

ASSESSMENT AND
MEASUREMENT
ISSUES WITH
GRADES

TxAHE 2022

Art Hernandez

University of the Incarnate Word

Background

The practice of grading is ubiquitous in higher education and grades themselves have been assigned multiple roles and meanings. However, the practice of grading and the interpretation of grades is often not informed by measurement or evaluation theory and the nature of grades themselves is not well understood by all interested and involved which can and often does result in misinterpretation and un-substantiable uses in data analysis for accreditation or other course or program level examinations and demonstrations (of effect for example). In addition, grade disputes between students and instructors are a common area of concern for institutions.

Focus and Purpose

This roundtable will provide a forum for practitioners and interested others to discuss the philosophy, conceptual frameworks and operationalization of grading practices focusing on considerations of meaning, utility, applicability, and interpretation as applicable to multiple audiences of and users of grade information. Technical issues which should inform interpretation of grades and grading practices will be the principal focus.

- Participants will be able to differentiate between measurement, assessment and evaluation.
- Participants will be able to apply considerations of measurement theory and practice as related to grades and grading.
- Participants will be able to describe some important considerations related to the interpretation of grades.

Goals

This session will provide an opportunity for discussion concerning how best to define and communicate practice and meaning in order to improve that practice, inform students (and interested others) and reduce the likelihood of such conflict. Concepts to be considered include fundamental psychometric considerations including demonstrations of validity and reliability as well as concepts associated with applicable evaluation theory. This roundtable will provide a forum for practitioners and interested others to discuss the philosophy, conceptual frameworks and operationalization of grading practices *focusing on considerations of meaning, utility, applicability, and interpretation* as applicable to multiple audiences of and users of grade information. Technical issues which should inform interpretation of grades and grading practices will be the principal focus.

ASSUMPTIONS ABOUT GRADES

The assignment of grades is presumed to relate to *student performance* on measures of designated student learning outcomes (SLO) for the course or program (i.e., some criteria), independent of observer (grader), ideally about which students and other stakeholders are pre-informed.

ASSUMPTIONS REGARDING ASSESSING STUDENT PERFORMANCE FOR GRADING

Student performance assessment is directly related to the criteria for grading and done in a way that results in a real/meaningful (unbiased) distinctions or differentiation between students relative to the assigned grade.

Reflection Questions

Do two students in the same class with the same grades know as much about the assessed content?

Does the same grade for the same course with the same instructor over two different semesters mean the same student knowledge/skill/performance?

GRADES AND GRADING

- Grades can be determined in multiple ways for multiple purposes and thus have multiple meanings. Without regard to the requirements for assignment (criteria) or assessment, they probably mean something different depending on instruction, assessments, program requirements, course requirements, student(s) characteristics, etc. Grades are often reified, that is interpreted independent of any consideration of influences.

ASSESSMENT AND GRADING

As with all indicators, grades should be valid, reliable and authentic. This means that difference in grade should reflect meaningful differences between student relative to the specified criteria. Unfortunately, grades are regularly reported independent of any statement of criteria or mechanism of assessment and thus convey very little context for meaningful interpretation.

ASSESSMENT AND GRADING

At a minimum, grades should be based on quality assessment instruments and methods which maximize the validity, reliability and authenticity of assessment upon which evaluative judgments (i.e., grades) are made. This requires the regular application of “tests” recognizing that *the characteristics of assessment tools (i.e., tests) are only measurable and understandable in the context of test takers.*

BASIC ASSUMPTION OF TRUE SCORE THEORY

$$O = T + e$$

Observed score equals true score *plus error*. It is presumed that any test score and thus grade will include some error. Grades are even more likely for error since they require standard setting that is usually arbitrary. The goal is to eliminate systematic error and control random error.

BASIC REQUIREMENTS FOR GOOD ASSESSMENT

- Reliability – pertains to consistency of measurement (i.e., limited error or “noise”)
- Validity – pertains to the accuracy of measurement (i.e., maximum relation to criteria)
- Authenticity – pertains to the legitimacy of measurement (i.e., describes equally well)

RELIABILITY

Reliability should be assessed in a way *congruent with purpose and population*. Two main foci: absolute (compared to a standard) and differential (compared to other tests with the same sample). The usual approach concerns absolute reliability but quite frequently grading depends on differences between students, assessments, courses, semesters, etc. However, *differential reliability is seldom examined*.

SOME APPROACHES FOR EXAMINING ASSESSMENT QUALITY

Examination for bias including differential reliability for groups, differences in Standard Error of Measurement, group regression differences, long term (program SLO's) differences between groups related to course grades, etc. It is important to note test-retest and internal consistency reliability are not interchangeable. (e.g., Schmidt FL, Le H, Ilies R. Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*. 2003;8:206-224).

VALIDITY

Validity pertains to the degree assessment relates to the purpose for that assessment – that the results of measurement pertain to what was intended rather than to something else. Validity is dependent on reliability and measures of reliability are indicative of the ceiling for validity. A test cannot be more valid than it is reliable.

AUTHENTICITY

Authenticity is closely related to validity but explicitly considers the diversity of the population being assessed. Authenticity pertains to the extent to which assessment is equally valid for all test takers and exams the extent of any bias – that is the consistent influence of some characteristic or behavior unrelated to the criteria.

SOME THINGS THAT AFFECT TEST QUALITY

- Item Irrelevance - check to be sure that every item is *directly* related to SLO or test purpose
- Item Heterogeneity – test contents should focus on specific instructional and curricular goals and ideally, item format is consistent throughout
- State Variation – tests should be administered as closely as possible under the same conditions for every test taker
- Respondents' Error – responses influenced by unrelated causes, naturally changing (e.g., “I feel” or “I think” etc.) random error, or deliberate underperformance are some examples
- Item Ambiguity - each test taker should similarly interpret the meaning of every item
- Sample Variance - diversity of test takers beyond differences in the object of measurement can affect the estimated reliability

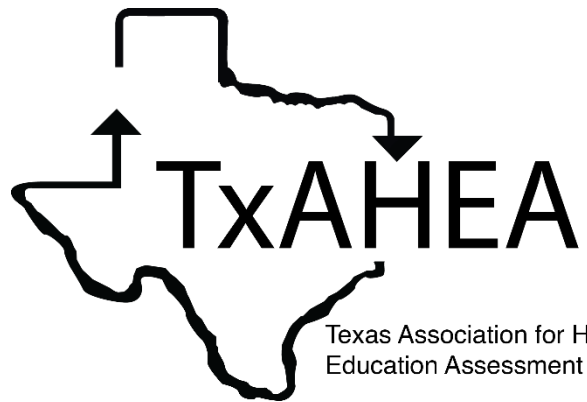
Reflection Questions

What is the difference and the relationship between authenticity, validity and reliability related to assessment?

How can you apply these concepts to grades and grading?

How often do you (or your program faculty) evaluate the quality of assessment and the meaning of grades?

Questions and Discussion



THANKS!

Presenter Email: aherna8@uiwtx.edu

*SESSION
EVALUATION*

SCAN

OR

ENTER THIS URL
INTO YOUR
BROWSER:
[TINYURL.COM/TX
AHEA](https://tinyurl.com/TXAHEA)

