Clarification on Some Misconceptions about Conditional Standard Errors of Measurement

Gregory M. Hurtz

California State University, Sacramento

SACRAMENTO STATE
Department of Psychology

MISCONCEPTION 1: TESTS ARE "MORE RELIABLE" AT THEIR UPPER AND LOWER EXTREMES

Misconception 1: Tests are "more reliable" at their upper and lower extremes

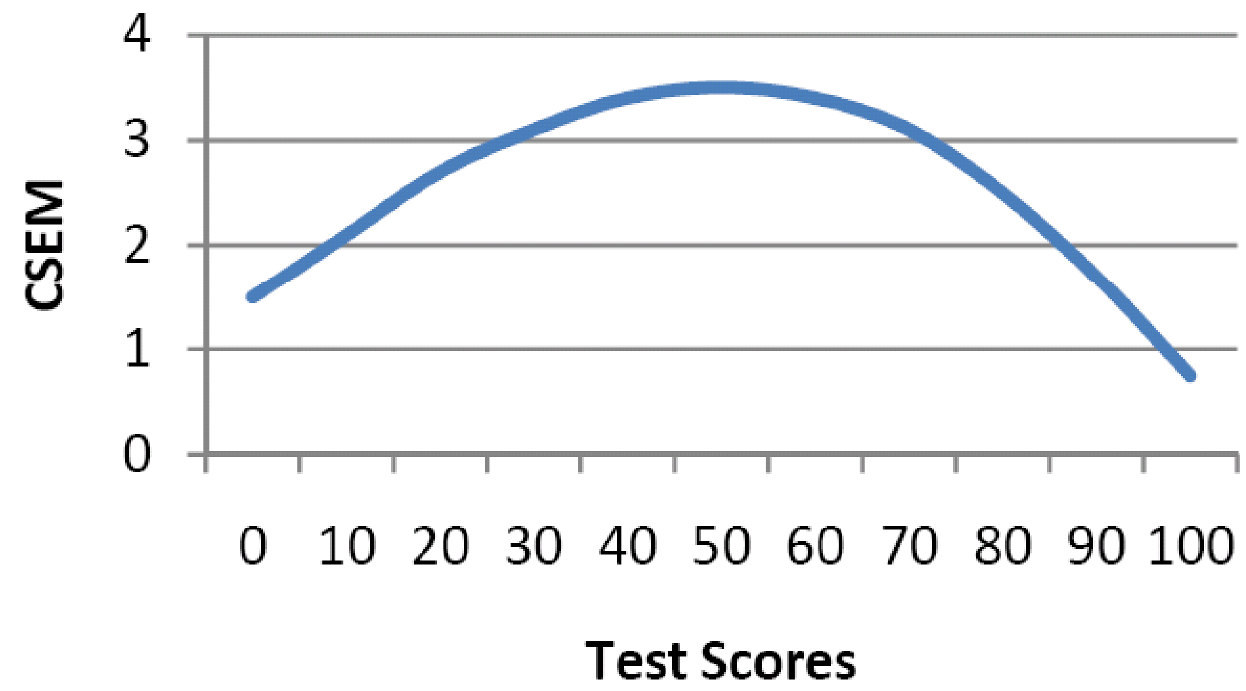
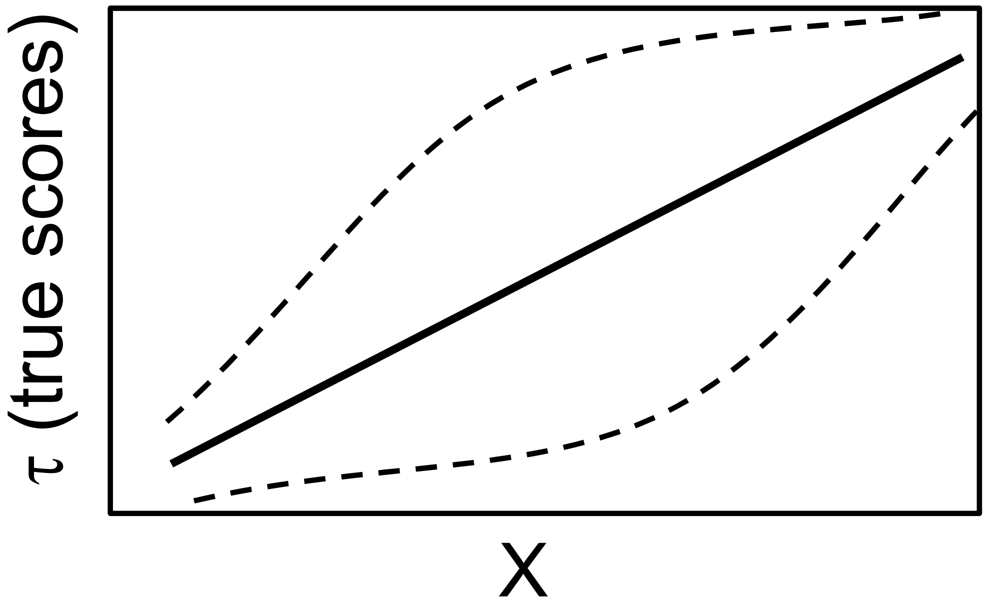


Figure 1. Characteristic shape of a CSEM function across test scores.

Misconception 1: Tests are "more reliable" at their upper and lower extremes



CTT model:

$X = \tau + e$

$\tau = X + e$

$$r_{xx} = \frac{\sigma_\tau^2}{\sigma_X^2}$$

- Linear regression roots of reliability:
  - r_{xx} = Estimated shared variance between X and τ
  - $\sqrt{r_{xx}}$ = Estimated correlation between X and τ

<u>Misconception 1:</u> Tests are "more reliable" at their upper and lower extremes

$$SEM = \sigma_X \sqrt{(1 - r_{XX})}$$

The variance in observed scores is what changes across X.

The slope can't change.

- In the linear regression model, this is heteroscedasticity.

<u>MISCONCEPTION 2:</u> TRUE SCORES ARE DIRECT INDEXES OF LATENT TRAIT LEVELS.

<u>Misconception 2:</u> True scores are direct indexes of latent trait levels.

- τ = "the mean of the theoretical distribution of X scores that would be found in repeated independent testing of the same person with the same test" (Allen & Yen, 1979, p. 57).

- τ is an estimate of a score <u>*on a particular test*</u>*.*

- It is tied directly to the operational definition of the construct, and only <u>in</u>directly to the construct itself.

Misconception 2: True scores are direct indexes of latent trait levels.

τ inherits the ceiling and floor of X.

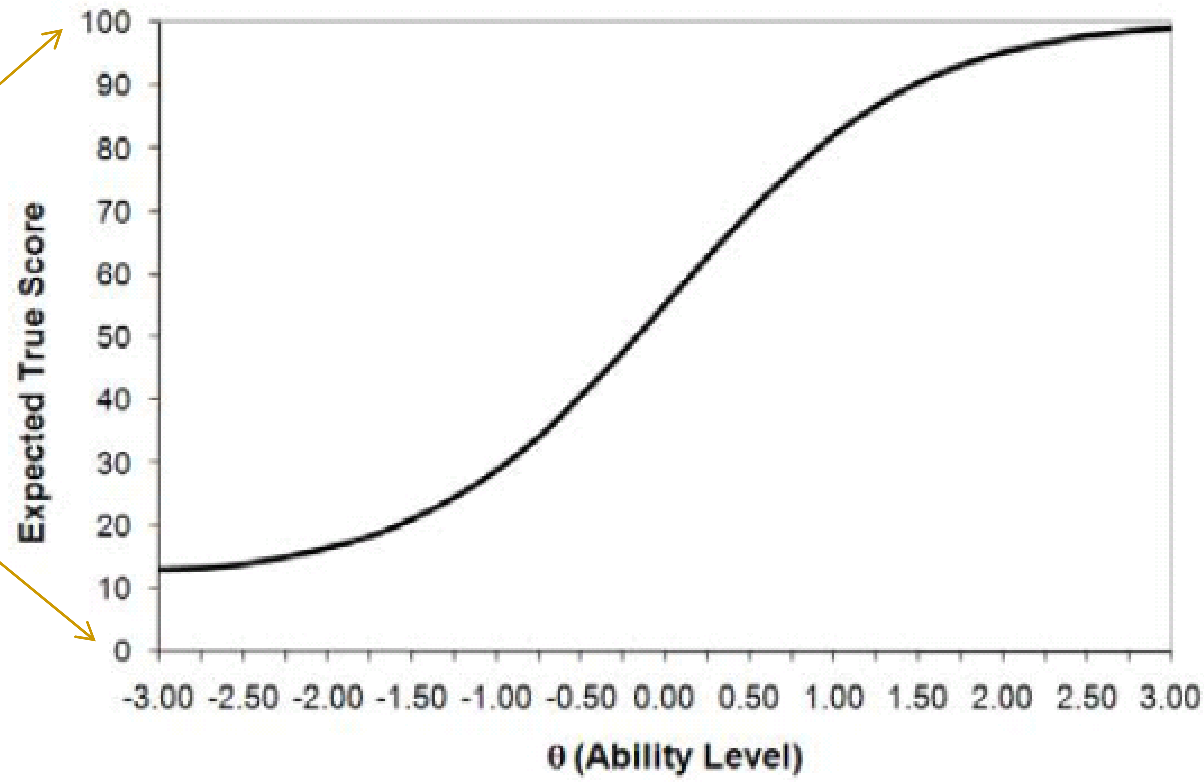


Figure 3. A test characteristic curve for a hypothetical 100-item test.

<u>MISCONCEPTION 3:</u>
CLASSICAL TESTING THEORY ONLY PROVIDES A STATIC SEM FORMULA.

Misconception 3: Classical Testing Theory only Provides a Static SEM Formula.

- Traditional SEM is static: $$\mathrm{SEM} = \sigma_x\sqrt{(1 - r_{xx})}$$

- SEE for τ is also static: $$\sigma_{\tau.x} = \sigma_\tau\sqrt{(1 - r_{xx})}$$

- These formulas are derived from linear regression: $$s_{y.x} = \sigma_y\sqrt{(1 - r_{xy}^2)}$$

Misconception 3: Classical Testing Theory only Provides a Static SEM Formula.

- Linear regression provides conditional standard errors:

$$s_{y'} = s_{y.x}\sqrt{\left[1 + \frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]}$$

- We can carry these over to the measurement context:

$$CSEE_{\tau'} = \sigma_{\tau.x}\sqrt{\left[1 + \frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]}$$

<u>Misconception 3:</u> Classical Testing Theory only Provides a Static SEM Formula.

This formula predicts the opposite pattern from traditional CSEM methods.

The compression of scores at the ceiling and floor reverse the pattern when estimating CSEMs from empirical data.

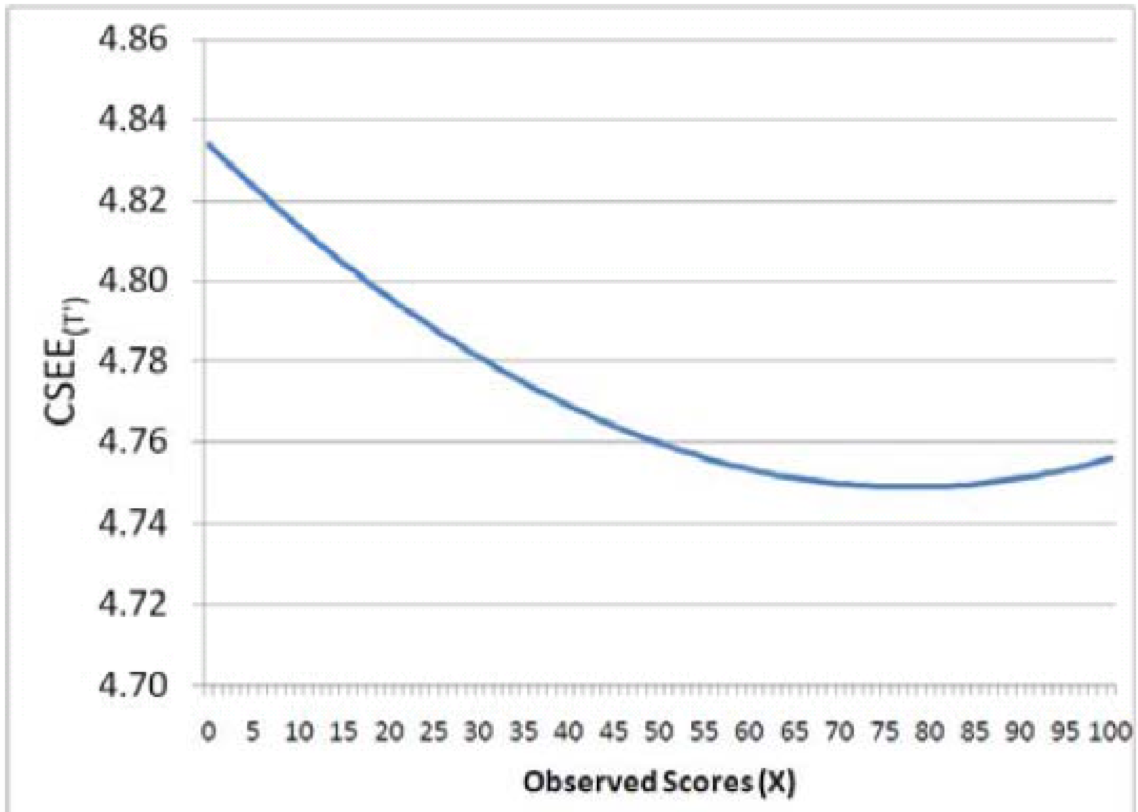
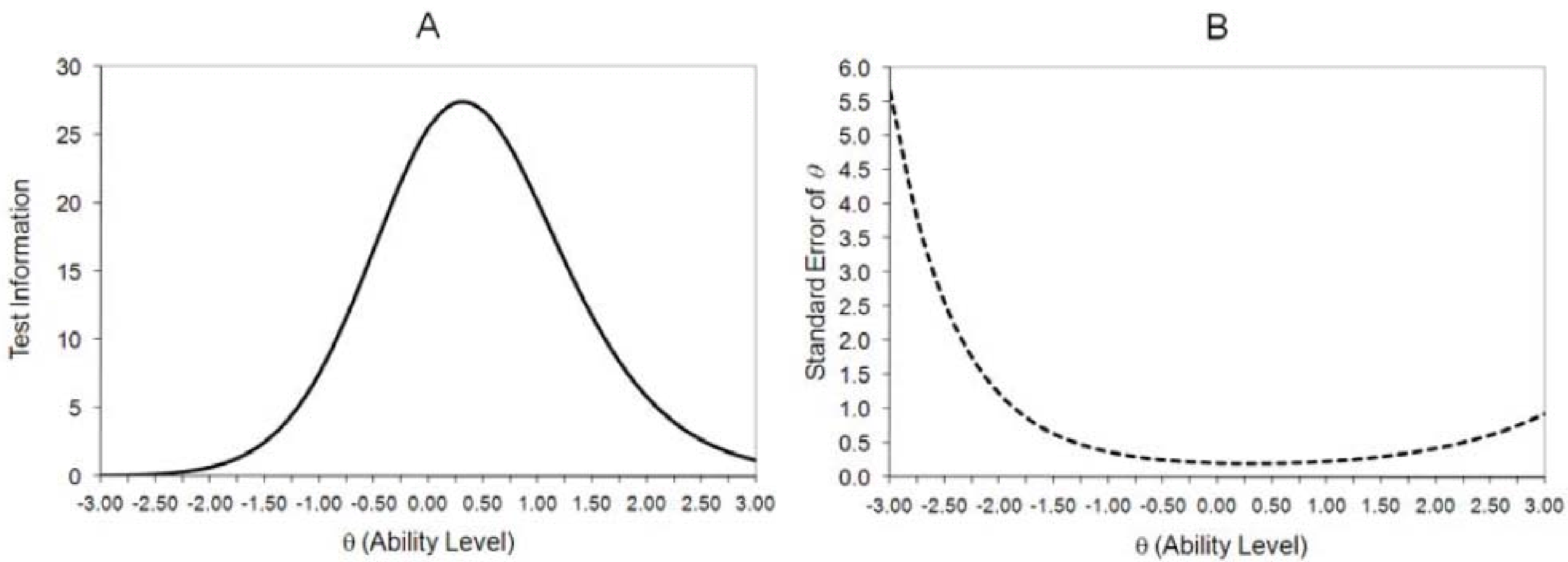


Figure 4. Regression-based conditional standard error of estimate calculations for individual scores on a test.

SACRAMENTO STATE
Department of Psychology

MISCONCEPTION 4: IRT CSEMS CONTRADICT CLASSICAL TEST THEORY CSEMS IN THEIR SHAPE.

Misconception 4: IRT CSEMs contradict classical test theory CSEMs in their shape.



<u>Misconception 4:</u> IRT CSEMs contradict classical test theory CSEMs in their shape.

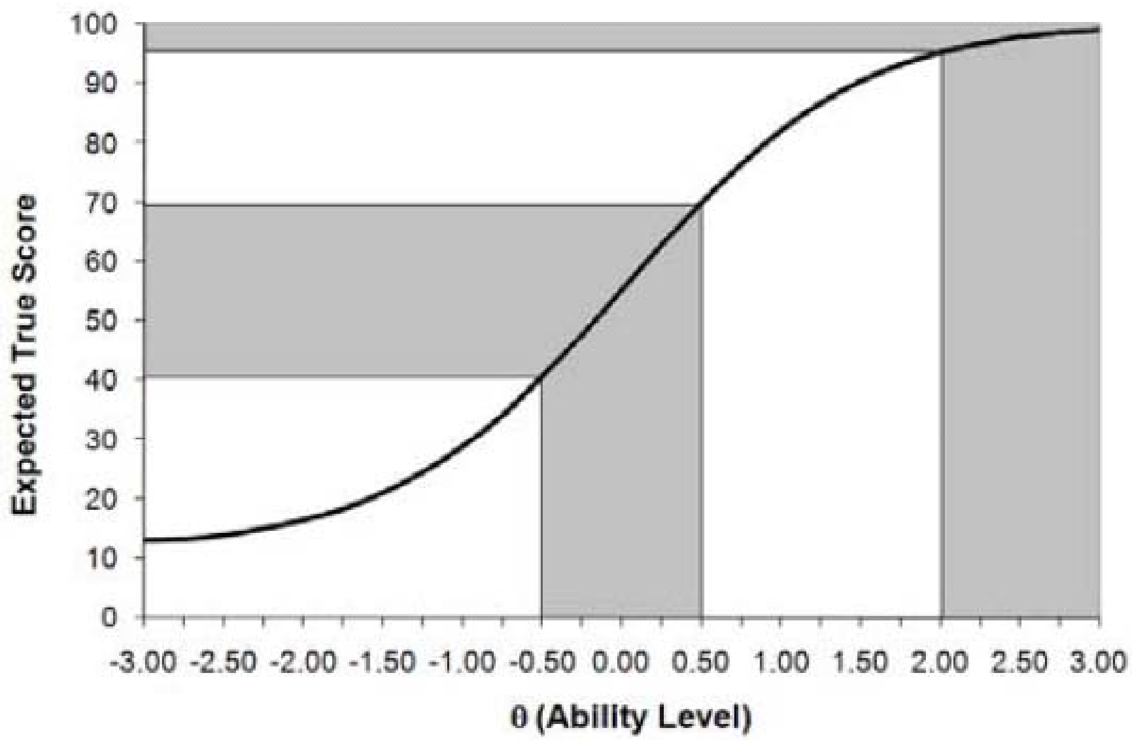


Figure 6. Compression of variance of high latent ability estimates when converted to the true score scale.

Practical Implications

- Be clear about the scale (τ vs. θ).
- If smaller CSEMs are found near the test ceiling, this isn't necessarily good for top-down selection.
- If score banding will be used, spuriously low CSEMs near the ceiling will yield spuriously narrow bands.
- If a cutoff score is applied somewhere in the middle of the score distribution, then the compression of top scores is of little concern.