THE PURPOSE AND VALIDITY OF LICENSURE EXAMINATIONS:
A Statistical Perspective

Board of Behavioral Sciences
Licensing Committee Meeting
October 27, 2023

Robert Calvert, PhD, Senior Statistician
Heidi Lincer, PhD, Chief

Overview

Examinations Race and Bias

Exploring Differences in Pass Rates

Caution when Consuming Research

Part 1: Examinations Race and Bias

Licensure examinations are designed to provide a standardized measure of minimum competence for independent practice, regardless of any variability in education, training, or experience programs.

- Standardized academic tests consistently show significant differences based on race and ethnicity.

- Association of Social Work Boards(ASWB) recently released data showing a large difference in first time pass rates for Black candidates when compared to other groups.

Do We Agree?

- Systemic racism is real, powerful, and ubiquitous
- Valid examinations show differences where they exist.
- Counselling is a skill, and that skill can be measured with an acceptable degree of error.
- Licensing entities exist to protect the public from unqualified practitioners.

The Problem

How can a licensing entity make licensure decisions without disproportionately excluding disenfranchised minorities from the profession?

The Solution

Choose the option that gives you the greatest ability to limit construct irrelevant variables while maximizing information used for licensure decisions.

Operational Definitions

- Difference: a statistically significant difference exists when one group is compared with others
- Construct-irrelevant variables: race, ethnicity, age, gender, test-taking ability, test-anxiety, etc.
- Biased: a difference exists based on construct irrelevant factors

Options for Licensure Decisions

- Standardized Tests

- Oral or Essay Examinations

- Internship/Experience

- Education

Tools to Minimize Bias of Standardized Tests

- **Representative SMEs (race, age, gender, experience)**
  - Item Writing (focus on limiting extraneous variables)
  - Exam Construction
  - Passing Score Determination
  - Item Review for content as well as potential bias
- **Editors/committees** that are trained to identify items that may be culturally biased
- **Statistical analysis** of item performance and examination performance

Recap of Options

	SME Involvement	Statistical Methods to Evaluate Bias	Relies on Differences for Potential Bias	Benefit to Entity for Candidate Passing	Licensing Entity Oversight
Standardized Examinations	High	Yes	No	Low	High
Written or Oral Examination	Medium	Difficult	No	Low	High
Internship/ Experience	Low	No	Yes	Medium	Medium
Education	Very Low	No	Yes	High	Low

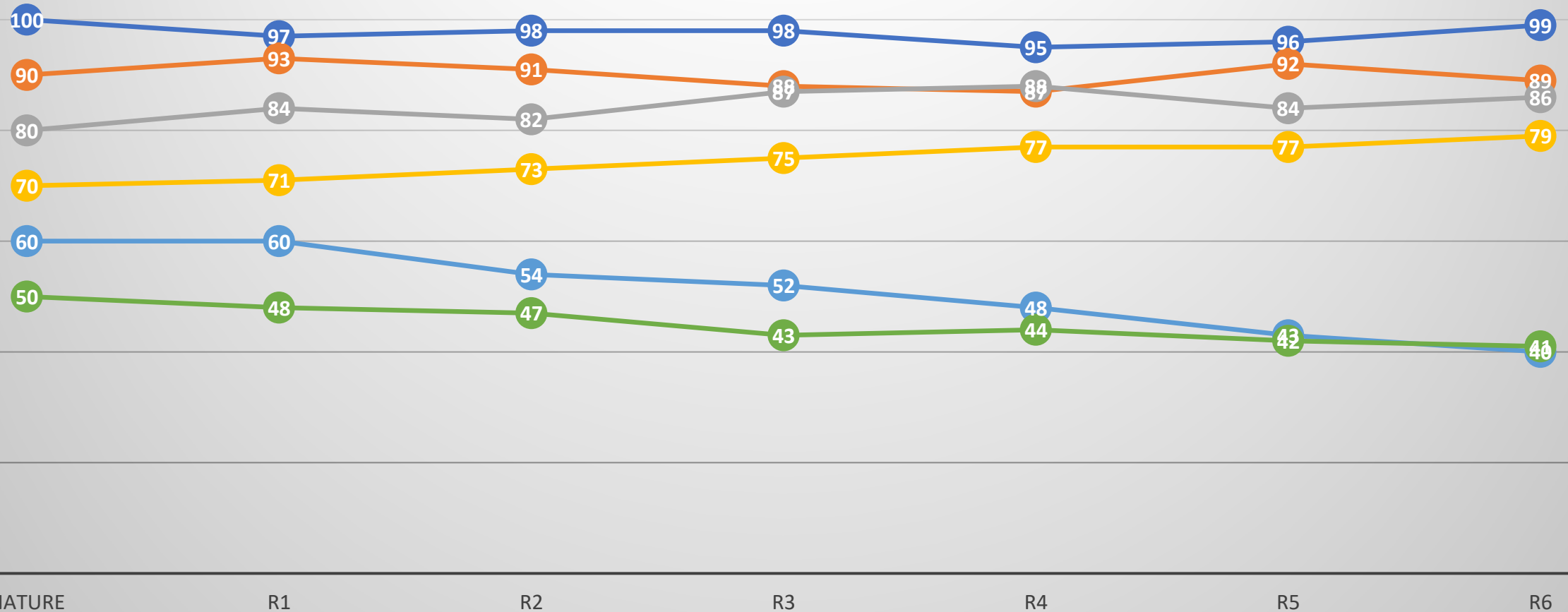# The Problem with Differences and Bias

- Group differences may exist without bias
- A valid examination will show differences that exist.
- There are many factors that can contribute to differences.
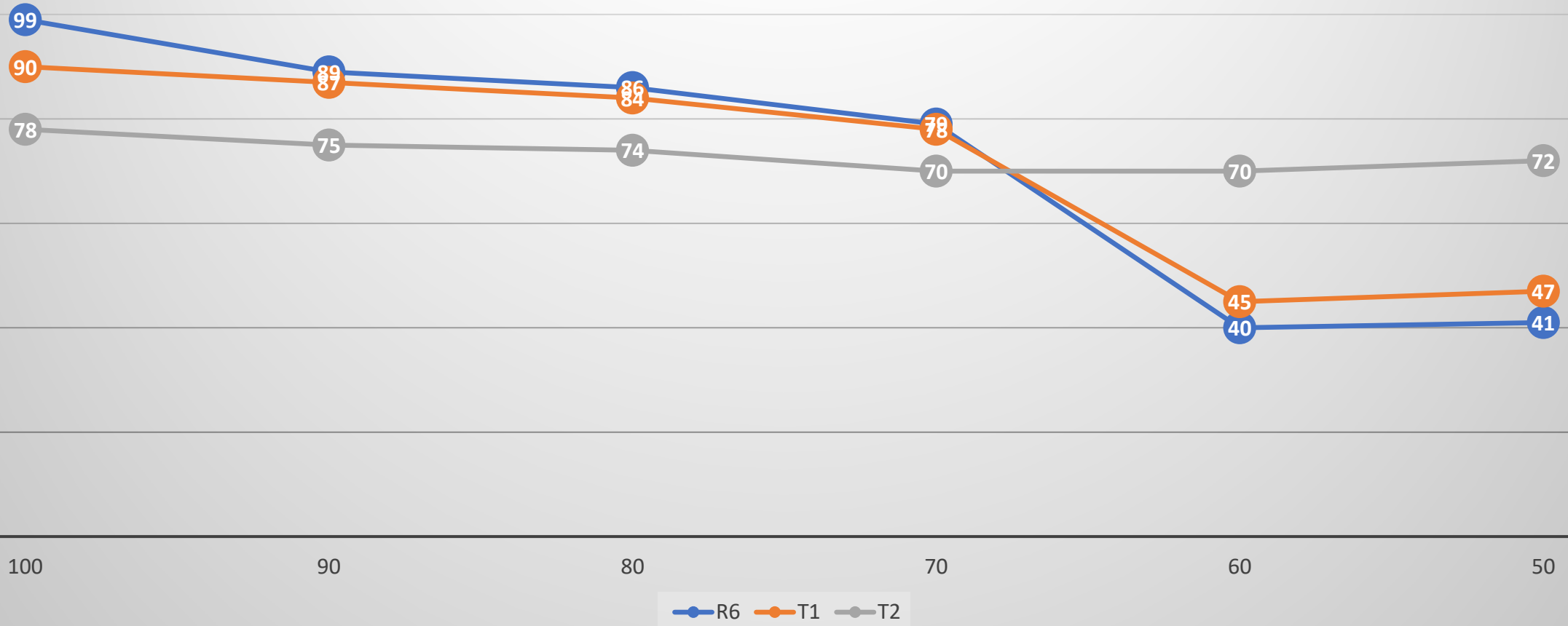
Part 2: Exploring Differences in Pass Rates

Hypothetical Examination Scenario

Examination	Test 1	Test 2
Group 1	72%	81%
Group 2	85%	80%

Group 1 Experiences

	NATURE	R1	R2	R3	R4	R5	R6
Blue	100	97	98	98	95	96	99
Orange	90	93	91	88	87	92	89
Gray	80	84	82	89	88	84	86
Yellow	70	71	73	75	77	77	79
Light Blue	60	60	54	52	48	43	40
Green	50	48	47	43	44	42	41

Group 1 Examinations

Group 2 Examinations
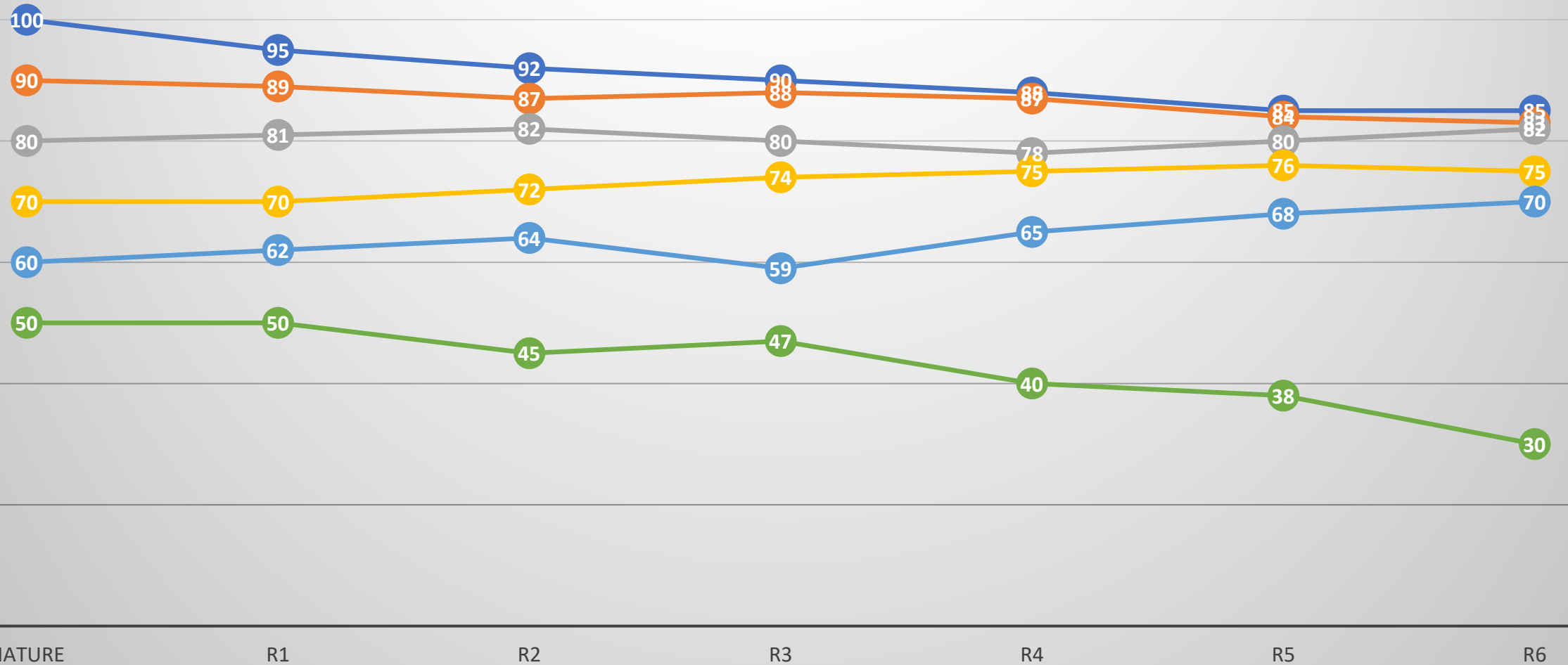
	100	90	80	70	60	50
R6	85	83	82	75	70	30
T1	86	80	83	77	73	45
T2	78	75	76	73	—	65

R6   T1   T2
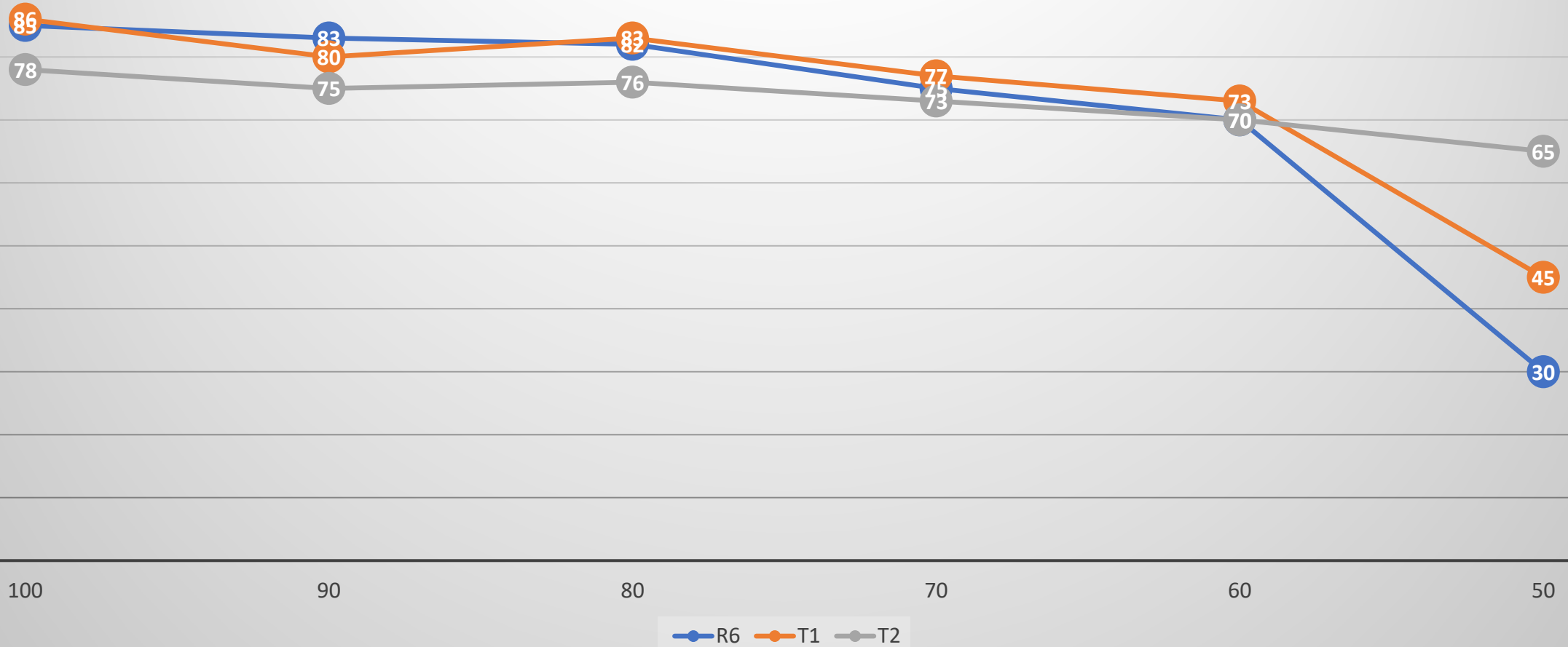
Hypothetical Examination Scenario
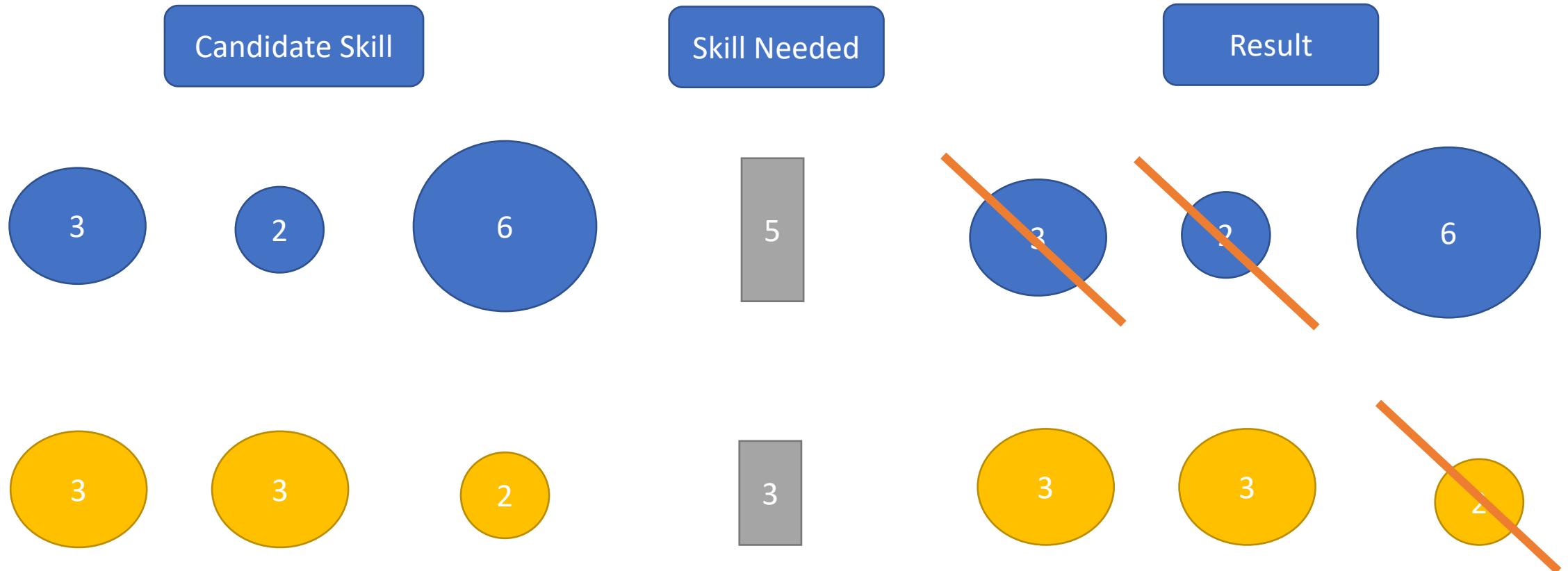
Examination	Test 1	Test 2
Group 1	72%	81%
Group 2	85%	80%

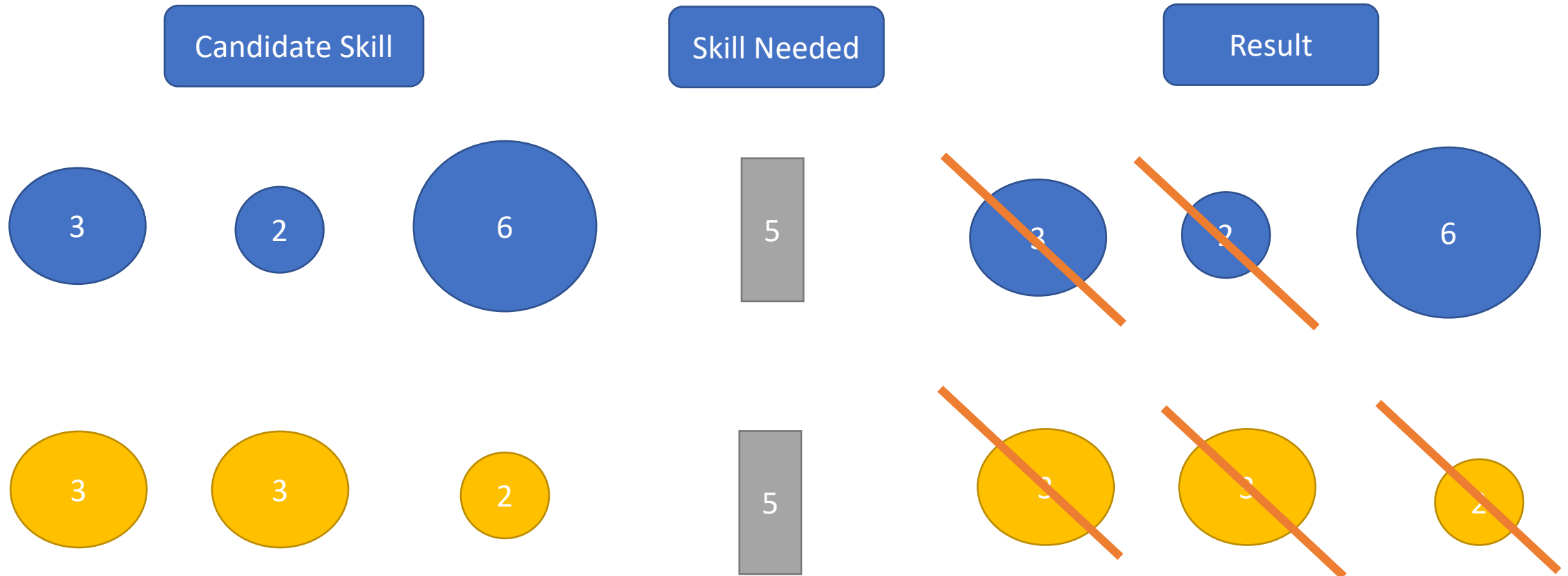# How to Distinguish Differences from Bias

- A standardized examination makes this easier
- Reminder: not all differences are indications of bias
- A valid examination will show differences that exist.
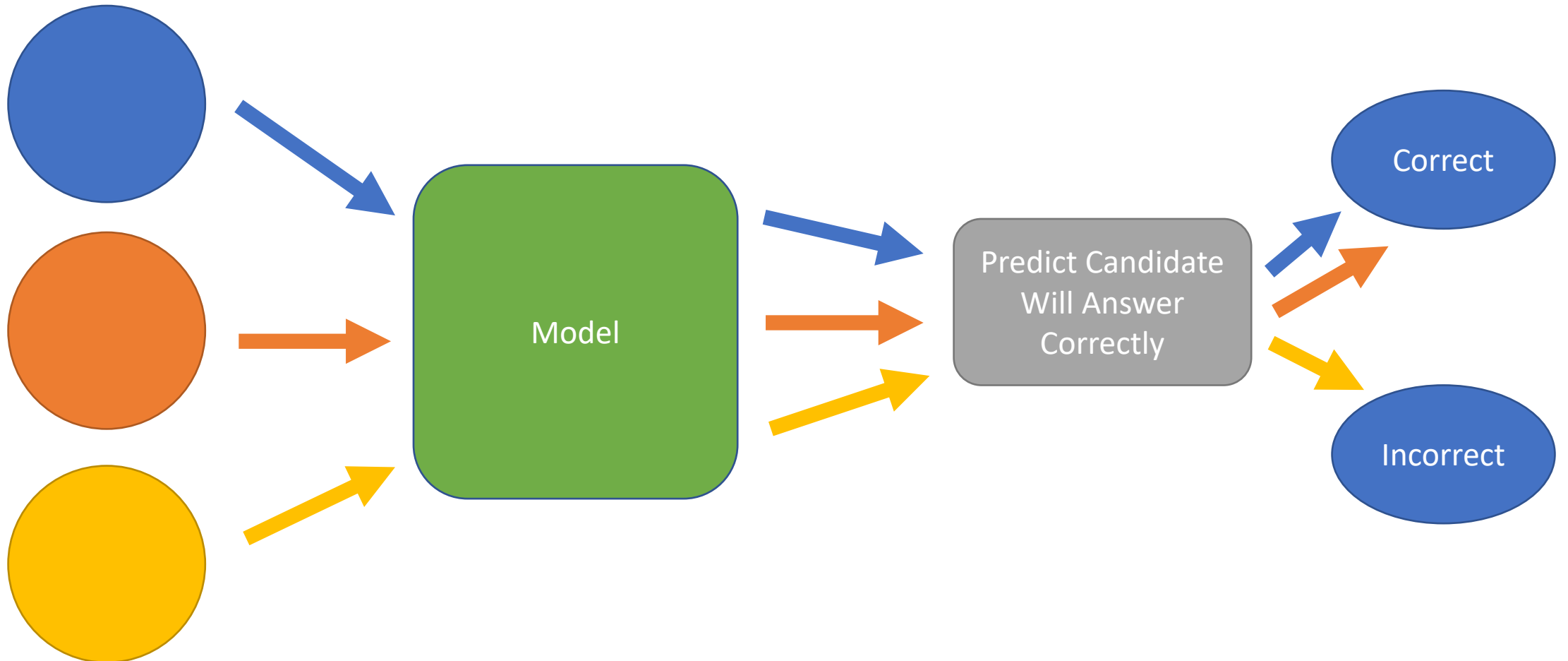
- **Answer: DIF analysis**

DIF = Differential *Item* Functioning

DIF = Differential *Item* Functioning

DIF = Differential *Item* Functioning

DIF with Data (Observed – Predicted)

Candidate	Item Error 1	Item Error 2	Item Error 3	Item Error 4
C1	0.7	-0.8	0.1	0.1
C2	-0.6	0.1	0.4	-0.9
C3	0.1	0.1	-0.2	0.1
Sum of errors^2	0.86	0.66	0.21	0.83

DTF Differential *Test* Functioning

Box full of Biased Questions and Unbiased Questions

How many biased questions do I need to remove to generate...

Results that are roughly equal to an unbiased test

Would only matter if there were more biased items than you want to remove...

Part 3: Caution when Consuming Research

Misleading Claims: Bias

Race as a predictor of outcome is not evidence of bias

Claims of pass rates not published by the examination

Claims that rely on anecdotal evidence

Claims that misrepresent the function of DIF and DTF

Misleading Claims: Examination Validity

Valid licensure examinations are currently the best method to prevent unqualified individuals from obtaining professional licenses and potentially causing public harm.

Lack of predictive validity is evidence of an invalid test

Examinations lack construct validity

Examinations lack content validity

Examinations lack construct clarity

Questions