# Introduction to Comparability

## Inclusive Assessment Seminar

Scott Marion

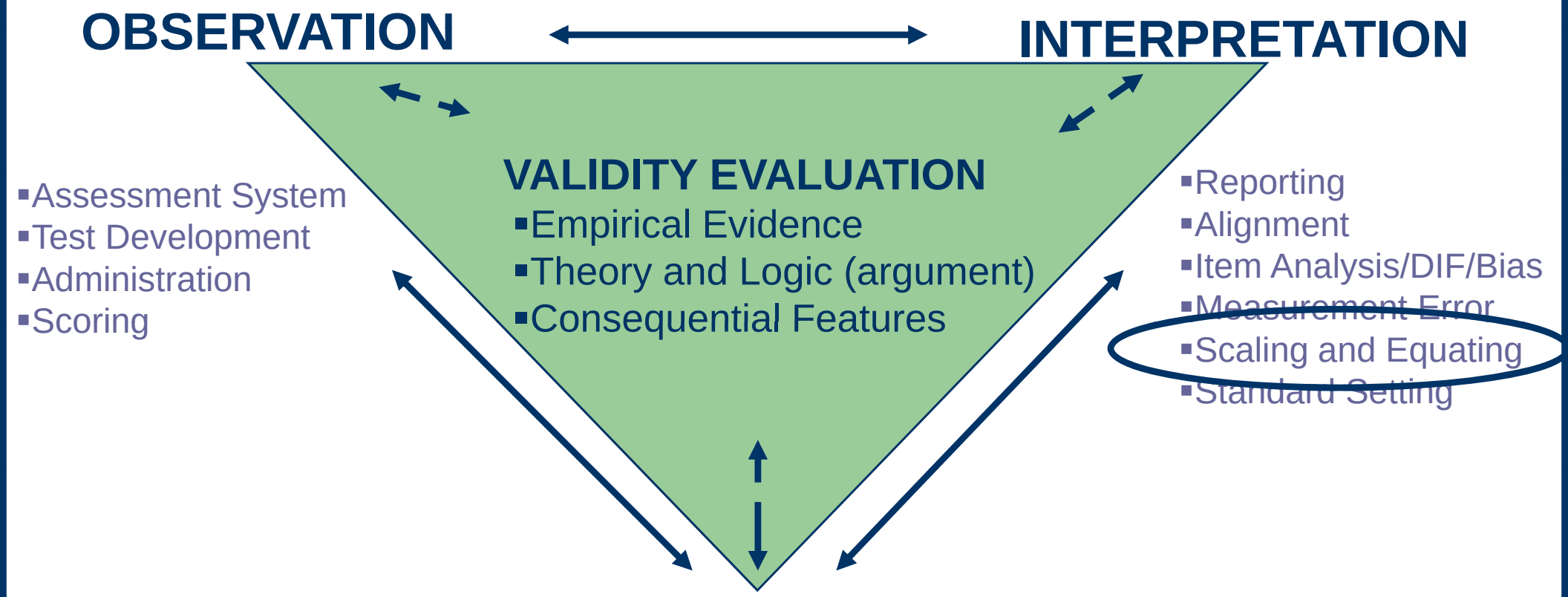
National Alternate Assessment Center

NHEAI

New Hampshire Enhanced Assessment Initiative:
Technical Documentation for Alternate Assessments

# The Assessment Triangle and Validity Evaluation

**(Marion, Quenemoen, & Kearns, 2006)**




## COGNITION

- Student Population
- Academic Content
- Theory of Learning

# What is comparability?

- In an assessment context, comparability means that the inferences from the scores on one test can be psychometrically related to a score on another “comparable” test.

# Why do we care about comparability?

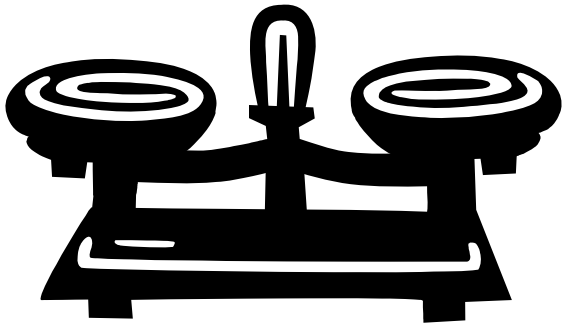
- In fully individualized assessments, we don't, BUT we need scores to be comparable when…
  - Judging two or more scores against a common standard,
  - Aggregating scores to the school or district level, (we are assuming that scores are comparable)
  - Judging scores for the same students and/or the same schools across years.

# Comparability and Flexibility

- Flexibility or individualization can pose challenges to comparability.
- Using the same items and the same (extended) content standards each year would appear to ameliorate any comparability concerns.
  - But, not everything is as it appears…issues with “teaching to the test” threaten comparability.
- Obviously, completely individualized tasks addressing non-systematic selection of standards raises considerable comparability concerns.

# Traditional Methods

- Scaling is the process by which we translate raw scores from an assessment into a more interpretable numerical scale such that similar scores have similar meanings.
- Linking describes a family of approaches (including equating) by which we can place the scores from one assessment on the same SCALE as another assessment (e.g., putting the 2006 scores on the 2005 scale).



# Scaling Requirements

- We can create scales from many different types of raw scores, but for the scale to lead to valid inferences, the original raw scores must have a similar conceptual foundation (i.e., the raw scores should be derived from similar assessment experiences, unless we move to a normative approach).

# Linking (Equating) Requirements

- There is a fairly extensive literature regarding the requirements for valid equating. Depending on content and contextual relationships between the two tests, the linking could be as strong as formal equating.
- If equating assumptions are not met, calibration, projection, or even judgmental methods could be applied to connect the two sets of test scores.

# But, we give the same test every year…

- Administering exactly the same test items each year certainly makes score comparability easier, but what do you do when…
  - you have to replace “tired” or poorly functioning items?
  - you find out that there is score inflation due to “teaching to the test”?

# But, we have considerable flexibility…

- In these cases, traditional, statistically-based methods for ensuring comparability might not work.
- We will have to rely on at least one of several judgmental methods.

# This Afternoon

- We will explore approaches for both:
  - Ensuring—to the extent possible—comparability among alternate assessment scores for a variety of alternate assessment approaches; and
  - Documenting the technical quality of the comparability approaches employed.