can be an explicit description of performance characteristics corresponding to a point on a rating scale.

**Sampling** A way to obtain information about a large group by examining a smaller, randomly chosen selection (the sample) of group members. If the sampling is conducted correctly, the results will be representative of the group as a whole.

**Scale Scores** Scores based on a scale ranging from 001 to 999. Scale scores are useful in comparing performance in one subject area across classes, schools, districts, and other large populations, especially in monitoring change over time.

**Scoring Criteria** Rules for assigning a score or the dimensions of proficiency in performance used to describe a student's response to a task. May include rating scales, checklists, answer keys, and other scoring tools.

**Self-Assessment** The learner uses an assessment list or rubric and benchmarks to assess his or her own work.

**Standardized Test** An objective test that is given and scored in a uniform manner. Standardized tests are carefully constructed and items are selected after trials for appropriateness and difficulty.

**Standards** Agreed upon values used to measure the quality of student performance, instructional methods, curriculum, etc.

**Subjective Test** A test in which the impression or opinion of the assessor determines the score or evaluation of performance. A test in which the answers cannot be known or prescribed in advance.

**Summative Assessment** Evaluation at the conclusion of a unit or units of instruction or an activity or plan to determine or judge student skills and knowledge or effectiveness of a plan or activity.

**Validity** The test measures the desired performance and appropriate inferences can be drawn from the results. The assessment accurately reflects the learning it was designed to measure.