**ACADEMIC ACHIEVEMENT**

Academic achievement can be defined as learned proficiency in basic skills and content knowledge. Documentation of achievement is an issue that has seen increasing concern among educators, policy makers, and the general public in recent decades and has been the driving force behind legislative and policy development in education since the turn of the20th century. Three aspects of academic achievement are addressed in this entry:

1. The format for collecting the observations, typically using academic achievement tests and

measures

2. The types of referents or comparisons for making interpretations

3. The purposes of academic achievement testing and the process for making decisions

ITEM FORMAT

In general, tests use either of two format types, depending on how the student is expected to respond. Selected response items present a (question) stem and a set of possible answers from which a response is chosen. In contrast, constructed response items require the student to provide an answer to a given question or prompt without any suggested responses.

*Measurement and Assessment in Teaching*,

teachers and researchers believe that both types of items can adequately measure the same content; however, information about student knowledge and/or ability resulting from student responses may be different. For example, selected responses may be more appropriate for understanding the factual knowledge from a student on a unit chapter, whereas a constructed

response format is better suited for assessing a student’s knowledge of and ability to develop an argument in a debate class. To appropriately influence and direct educational decisions, the intended references and uses of the results must be considered prior to test construction and administration. Using selected and constructed response items, tests can be designed to measure both declarative knowledge and conditional and procedural skills. The

format depends on the nature of the decisions that are to be based on the outcomes.Regardless of format choice, the test items must target the ability of the student (group) and provide enough information to discriminate among different performances to make a valid decision.

**Selected Responses**

Selected response items provide the test taker with all of the information needed to choose an answer. The item contains a question stem or phrase and a set of possible choices, one of which is correct. Responses include multiple-choice, true/false, and matching questions, and the answers are scored objectively as correct or incorrect using an answer key.

**Selection response tests** are not limited to measuring factual knowledge, according to most measurement experts. The students’ understanding of instructional content and their ability to perform complex analyses—such as judging possible outcomes and interpreting results—can also be measured using selected response items. However, the dependency on reading skills and fixed answer choices may restrict students’ responses in a way that limits information for decision making. Apparent correct answers may be the result of guessing by the test taker and can have adverse effects on interpreting a student’s actual knowledge and ability. Because of this, positive results obtained from selected response items must be analyzed with caution. With some tests, this type of error is corrected by subtracting an additional fraction of an incorrect response each incorrect response.

**Constructed Responses**

Test items that use constructed responses ask students to create an answer to a prompting question or issue without any suggested answers provided. Responses may be written, oral, or performancebased. Typical items that require written responses include fill-in-the-blank, short answer, and extended essay. Oral presentations may include speeches and other performances that require demonstration of a physical or motor skill within a specific set of criteria, such as reading fluency within a one-minute time limit. Although most educators believe that responses to these items typically are scored subjectively using a rubric for determining the level of correctness, many performances can actually be scored objectively. Complex understanding and critical thinking skills can be demonstrated through a variety of production responses. The design of constructed response items enables the student to demonstrate proficiency with skill and content while minimizing the threat of guessing. The format of these items, however, limits the number of opportunities to assess student understanding across broad content areas. Furthermore, reliance on skills of self-expression may interfere with students’ ability to demonstrate mastery (e.g., present problems with access skills). Data regarding student knowledge and ability may be further obscured by subjective scoring methods that reduce the repeatability and generalizability of the results.

**TEST REFERENTS**

**Norm-Referenced**

Norm-referenced tests compare students’ scores relative to students’ scores in another, similar group (a normative sample) on the same items and tasks. Norm-referencing is relative in that the interpretations are quantified with a specific metric that positions the student within a group. A number of different metrics typically are offered, including: (a) ranks in the form of percentile and stanines, all of which provide ordinal outcomes; and (b) scale scores such as z- and t-scores as well as normal curve equivalent scores, all of which provide interval outcomes. For example, a student scoring in the 90th percentile on a given standardized test means that that student scored better than 90% of the students in the normative sample. In contrast, a z-score of –1.0 means that the student performed 1 standard deviation below the average of

others in the norm group.

**Criterion-Referenced**

Educational decisions based on interpretations of criterion-referenced tests focus on a specific set of items or problems that have been carefully sampled from a well-defined domain to judge student mastery relative to a set of absolute criteria. When states test students to determine if they have met specific standards, they are using criterion-referenced tests. Criterion-referenced interpretations do not depend on how other students performed; rather it is the amount of content or skill mastery that the student possesses that is being measured. For example, teachers typically make criterion-referenced decisions based on chapter tests to rank student performance levels. If a student scores 85% on a chapter test, the teacher can infer that the student possesses 85% mastery of the content or skill tested, from which the teacher can then make qualitative decisions (e.g., grades). Many norm-referenced tests purport to be criterion-referenced by the nature of the alignment with specific objectives.

**Individual-Referenced**

A third approach to instructional decision making leads to individual-referenced interpretations. In this case, instructional decisions are made based on data obtained from repeated measures to determine an individual student’s improvement over time.