Consistency/Reliability

Inclusive Assessment Seminar

Charlie DePascale

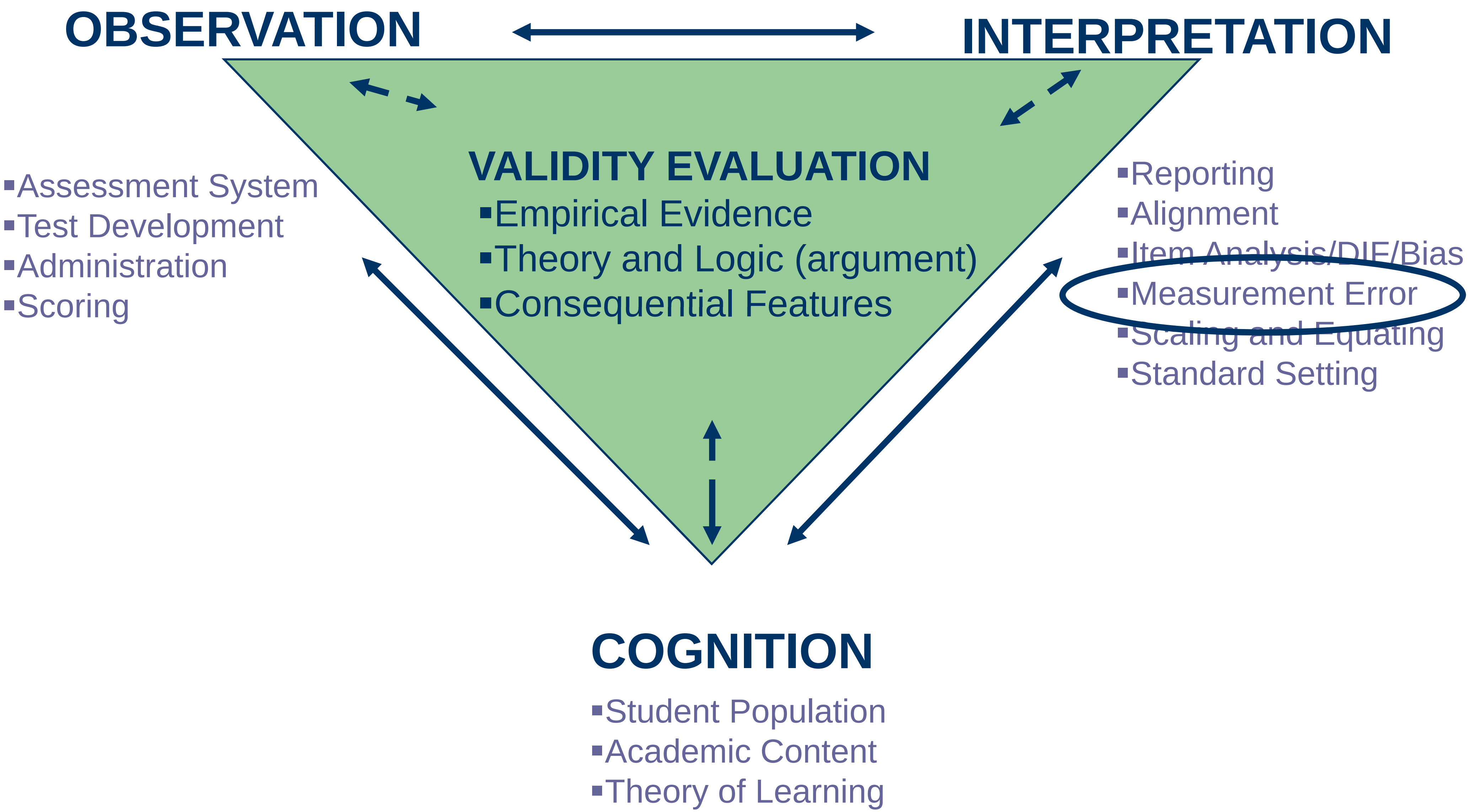
National Alternate Assessment Center

NHEAI

New Hampshire Enhanced Assessment Initiative:
Technical Documentation for Alternate Assessments

The Assessment Triangle and Validity Evaluation

(Marion, Quenemoen, & Kearns, 2006)



Measurement Error

- All measurements contain error
- Many sources contribute to measurement error
  - Features/traits/characteristics being measured
  - Individuals or object whose features are being measured
  - Measurement tool
  - People using the measurement tool

Common Sources of Error

- General education assessments:
  - Test content (items, length, format)
  - Administration conditions
  - Scoring
  - Student (preparation, motivation, anxiety)

- Added sources for Alternate Assessments:
  - Individual test administrators
  - Confounding medical conditions
  - Depth and breadth of the construct
  - Complexity of the assessment instruments

Interpreting Measurement Error: Two Key Conceptions

Classical Test Theory

- Measurement error is viewed as a *unitary, global entity*, though it is acknowledged to arise from a combination of sources.

  (Feldt and Brennan, 1989)

- Concern with estimating the total amount of error present.

Generalizability Theory

- Concerned with the errors from multiple sources as *separate entities*.

  (Feldt and Brennan, 1989)

- Estimates the magnitude of various sources of error

- Pinpoints sources of error so cost-efficient measurements can be built

  (Brennan, 1995)

A little background on reliability

- Reliability has traditionally been conceptualized as the *tendency toward consistency from one set of measurements to another.*

- “The fact that repeated sets of measurements never exactly duplicate one another is what is meant by unreliability” (Stanley, 1971).

- There are many approaches to interpreting measurement error and quantifying reliability.

How much should we worry?

- Acceptable levels of reliability are contingent upon the uses of test scores.
  - If scores are used only for school accountability purposes, reliability at the student level is not as critical.
  - If scores are used to make decisions about students, then student-level reliability is much more of a concern.
- Test users need to be explicit about the intended purposes of the test and acceptable uses of the results.

Alternate Assessment Reporting Requirements

- Individual student-level results must be reported to students and parents. (IDEA and NCLB)
- There is also an expectation that the results can be used to improve individual instruction.
- Therefore, although student test results might not be used for accountability, we must be concerned with their reliability and report the amount of measurement error associated with each score.

Reliability of Alternate Assessments

- We need to find the most appropriate method to define reliability.
- We need to find the most accurate way to calculate and report measurement error.
- Each type of alternate assessment presents a unique set of challenges.

Reliability Challenges with AA-AAS

- Attaining and calculating reliability with small numbers of tasks – particularly when tasks are individualized for students.
- Determining the impact of the test administrator.
- Evaluating the error associated with the student based on a single test administration – some assessment designs (e.g., portfolios, observation, checklists) build in multiple test occasions.

Characterizing Measurement Error

- Ideally, we want to calculate and report the amount of error associated with each of the major sources. In practice, however, this is hard to accomplish.
- Reporting results from a single source of error (e.g., inter-rater agreement) is not nearly sufficient.
- This afternoon’s session will present some strategies for approaching these challenges