

Bibliography

References used in development of Extended Grade Band Standards and Achievement Descriptors:

- Alternate Assessment Collaborative. (2004). *Enhanced assessment grant products*. Denver, CO: Colorado Department of Education. Available at <http://www.cde.state.co.us/cdesped>
- Council of Chief State School Officers. (2003). *Glossary of assessment terms and acronyms used in assessing special education students: A report from the Assessing Special Education Students (ASES) State Collaborative on Assessment and Student Standards (SCASS)*. Washington, DC: Author.
- Education Commission of the States. (1998). *Designing and implementing standards-based accountability systems*. Denver, CO: Author.
- Hansche, L. (1998). *Handbook for the development of performance standards: Meeting the requirements of Title I*. Washington, DC: U.S. Department of Education and Council of Chief State School Officers.
- Jaeger, R. M., & Tucker, C. G. (1998). *Analyzing, disaggregating, reporting, and interpreting students' achievement test results: A guide to practice for Title I and beyond*. Washington, DC: Council of Chief State School Officers.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Lehr, C., & Thurlow, M. (2003). *Putting it all together: Including students with disabilities in assessment and accountability systems* (Policy Directions No. 16). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Linn, R. L., & Herman, J. L. (1997). *A policymaker's guide to standards-led assessment*. Denver, CO: CRESST and ECS.
- Quenemoen, R., Thompson, S., & Thurlow, M. (2003). *Measuring academic achievement of students with significant cognitive disabilities: Building understanding of alternate assessment scoring criteria* (Synthesis Report 50). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Rabinowitz, S., Roeber, E., Schroeder, C., & Sheinker, J. (2006) *Creating aligned standards and assessment systems under NCLB*. Washington, DC: Council of Chief State School Officers.
- Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Sheinker, J. (2004, April 26). *Achievement standards for alternate assessments: What is standard setting?* A teleconference presentation for the National Center for Educational Outcomes to 38 State Departments of Education, Minneapolis, MN.

- Sheinker, J. M., & Redfield, D. (2001). *Handbook for professional development on assessment literacy*. Washington, DC: Council of Chief State School Officers.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Wisconsin Department of Public Instruction. (1998). *Wisconsin's model academic standards for mathematics*. Madison, WI: Author. Available at <http://www.dpi.wi.gov/standards/mathintro.html>
- Wisconsin Department of Public Instruction. (1998). *Wisconsin's model academic standards for English language arts*. Madison, WI: Author. Available at <http://www.dpi.wi.gov/standards/elaintro.html>
- Wisconsin Department of Public Instruction (1998). *Wisconsin's model academic standards for science*. Madison, WI: Author. Available at <http://www.dpi.wi.gov/standards/sciintro.html>
- Wisconsin Department of Public Instruction. (2005). *Wisconsin student assessment system assessment framework for mathematics in grades 3 through 8 and 10*. Retrieved May 21, 2007, from http://www.dpi.wi.gov/oea/wkce-crt/math_framework.html
- Wisconsin Department of Public Instruction. (2005). *Wisconsin student assessment system assessment framework for reading in grades 3 through 8 and 10*. Retrieved May 21, 2007, from www.dpi.wi.gov/oea/wkce-crt/read_framework.html
- Wisconsin Department of Public Instruction. (2005). *Wisconsin student assessment system assessment framework for science in grades 4, 8 and 10*. Retrieved May 21, 2007, from www.dpi.wi.gov/oea/wkce-crt/read_framework.html
- Ysseldyke, J., Krentz, J., Elliott, J., Thurlow, M. L., Erickson, R., & Moore, M. L. (1998). *NCEO framework for educational accountability*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Glossary of Assessment and Related Education Terms

Accommodations are changes in the administration of an assessment, such as setting, scheduling, timing, presentation format, response mode, or others, including any combination of these that does not change the construct intended to be measured by the assessment or the meaning of the resulting scores. Accommodations are used for equity, not advantage, and serve to level the playing field. To be appropriate, assessment accommodations must be identified in the student's Individualized Education Plan (IEP) or Section 504 plan and used regularly during instruction and classroom assessment.

Accountability is the use of assessment results and other data to ensure that schools are moving in desired directions. Common elements include standards, indicators of progress toward meeting those standards, analysis of data, reporting procedures, and rewards or sanctions.

Accountability System is a plan that uses assessment results and other data and outlines the goals and expectations for students, teachers, schools, districts, and states to demonstrate the established components or requirements of accountability. An accountability system typically includes rewards for those who exceed the goals and sanctions for those who fail to meet the goals.

Achievement Descriptors are narrative descriptions of performance levels that convey student performance at each achievement level. They further define content standards by connecting them to information that describes how well students are learning the knowledge and skills contained in the content standards. (See also Performance Descriptors.)

Achievement Levels are measurements that distinguish an adequate performance from a novice or expert performance. Achievement levels provide a determination of the extent to which a student has met the content standards. (See also Performance Levels.)

Achievement Standards are a system that includes performance levels (e.g., unsatisfactory, proficient, advanced), descriptions of student performance for each level, examples of student work representing the entire range of performance for each level, and cut scores. A system of performance standards operationalizes and further defines content standards by connecting them to information that describes how well students are doing in learning the knowledge and skills contained in the content standards. (See also Performance Standards.)

Achievement Test is an instrument designed to efficiently measure the amount of academic knowledge and/or skill a student has acquired from instruction. Such tests provide information that can be compared to either a norm group or a measure of performance, such as a standard.

Adaptations is a generalized term that describes a change made in the presentation, setting, response, timing, or scheduling of an assessment; it may or may not change the construct of the assessment.

Adequate Yearly Progress (AYP) is a provision of the federal No Child Left Behind Act of 2001 (NCLB, 2001), which is legislation that requires schools, districts, and states to demonstrate, using students' test scores, that their students are making academic progress based on the percentage of students attaining proficiency on state standards.

Age Appropriate refers to the characteristics of the skills taught, the activities and materials selected, and the language level employed; each should reflect the chronological age of the student.

Alignment refers to the similarity or match between or among content standards, achievement (performance) standards, curriculum, instruction, and assessments in terms of breadth, depth, and complexity of knowledge and skill expectations.

Alternate Assessment is an instrument used in gathering information on the standards-based performance and progress of students whose disabilities preclude their valid and reliable participation in general assessments. Alternate assessments measure the performance of a relatively small population of students who are unable to participate in the general assessment system, even with accommodations, as determined by the IEP Team.

Assessment is the process of collecting information about individuals, groups, or systems. It relies on a number of instruments, one of which may be a test. Therefore, *assessment* is a more comprehensive term than *test*.

Assessment Literacy is the knowledge of the basic principles of sound assessment practice, including terminology, development, administration, analysis, and standards of quality.

Assistive Technology is a device, piece of equipment, product system, or service that is used to increase, maintain, or improve the functional capabilities of a student with a disability. (See 34 CFR §§300.5 and 300.6.)

Authentic Assessment is an evaluation of student learning that requires demonstration of the behavior the learning is intended to produce.

Baseline Data are the initial measures of performance against which future measures will be compared.

Benchmarks are specific statements of knowledge and skills within a content area continuum. These indicate what a student must possess to demonstrate a level of progress toward mastery of a standard.

Bias (test bias), in a statistical context, is a systematic error in a test score. In discussing test fairness, bias is created by not allowing certain groups into the sample, not designing the test to allow all groups to participate equitably, selecting discriminatory material, testing content that has not been taught, etc. Bias usually favors one group of test takers over another, resulting in discrimination.

Body of Evidence constitutes information or data that establish that a student can perform a particular skill or has mastered a specific content standard. The evidence must be either produced by the student or collected by someone who is knowledgeable about the student.

Breadth indicates the comprehensiveness of the content and skills embodied in the standards, curriculum, or assessments.

Criterion-Referenced Tests (CRT) is an assessment that measures a student's achievement of content standards and objectives rather than comparing a student's performance to the performances of other similar test takers.

Cues are assistance, words, or actions provided to a student to increase the likelihood that the student will give the desired response.

Curriculum is a document that describes what teachers do to convey grade-level knowledge and skills to a student.

Curriculum-based Assessments (instructionally supportive tests) are assessments that mirror instructional materials and procedures related to the curriculum and result in an ongoing process of monitoring progress and guiding adjustments in instruction, remediation, accommodations, or modifications provided to the student.

Cut Score is a specified point on a score scale. Scores at or above that point are interpreted differently from scores below that point.

Depth indicates the level of cognitive processing (e.g., recognition, recall, problem solving, analysis, synthesis, evaluation) required for success relative to the performance standards.

Disaggregation refers to the collection and reporting of student achievement results by particular subgroups (e.g., students with disabilities, limited-English proficient students) to ascertain the subgroup's academic progress. Disaggregation makes it possible to compare subgroups or cohorts.

Eligible Students with Disabilities are students who, in accordance with the definitions, eligibility criteria, procedures, and safeguards contained in Wisconsin Board of Education Policy, are designated as in need of specially designed instruction, or special education and related services.

Essence of the Standard is that which conveys the same ideas, skills, and content of the standard, expressed in simpler terms.

Errors of Measurement are the differences between observed scores and the theoretical true scores; the amount of uncertainty in reporting scores; the degree of inherent imprecision based on test content, administration, scoring, or examinee conditions within the measurement process that produce errors in the interpretation of student achievement.

Exemplar refers to scored student work that evidences or exhibits the ideal for a particular rubric score point.

Extended Grade Band Objectives are content standards that have been extended from or linked to the Content Standards and Grade-Level Objectives while maintaining the essence of those standards, thereby identifying the grade-level expectations for students with significant cognitive disabilities to access and make progress in the general curriculum.

Gap Analysis is an investigation of differences in achievement performance between two or more different groups of students, such as general education students and students with disabilities.

Grade Level is the grade in which a student is enrolled.

Grade Band consists of two or more consecutive grade levels that are tested together. Wisconsin has chosen four extended grade bands for reading and mathematics. These are Grades 3-4, Grades 5-6, Grades 7-8, and Grade 10. Science is tested only at Grade 4, Grade 8, and Grade 10.

High Stakes Testing refers to a test that has important consequences for students, teachers, schools, districts, and/or states attached to the results. Consequences may include promotion, graduation, rewards, or sanctions.

Individualized Education Program (IEP) is a written plan, developed by a team of regular and special educators, parents, related service personnel, and the student, as appropriate, describing the specially designed instruction needed for an eligible exceptional student to progress in the content standards and objectives and to meet other educational needs.

Large-Scale Assessments are tests that are administered simultaneously to large groups of students within the district or state.

Linked defines a relationship between Grade-Level Objectives for Content Standards and Extended Grade Band Standards that reflect similar content and skills but do not match in terms of breadth, depth, and complexity.

Modifications are changes in the administration of an assessment that include changes in the setting, scheduling, timing, presentation format, response mode, or any combination thereof that modify what is intended to be measured by the assessment or the meaning of the resulting scores.

Multiple Measures are measurements of student or school performance through more than one form or test.

- For students, these might include teacher observations, performance assessments or portfolios.
- For schools, these might include dropout rates, absenteeism, college attendance rates, or documented behavior problems.

Natural Cues are assistance given to a student to provide a flow among the expectations presented by the educator, opportunities to learn, and the desired outcome exhibited by the student.

Opportunity to Learn refers to the provision of learning conditions, including suitable adjustments, to maximize a student's chances of attaining the desired learning outcomes, such as the mastery of content standards.

Performance Assessment is a task or series of tasks requiring a student to provide a response or create a product to show mastery of a specific skill or content standard.

Performance Descriptors are narrative descriptions of performance levels that convey student performance at each achievement level. They further define content standards by connecting them to information that describes how well students are learning the knowledge and skills contained in the content standards. (See also Achievement Descriptors.)

Performance Levels are measurements that distinguish one performance from others (e.g., an adequate performance from a novice or expert performance). Performance levels provide a determination of the extent to which a student has met the content standards. (See also Achievement Levels.)

Performance Standards are a system that includes performance levels (e.g., unsatisfactory, proficient, advanced), descriptions of student performance for each level, examples of student work representing the entire range of performance for each level, and cut scores. A system of performance standards operationalizes and further defines content standards by connecting them to information that describes how well students are doing in learning the knowledge and skills contained in the content standards. (See also Achievement Standards.)

Portfolio is a deliberate collection of student-generated or student-focused evidence that provides the basis for demonstrating the student’s mastery of a range of skills, performance level, or improvement in skills over time. The portfolio evidence may include student work samples, photographs, videotapes, interviews, anecdotal records, interviews, and observations.

Portfolio Assessment is an organized collection or documentation of student-generated or student-focused work that typically depicts the range of individual student skills.

Readability refers to the formatting of presented material that considers several factors—the organization of text, syntactic complexity of sentences, use of abstractions, density of concepts, sequence and organization of ideas, page format, sentence length, paragraph length, variety of punctuation, student background knowledge or interest, and use of illustrations or graphics—in determining the appropriate level of difficulty of instructional or assessment materials.

Real-world Application refers to the opportunity for a student to exhibit a behavior or complete a task that he or she would normally be expected to perform outside of the school environment.

Response Requirements are the types, kinds, or methods of action required of a student to answer a question or testing item. The response may include, but not be limited to, reading, writing, speaking, creating, and drawing.

Rubric is a scoring tool based on a set of criteria used to evaluate a student’s test performance. The criteria contain a description of the requirements for varying degrees of success in responding to the question or performing the task. Rubrics may be diagnostic or analytic (providing ratings of multiple criteria), or they may be holistic (describing a single, global trait).

Stakeholders are groups of individuals perceived to be vested in a particular decision (e.g., a policy decision).

Standardized is defined as an established procedure for assuring that a test is administered with the same directions and under the same conditions (time limits, etc.) and scored in the same manner for all students to ensure the comparability of scores. Standardization supports reliable and valid comparisons among students taking the test. The two major types of standardized tests are norm-referenced and criterion-referenced.

Standards come in two types—content and achievement (performance).

- **Content standards** are statements of the subject-specific knowledge and skills that schools are expected to teach students and indicate what students should know and be able to do.
- **Achievement (Performance) standards** are indices of qualities that specify how adept or competent a student demonstration must be. They consist of the following four components:
 1. levels that provide descriptive labels or narratives for student performance (e.g., advanced, proficient);
 2. descriptions of what students at each particular level must demonstrate relative to the task;
 3. examples of student work at each level that illustrate the range of performance within each level; and
 4. cut scores that clearly separate each performance level from others.

Standards-based Assessments are assessments constructed to measure how well students have mastered specific content standards or skills.

Starting Point is the accepted baseline that is referenced to determine future levels of performance or progress.

Student Self-reflection is assessment by a student of his or her own instruction or work.

Subgroup refers to a well-defined group of students. For example, the NCLB Act identifies the following specific subgroups that must achieve adequate yearly progress: race/ethnicity groups, students with disabilities, limited-English proficient (LEP) students, and economically disadvantaged students.

Test is a measuring device or procedure. Educational tests are typically composed of questions or tasks designed to elicit predetermined behavioral responses or to measure specific academic content standards.

Test Forms are parallel or alternate versions of a test that are considered interchangeable; that is, they measure the same constructs, are intended for the same purposes, and are administered using the same directions.

Test Presentation is the method, manner, or structure by which test items or assessments are administered to the student.

Test Security refers to procedures established to ensure current or future confidentiality, fidelity, and integrity of a test. Public access is limited and strictly monitored, with clearly outlined consequences for breaches in test security.

Universal Design of Assessment is a method for developing an assessment to ensure accessibility by all students, regardless of ability or disability. Universal design is based on principles that originated in the field of architecture to consider user diversity during the conceptual stage of development.

Validity is the extent to which a test measures what it was designed to measure. Multiple types of validity exist. Common types of validity include the following:

- **Construct Validity** refers to the extent to which the characteristic to be measured relates to test scores that measure the behavior in situations where the construct is thought to be an important variable.
- **Content Validity** refers to the extent to which the stimulus materials or situations that compose the test call for a range of responses that represent the entire domain of skills, understandings, or behaviors that the test is intended to measure.
- **Convergent Validity** refers to the extent to which the assessment results positively correlate with the results of other measures designed to assess the same or similar constructs.
- **Criterion-related Validity** refers to the extent to which test scores of a group or subgroup are compared to other criterion measures (ratings, classifications, other tests) assigned to the examinees.
- **Face Validity** is a concept based on a judgment concerning how relevant the test items appear to be. It relates more to what a test appears to measure than to what the test actually measures.
- **Consequential Validity** is the extent to which the assessment results in the intended positive outcomes for students, e.g., results in improved instruction and improved student achievement.

These definitions are adapted from the *Glossary of Assessment Terms and Acronyms Used in Assessing Special Education Students: A Report from the Assessing Special Education Students (ASES) State Collaborative on Assessment and Student Standards (SCASS)* (Council of Chief State School Officers, 2003).

