Glossary of Assessment Terms
by
Jay McTighe and Judy Arter

Analytical Trait Scoring – a scoring procedure in which performances are evaluated for selected traits, with each trait receiving a separate score. For example, a piece of writing may be evaluated according to organization, use of details, attention to audience, and language usage/mechanics. Trait scores may be weighted and/or totaled. (see *Holistic* and *Primary Trait Scoring)*

Anchor(s) – the representative products or performances used to illustrate each point on a scoring scale. The top anchor is sometimes called the exemplar. (see *Criteria, Rubric, and Scoring Guide)*

Assessment – any systematic basis for making inferences about characteristics of people, usually based on various sources of evidence; the global process of synthesizing information about individuals in order to understand and describe them better.

Authentic – refers to assessment tasks that elicit demonstrations of knowledge and skills in ways that they are applied in the "real world." An "authentic assessment" task is also engaging to students and reflects the best current thinking in instructional activities. Thus, teaching to the task is desirable. *(see Performance Assessment* and *Task*)

Bias and Distortion – factors, unrelated to the skill being assessed, that interfere with a valid inference regarding a student's true ability. For example, too much reading on a mathematics test might result in a distorted vision of a student's mastery of mathematics content. *(see Reliability* and *Validity)*

Constructed Response Assessment– assessment questions that require a student to produce a response rather than selected it from a list. For example, essays, reports, oral presentations, reading fluency, open-ended mathematics problems, etc.

Content Standards – goal statements identifying the knowledge, skills, and dispositions to be developed through instruction. *(see Target)*

Criteria – guidelines, rules, or principles by which student responses, products, or performances are judged. *(see Criterion-Referenced* and *Evaluation)*

Criterion-Referenced – an approach for describing a student's performance according to established criteria; e.g., she typed 55 words per minute without errors. *(see Criteria, Norm-Referenced)*

Dispositions – refers to the affective dimensions of students in school; e.g., motivation to learn, attitude toward school, academic self-concept, flexibility, persistence, and locus of control. Some

scoring guides are designed to assess dispositions. These provide specific, observable indicators of the disposition being assessed.

Evaluation – judgment regarding the quality, value, or worth of assessment results; e.g., "the information we collected indicates that students are reading as well as we would like." Evaluations are usually based on multiple sources of information. *(see Criteria)*

Generic (General) Criteria – criteria that can be used to score performance on a large number of related tasks; e.g., a mathematics rubric that can be used to score problem solving regardless of the specific content of the problem. *(see Task-Specific Criteria)*

Holistic Scoring – a scoring procedure yielding a single score based upon an overall impression of a product or performance. *(see Analytic Trait and Primary Trait Scoring)*

Performance List – a scoring guide consisting of designated criteria, but without descriptive details. For example, a performance list for writing might contain six features—ideas, organization, voice, word choice, sentence fluency, and conventions. Unlike a rubric, a performance list merely provides a set of features without defining the terms or providing indicators of quality. *(see Anchor, Criteria, Rubric, and Scoring Guide)*

Norm-Referenced – describing a student's performance by comparison to other, similar students; e.g., she typed better than 80 percent of her classmates. (see *Criterion-Referenced*)

Performance Assessment – an assessment activity that requires students to construct a response, create a product, or perform a demonstration. Since performance assessments generally do not yield a single correct answer or solution method, evaluations of student products or performances are based on judgments guided by criteria. *(see Assessment, Criteria, and Task)*

Performance Standard – an established level of achievement, quality of performance, or degree of proficiency. Performance standards specify how well students are expected to achieve or perform. *(see Anchor and Content Standard)*

Primary Trait(s) Scoring – a scoring procedure by which products or performances are evaluated by limiting attention to a single criterion or a few selected criteria. These criteria are based upon the trait or traits determined to be essential for a successful performance on a given task. For example, a note to a principal urging a change in a school rule might have persuasiveness as the primary trait. Scorers would attend only to that trait. *(see Analytical Trait and Holistic Scoring)*

Reliability – the degree to which the results of an assessment are dependable and yield consistent results across raters (inter-rater reliability), over time (test-retest reliability), or across different versions of the same test (internal consistency or inter-form reliability). Technically, this is a statistical term that defines the extent to which errors of measurement are absent from an assessment instrument. *(see Bias and Distortion and Validity)*

Rubric – a set of general criteria used to evaluate a student's performance in a given outcome area. Rubrics consist of a fixed measurement scale (e.g., 4-point) and a list of criteria that describe the characteristics of products or performances for each score point. Rubrics are frequently accompanied by examples (anchors) of products or performances to illustrate the various score points on the scale. *(see Anchor, Criteria, Performance List, Scoring Guide)*

Selected Response Assessments – assessment questions that ask students to select an answer from a provided list. For example, multiple-choice, matching, and true-false.

Scoring Guide – a generic term for a criterion-based tool used in judging performance. In this book, we are using scoring guide synonymously with criteria and rubric. *(see Criteria, Performance List, and Rubric)*

Standardized – a set of consistent procedures for constructing, administering, and scoring an assessment. The goal of standardization is to ensure that all students are assessed under uniform conditions so that interpretation of their performance is comparable and not influenced by differing conditions. Both norm-referenced and criterion-referenced assessments can be standardized.

Task – an assessment exercise involving students in producing a response, product or performance; e.g., solving a mathematics problem, conducting a laboratory in science, or writing a paper. Since tasks are associated with performance assessments, many are complex and open-ended, requiring responses to a challenging question or problem. However, there can be simple performance tasks, such as reading aloud to measure reading rate. Tasks don't have to be exclusively used as stand-alone activities that occur at the end of instruction; teachers can observe students working on tasks during the course of regular instruction in order to provide on-going feedback. *(see Assessment and Test)*

Target (as in "learning target")—statements of what we want students to know and be able to do. *(see Content Standards)*

Task-Specific Criteria – a scoring guide or rubric that can only be used with a single exercise or task. Since the language is specific to a particular task (e.g., to get a "4" the response must have "accurate ranking of children on each event, citing Zabi as overall winner"), task-specific guides cannot be applied to any other task without modification. *(see Generic Criteria)*

Test – a set of questions or situations designed to permit an inference about what an examinee knows or can do in an area of interest. *(see Assessment, Performance Assessment, and Task)*

Validity – an indication of how well an assessment measures what it was intended to measure; e.g., does a test of laboratory skills really assess laboratory skills or does it assess ability to read and follow instructions? Technically, validity indicates the degree of accuracy of predictions or inferences based upon an assessment measure. *(see Bias and Distortion and Reliability)*