



# **The Right Answer is Not Always C: Inside the Test Development Team at SSATB**

Keith D. Wright, Ph.D.

Secondary School Admission Test Board

Presentation at the SSATB Annual Meeting 2015

(Las Vegas)



# Overview/Learning Objectives

- Standardized Testing
  - Test optional debate, why testing matters?
  - How do we ensure fairness?
- Common Misconceptions about SSAT Scores
- Repeater Analysis
- Construct Validity, Predictive Validity
- Q&A

# STANDARDIZED TESTING

# Test Optional Debate **Equity Era (1960s)**

- The Civil Rights Act of 1964
- The Coleman Report
- ETS Awarded Contract- Coleman Report Survey
- Title I Elementary and Secondary Education Act of 1965
- Title I / Standardized Testing

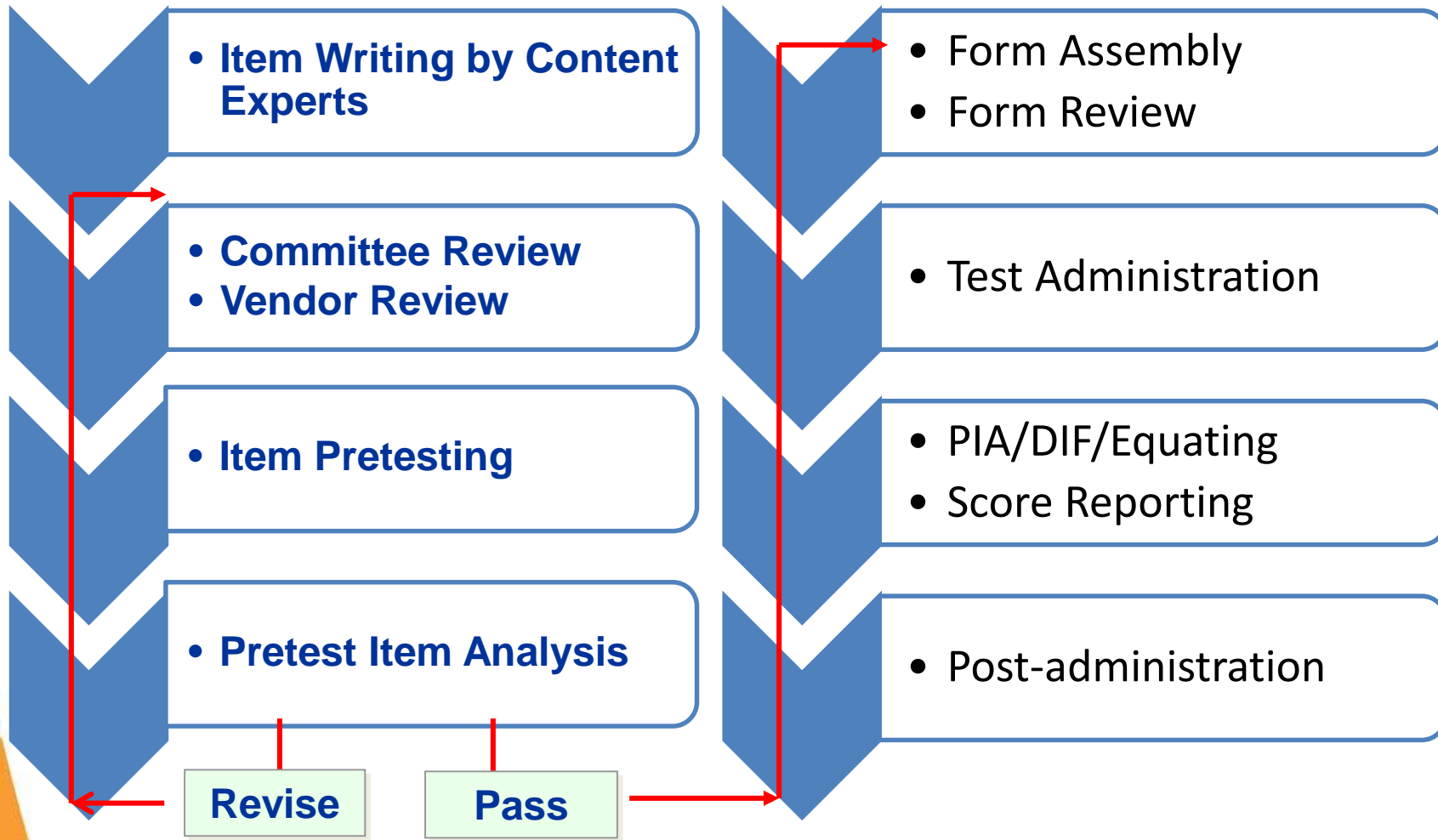
# Why Testing Matters?

- Don't "Kill The Messenger" it is about fairness
- What other standard method exists to assess cognitive ability?
- The debate should be around.....the societal factors
- How are you using this one important piece of information?
- See Howard Wainer, "Uneducated Guesses- Using Evidence to Uncover Misguided Education Policies"

# How Do We Ensure Fairness?

- SSATB Test Cycle
- DIF Analysis
- Score Equating

# SSATB Test Cycle



# Item Fairness

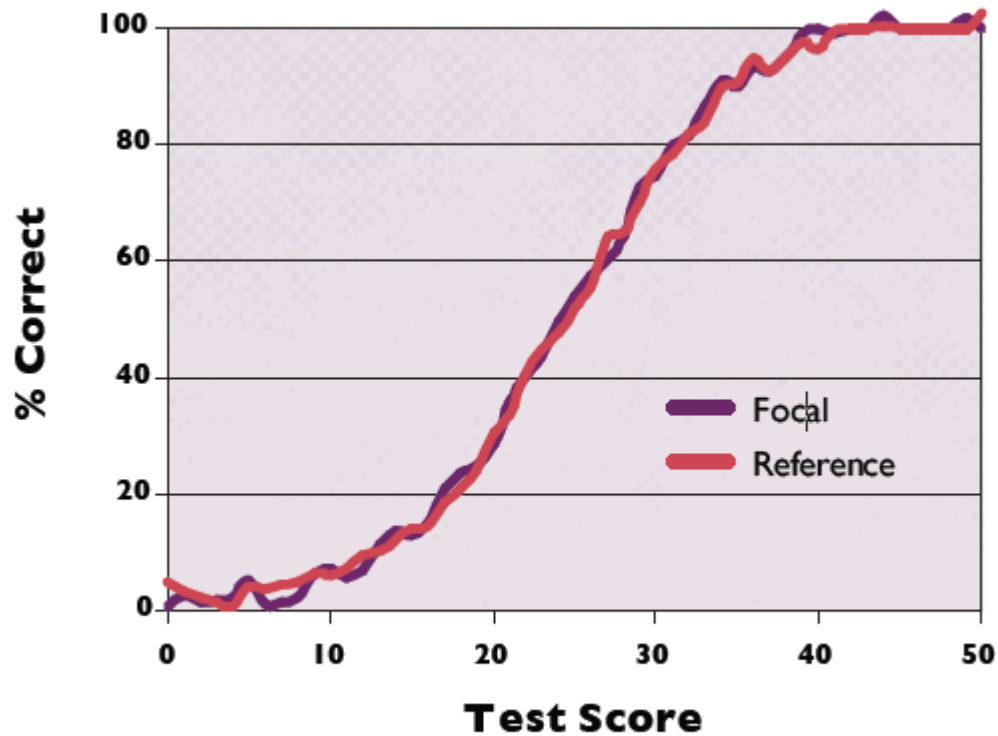
## Differential Item Functioning (DIF)

- A statistical measure of whether items are potentially “biased” towards/against one group versus another
  - Reference Group, Focal Group
  - Item Impact (Desirable)—an item performs differently in different subgroups of examinees based on proficiency
  - Item DIF (Not Desirable)—Differential Item Functioning; again, a statistical measure indicating that impact remains after the reference and focal groups have been matched on proficiency



# DIF (cont.)

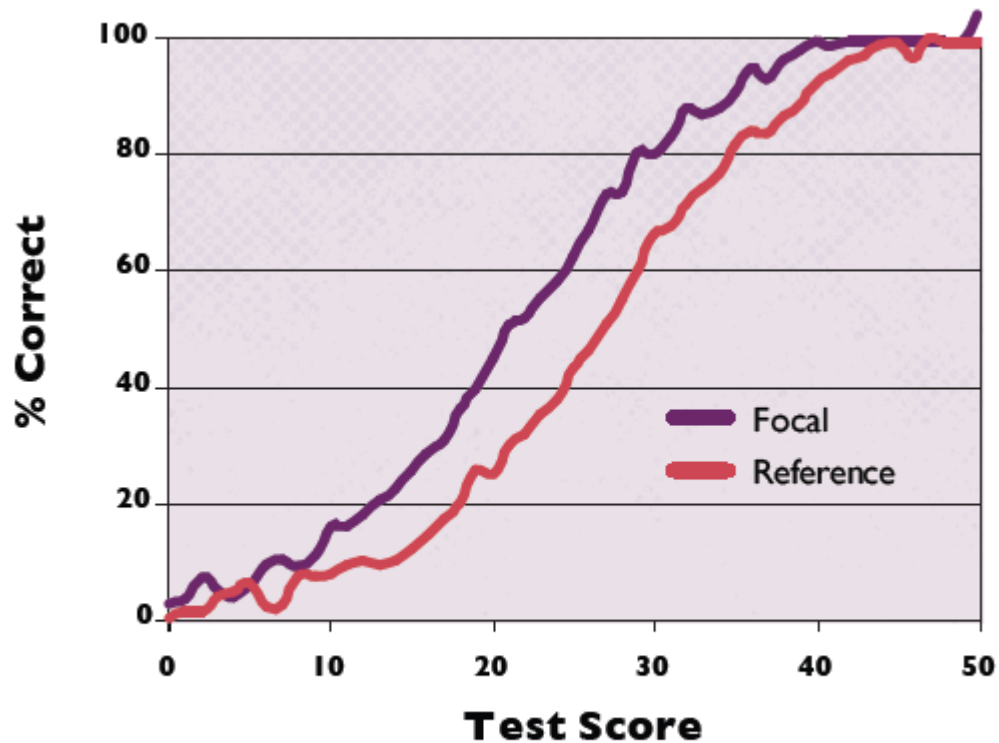
Test Item – Displaying No DIF



**Figure I. Differential item functioning –  
An item showing no DIF**

# DIF (cont.)

Test Item – Displaying DIF



**Figure 2. Differential item functioning –  
An item showing DIF**

# Test Score Equating

- Equating is a statistical procedure to adjust the form difficulty differences so that scores from different test forms have the same meaning and can be used interchangeably

# Why equating? Let's take a look!

Group A answered 50 items correctly on Form X;  
Group B answered 50 items correctly on Form Y.  
Do the two groups, A and B, have the same ability level?

- Forms X and Y have the same difficulty level;  
A is as capable as B;  $50 = 50$ ;
- Form X is easier than Form Y;  
A is less capable than B;  $50 \neq 50$ ;
- Form X is harder than Form Y;  
A is more capable than B;  $50 \neq 50$

# Example

DEC '13

(Dave)

25/50

DEC '14

(Mike)

25/50

MAR '15

(Carinne)

30/50

(Julie)

25/50

# Oversight





# **COMMON MISCONCEPTIONS ABOUT SSAT SCORES**

# Number of Rights, Wrongs and Omits

“Can the the number of rights, wrongs and omits be used to

- Evaluate a student’s growth over time on a particular sub-content area (e.g., algebra), and
  - Compare different students (who may take different forms)?”
- No. Inappropriate use
    - Number of rights, wrongs and omits are confounded results of item difficulties and test takers’ abilities
    - They should NOT be compared across different forms, either on the same student or between different students



# Is there an Even Distribution of Percentiles for Each Administration?

- Not necessarily
  - Due to the nature of self-selection, groups who choose to take different administrations are different. Certain groups are more capable than others. Hence, it is possible that more test takers from a more capable group may reach 99 percentile rank than a less capable group.

# REPEATER ANALYSIS

# Will My Score Stay the Same if I Take the Test Again?

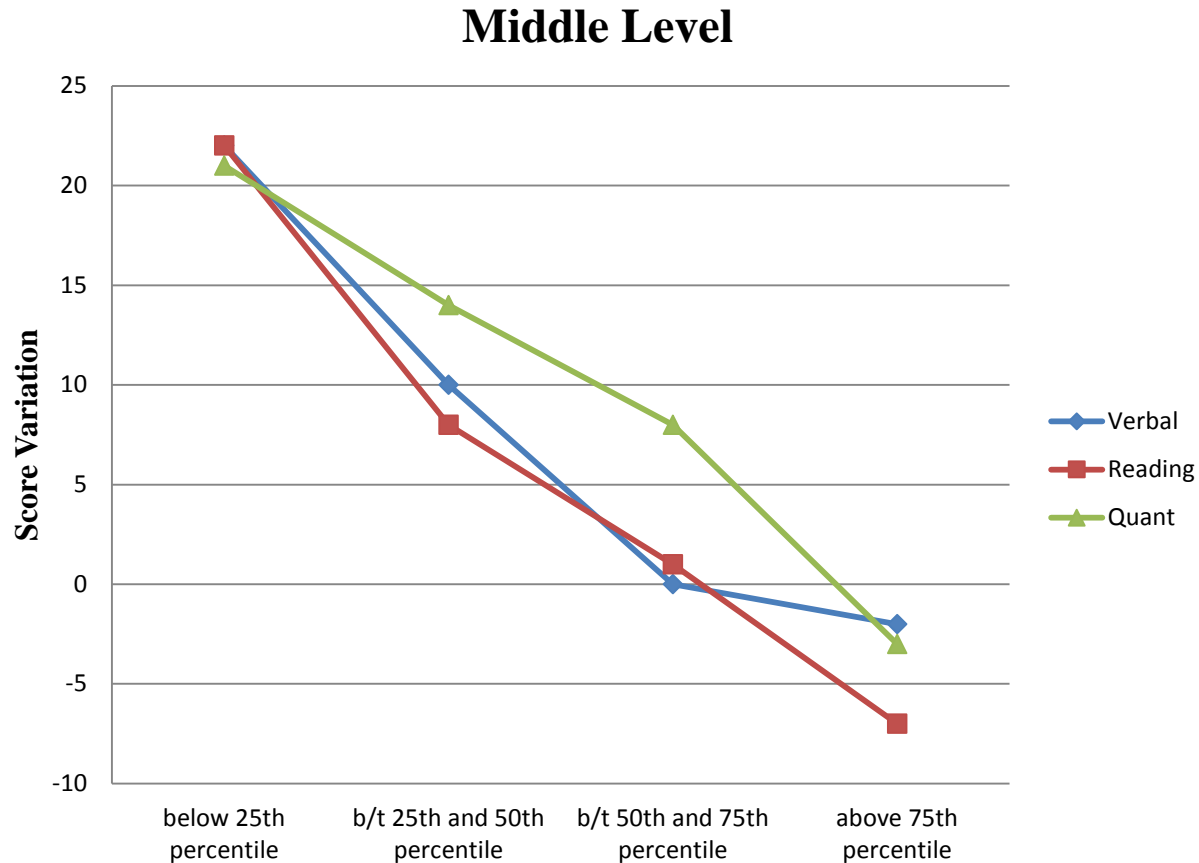
- Not necessarily
- Test takers' score tend to change if they re-take the test
- In general, the lower the test takers' initial scores, the more likely their 2<sup>nd</sup>-time scores will increase; the higher the test takers' initial scores, the more likely their 2<sup>nd</sup>-time scores will decrease

# The Research Study

- We examined test taking patterns and the score changes on the middle level and upper level test across testing year 2013-14 (October 2013 to June 2014).
- The research study attempted to answer some of the following questions:
  - How many test takers take the SSAT more than once?
  - What are the most common test taking patterns? How is the test taking pattern related to the initial scores?
  - How do SSAT scores change between consecutive tests?
  - How do SSAT scores change conditioned on initial scores?
  - What is the effect of the interval length between repeating the test on score change?

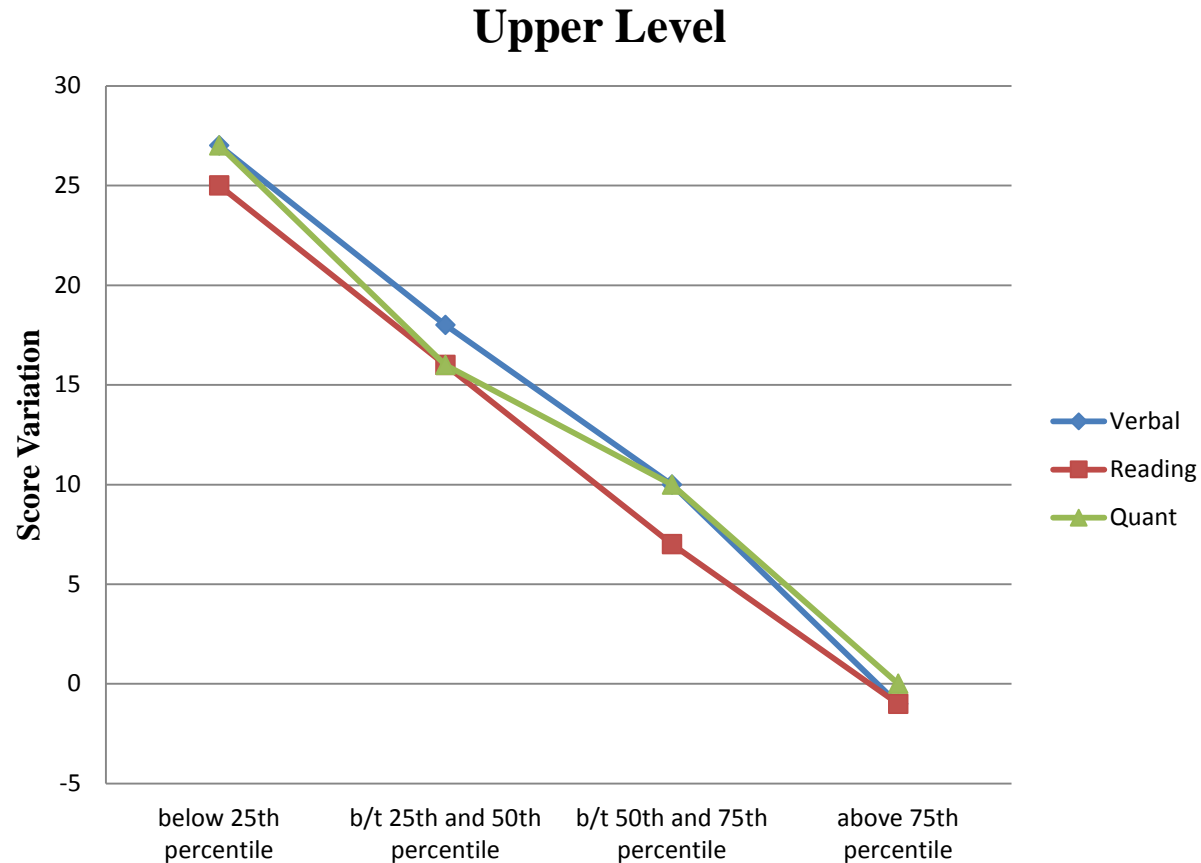
# Middle-Level Summary (2013-14)

*Average Score Change for Those Who Repeated a Second Time*



# Upper-Level Summary (2013-14)

*Average Score Change for Those Who Repeated a Second Time*



# TEST VALIDITY

- If reliability tells us that we are measuring something consistently, validity deals with the issue of whether or not we are actually measuring what we want to measure.



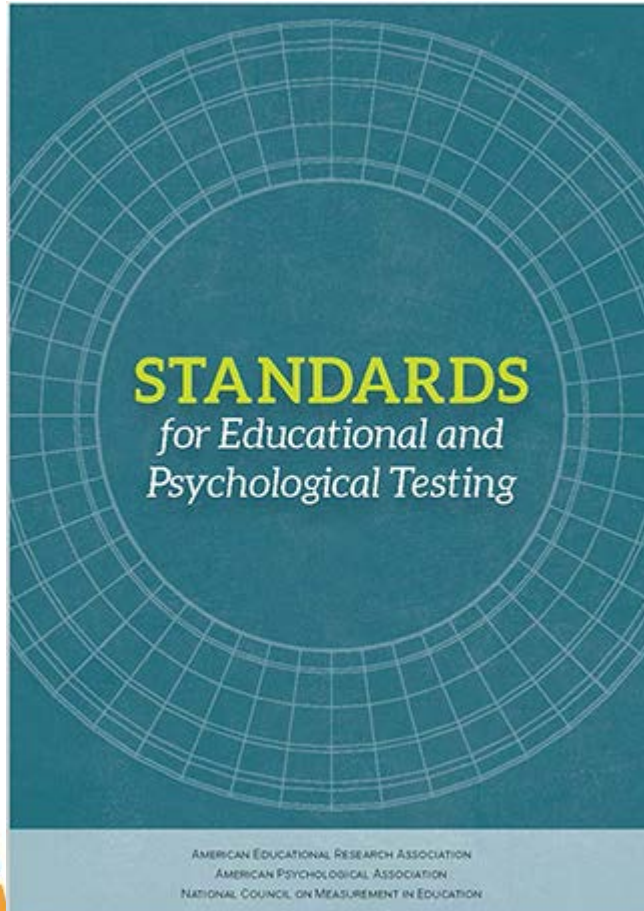
# Three Types of Test Validity

- Content validity (Committees)
- Construct validity (Research)
- Predictive validity (Research)

# Predictive validity (Research)

- The SSAT is designed to measure the basic verbal, quantitative, and reading skills students develop over time — skills that are needed for successful performance in private schools. Independent school admission professionals use the SSAT to project academic success for applicants to their schools.
- The SSATB is currently conducting a predictive validity study to examine how well the SSAT can predict student performance, both academically and in other standardized tests. Your support and participation in this study will be crucial in helping us gather evidence of test validity.
- The study will analyze the ability of the SSAT to predict first-year GPA (FYGPA) and the ability of each SSAT section to predict FYGPA.

# In Summary



- Standards for Educational and Psychological Testing (AERA, APA, NCME)
- Highest quality, reliable, valid and most importantly FAIR for all students taking the test.

# Questions & Answers



**Thank you!**

