Interpreting GRE Scores: Reliability and Validity

Dr. Lesa Hoffman
Department of Psychology
University of Nebraska-Lincoln

All data from: *GRE Guide to the Use of Scores 2010-2011*
http://www.ets.org/s/gre/pdf/gre_guide.pdf

GRE Item Selection and Scoring (p. 4-5)

- **Verbal and Quantitative Tests are *computer-adaptive*:**
  - Start with an "average" item
  - Administer more/less difficult items when correct/incorrect
  - Most likely score after completing (?) number of items is given

- **Analytical Writing (2 essays) Test is graded by a trained reader**
  - Computer version is also graded by a computer program
  - If human and computer agree, uses human score
  - If disagree, add second human rater, average human ratings
  - Paper version: average two human ratings (add third for discrepancies)

- **Psychology Subject Test is paper only (same items for all)**
  - Score = (# correct) – (¼ * # incorrect)

GRE Score Units: Current (p. 13) vs. New

- **Current and New GRE Scoring: Analytical Writing**
  - 0 to 6 in 0.5-point increments (13 different scores)
  - Mean ≈ 3.8, SD ≈ 0.9

- **Current GRE Scoring: Verbal and Quantitative**
  - 200 to 800 in 10-point increments (61 different scores)
  - Verbal Mean ≈ 456, SD ≈ 120; Quant Mean ≈ 590, SD ≈ 150

- **New GRE Scoring: Verbal and Quantitative**
  - 130 to 170 in 1-point increments (41 different scores)
  - **Wait… Isn't fewer scores (less precision) going to be bad?**

Validity and Reliability of GRE Scores

- **Validity: *Measuring the right thing…***
  - Extent to which a test measures *what it is supposed to* (i.e., the test does what it is intended to do)
  - **GRE is supposed to predict performance in graduate school** and thus serves as a criterion for entry in most programs

- **Reliability: *Measuring the thing right…***
  - Extent to which a test does what it is supposed to *with sufficient consistency* for its intended usage
  - Extent to which **same GRE test score** would be obtained after repeated trials by the same person…
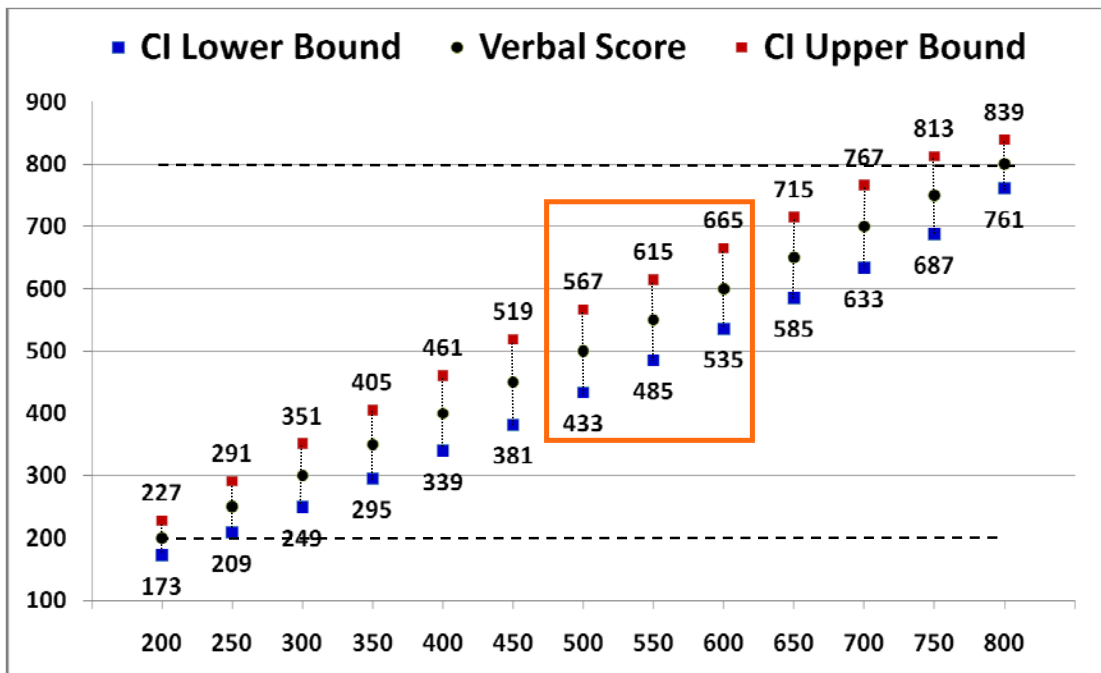
Assessing Reliability of GRE Scores

- ETS estimates the reliability of the Verbal and Quantitative Tests through **computer simulation** (p. 20)
  - ➢ Administer GRE items to sample of fake people w/ known scores
  - ➢ See how far off the predicted scores are from the known scores
  - ➢ No details are given about the simulation process in this guide

- How? Predictions from **Item Response Theory Model (IRT)**
  - ➢ Prob(item i is correct) =

  $$c_{item} + \left(1 - c_{item}\right)\left\{\frac{\exp\left[a_{item}(Score_{person} - b_{item})\right]}{1 + \exp\left[a_{item}(Score_{person} - b_{item})\right]}\right\}$$

  - ➢ a_{item} = item discrimination (how good item is)
  - ➢ b_{item} = item difficulty, c_{item} = lower asymptote (guessing)
  - ➢ $Score_{person}$ = ability of person answering the item

Quantifying Reliability: SEM (p. 12)

- **Conditional Standard Error of Measurement (SEM)**
  - ➢ Standard deviation of the sampling distribution around each possible test score (given on p. 20)
  - ➢ Average discrepancy between "true" and "observed" test scores across repeated assessments for a given person
  - ➢ "Conditional" ➔ Differs across range of test scores

- **95% confidence interval (CI) around each score**:
  - ➢ CI = Obtained Score ± 1.96*SEM
  - ➢ Expected range in which actual scores should fall
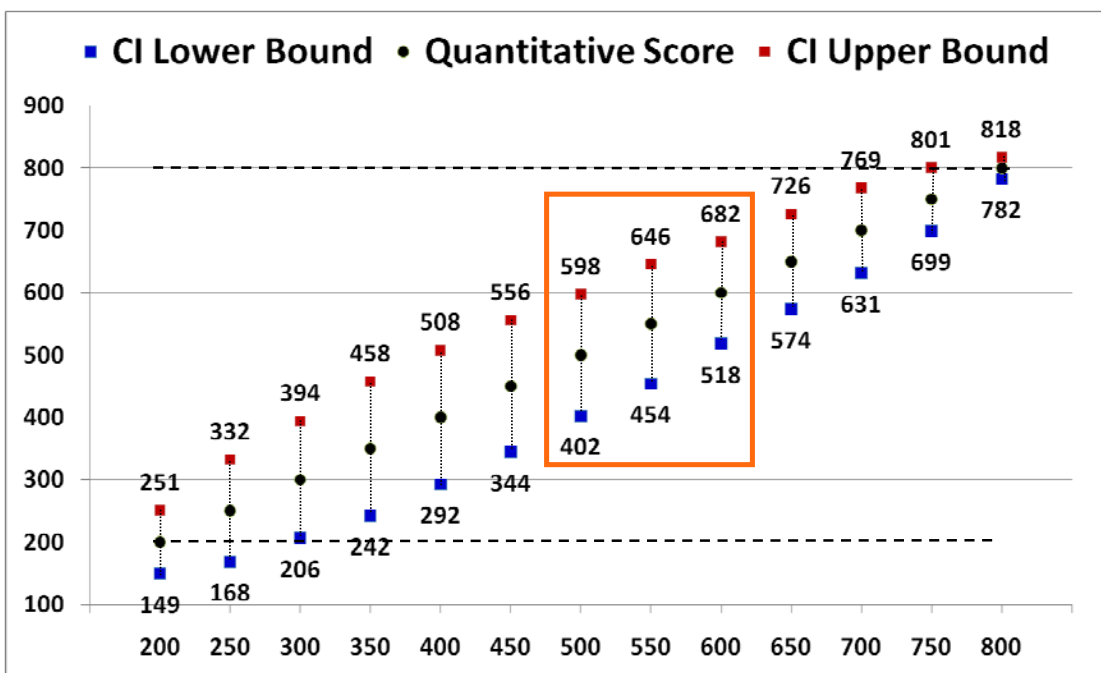   *if the test was given repeatedly to the same person*

95% Confidence Intervals: Verbal
SEM ranges from 14 to 35

95% Confidence Intervals: Quantitative
SEM ranges from 9 to 55

Reliability of GRE Scores

- **Analytical Writing SEM = 0.40** (out of 0 to 6 possible)
  - ➢ Score of 1 ≈ 0.2 to 1.8   Score of 2 ≈ 1.2 to 2.8
  - ➢ Score of 3 ≈ 2.2 to 3.8   Score of 4 ≈ 3.2 to 4.8
  - ➢ Score of 5 ≈ 4.2 to 5.8   Score of 6 ≈ 5.2 to 6.8

- So what do these assessments of reliability mean?
  - ➢ **GRE scores are far less precise** than most people realize
  - ➢ According to ETS (p. 6): "**Small differences** in GRE scores (**as defined by the SEM**) should not be used to make distinctions among examinees"
    - ◆ A "**small difference**" at Verbal = 450 is **35**!
    - ◆ A "**small difference**" at Quant = 500 is **50**!

Validity of GRE Scores (p. 21)?

- GRE scores are used to predict success in grad school… What is the **evidence for their validity** in doing so?

Indicators of Success	R^2 for Verbal	R^2 for Quant.
First Year Graduate GPA	12%	14%
Comprehensive Exam Scores	19%	7%
Faculty Ratings	18%	22%
Degree Attainment	3%	4%

- So what does ETS recommend about the use of their GRE scores in admitting students?

GRE for Admission into Grad School?

- p. 3: "The content validity of the tests for a graduate department should be determined by reviewing each test carefully and then making subjective decisions as to the weight, **if any**, the scores on the GRE tests should receive in relation to other admissions factors."

- p 9: "The GRE board believes that GRE scores should never be used as the sole basis for an admissions decision and that it is inadvisable to reject an applicant on the basis of GRE scores. **A cutoff score** below which every applicant is categorically rejected without consideration of any other information **should not be used**."

- p. 6: "Since the level of skills in verbal, quantitative, and analytical writing abilities required for success in graduate school varies by field … [these] **scores should not be combined**."

Summary and Conclusions

- GRE Verbal and Quantitative Test Scores:
  - **Current**: 200 to 800 in 10-point increments (61 different scores)
  - **New**: 130 to 170 in 1-point increments (41 different scores)

- Reliability of Current GRE Scores

  - GRE scores are **far less precise** than most people realize
    - e.g., 500 quant really means somewhere between 400 and 600

  - So reducing the number of possible new GRE scores may not be problematic if the distinctions between GRE scores were not that reliable to begin with!

Summary and Conclusions

- GRE scores also have **minimal predictive validity**
  - Good GRE scores do not necessarily mean you are prepared to do well in grad school (and vice-versa)
  - The GRE creators (ETS) explicitly state that other factors should be considered in admitting grad students…

- So don't neglect those other factors!
  - Classes and GPA only matter so much
  - **Experiences** that are relevant for grad school matter more
    - Research, internship, clinical/volunteer experience
  - **FIT FIT FIT** to the program and potential faculty supervisor
    - Will be the deciding factor between equally qualified applicants

Further Questions?

- Psi Chi Meeting Monday March 7 at 4 PM in 79 Burnett
  - Presentations by and Q&A with faculty about how to select graduate programs and what they look for in an applicant
  - Anyone is welcome to attend!

- Questions or comments about the presentation?
  - Email Dr. Lesa Hoffman: **Lhoffman2@unl.edu**
  - Handouts for both presentations can be found at the UNL Psi Chi web page: **http://www.unl.edu/psypage/psichi/index.shtml**

- Thank you for your attention!