Equating UDS Neuropsychological Tests:

3.0>2.0, 3.0=2.0, 3.0<2.0?

Dan Mungas, Ph.D.
University of California, Davis

There are three kinds of statisticians. Those who can count and those who can't.

Gerald Van Belle, (with apologies to his original source)

Short Version - Conclusions

- Measure selection
- Study design
- Statistical and psychometric methods

Overview

- Test Equating: What does this mean?
- Content issues
- Study design issues
- Statistical / psychometric issues
- Implications for UDS transition

How/when are tests equivalent?

- Have same content
- Have same statistical properties
  - Mean
  - Variance
- Have same psychometric properties
  - Reliability
  - Validity

Equivalent Content in Equated Tests

- Equivalent tests should measure the same domain
- Trivial example
  - A test of object naming is never equivalent to a list learning test
- Less trivial example
  - A test of global cognition with episodic memory (or object naming) content is not the same as a test of global cognition without episodic memory (or object naming) content

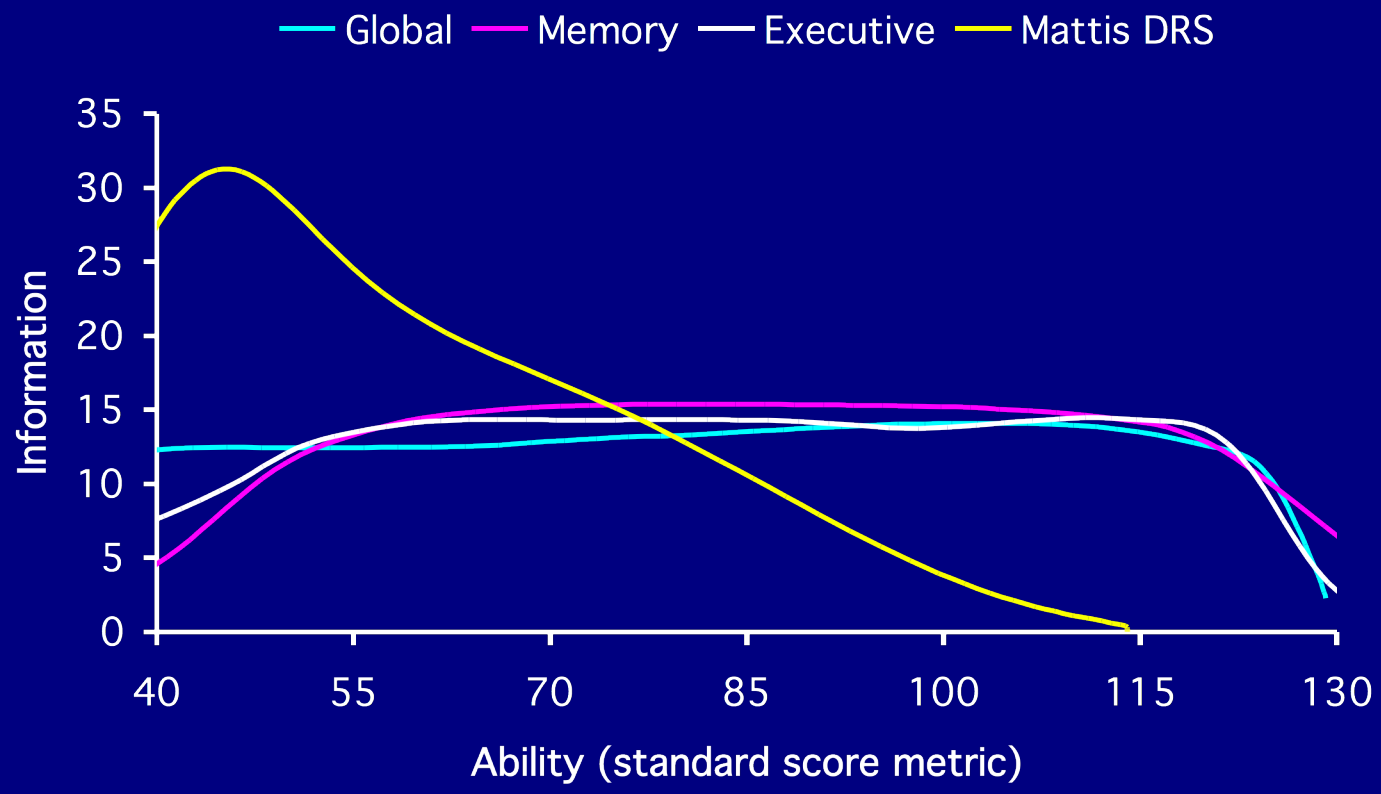
Equivalent Statistical Properties in Equated Tests

- Equivalent tests should have the same distributions
  - in equivalent samples
- Minimally, means and variances should be the same
- Distribution shape must also be the same

Equivalent Psychometric Properties in Equated Tests

- Minimally, classical test theory reliability should be the same
  - Internal consistency reliability
  - Test-retest reliability
- Reliability at different points of the ability continuum must also be the same
  - Psychometric matching
  - IRT

TICs from an Existing Global Cognition Scale and Re-Calibrated Existing Cognitive Tests



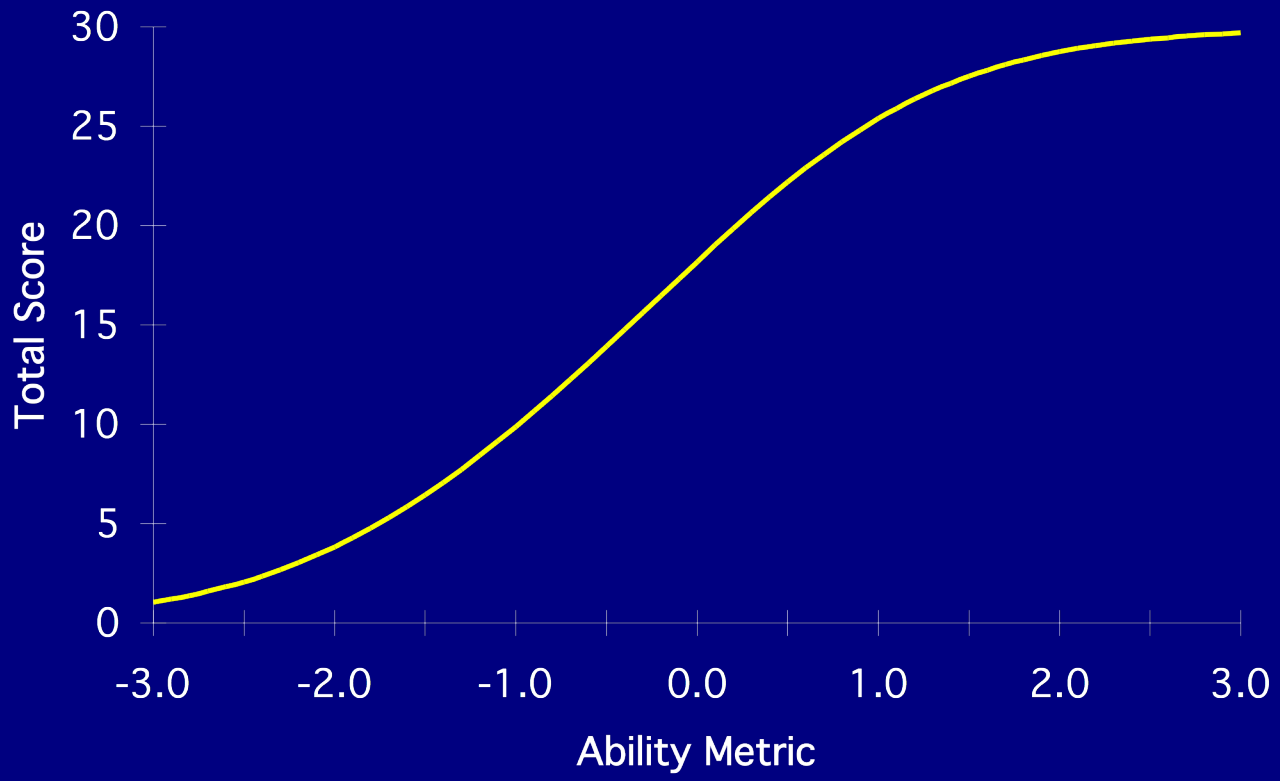
Mungas et al., Neuropsychology, 2003

Psychometric Properties of Equated Tests

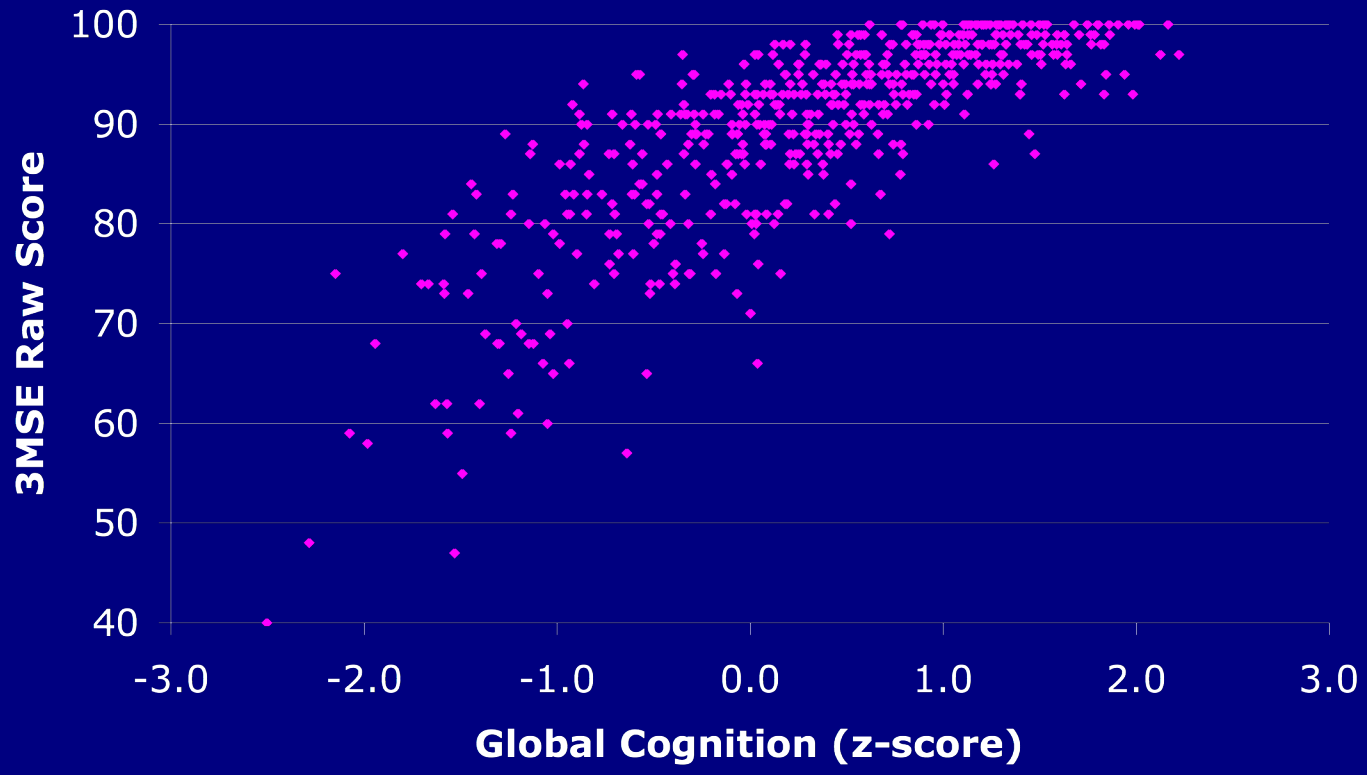
- Linear measurement is a nice property
  - Especially for longitudinal studies

Test Characteristic Curve

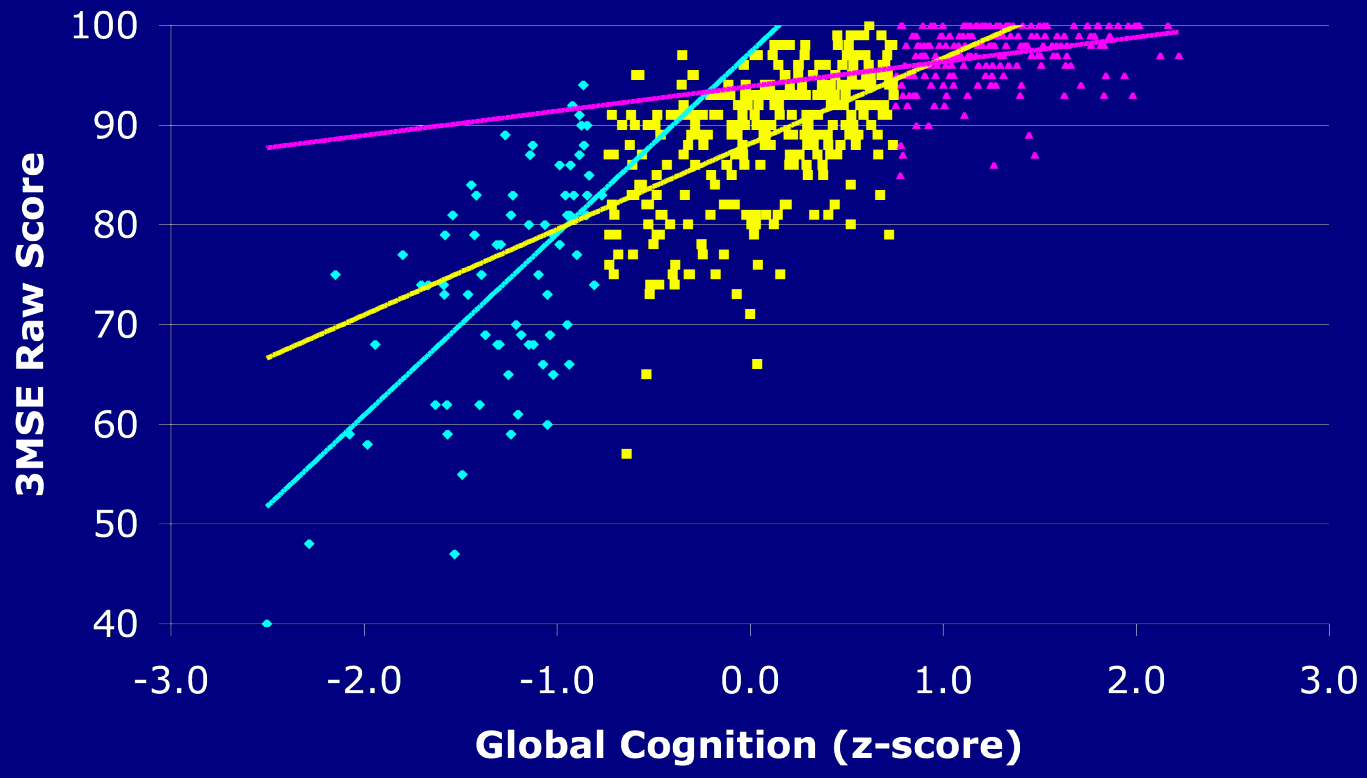
Mini-Mental State Examination



Relationship of 3MSE to Global Cognitive Ability



Relationship of 3MSE to Global Cognitive Ability



Designs for Test Equating

- Common item equating
  - Anchor items that are common to tests being equated
  - Sample overlap not necessary
- Common sample equating
  - Both tests administered to same sample
  - Item overlap not necessary
  - Sample should cover range of variability of target population
    - Distributions need not be identical

Statistical/Psychometric Approaches

- Item response theory (IRT)
  - Require item level data
- Non-IRT
  - Can be used with scale level scores

Statistical/Psychometric Approaches

- Test based
  - Cross-walk between tests used to create recoded test scores
  - Recoded scores entered into analysis
- Model based
  - Original test scores entered into analysis
  - Linking of scores occurs within analytic model

Issues for UDS

- Content
  - Close correspondence for some measures (Digit Span, Story Recall)
  - Apparent differences for MMSE and MOCA
- Psychometric characteristics
  - Empirical question
  - Design/selection suggests that this may be greater concern for MMSE - MOCA

Study Design Issues UDS

- Common sample equating most applicable
  - Sample size depends on method used for equating
  - Can have different samples for different test pairs
    - Common sample needed for to be equated pairs
    - Different samples could be used for different pairs
- Sample composition should roughly match target population (UDS enrollees)
- Order of test administration is important concern
  - Especially for Story Recall measures

Study Design Issues UDS

- Practice effects
  - Content learned over repeated administrations
    - Especially relevant to memory tasks
  - Familiarity with task
  - Sample with previous exposure to UDS tests is problematic
    - Familiarity with UDS 2.0, naive to UDS 3.0

Challenges for Equating UDS Tests

- Measure selection
- Study design
- Statistical and psychometric methods