

Testing Bias

- I. Ability tests
- II. Testing bias definitions
- III. Testing bias consequences
- IV. Detecting bias

I. Ability tests

Ability, sometimes called achievement, aptitude, or cognitive tests are standardized tests with correct/incorrect answers or timed metrics

Used for job recruitment and skills assessment, school admissions or placement, neurologic evaluations

Commonly have numeric and visual-spatial, verbal components

May include reasoning, memory, perceptual tasks, problem solving

Can be group or individual administered via paper and pencil, computer, or individual interview/examination

I. Ability tests

Ability tests have their root in earlier research on intelligence (Cattell; Binet & Simon; Terman; Otis)

¹

Binet & Simon's tests

name as many words as possible in a given timeframe, naming the days of week, unscrambling sentences, executing three verbal commands sequentially, and detecting absurdities.

Otis group intelligence for military

oral directions, arithmetical reasoning, practical judgment, synonym/antonym, disarranged sentences, number series completion, analogies, and information

¹ A good brief summary of history of intelligence can be found in Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.

I. Ability tests

Intelligence more recently considered by some to include emotional intelligence, creativity, social, and moral dimensions

II. Testing Bias Definitions

Testing bias refers to threats to the validity of a test, which means that

Scores are systematically affected (higher or lower scores), not randomly affected

This implies impact on the mean score rather than variance of scores

Usually concerning groups and usually implying unfair testing results for disadvantaged or protected groups

Implies that scores show group differences even though values for the hypothetical construct should be the same across groups

II. Testing Bias Definitions

The term *bias* is used by psychometricians in a scientific sense to describe inequities in test scores or particular items that can be evaluated empirically

Just because groups differ on average, even consistently across studies or over time, does not mean that the test is biased

II. Testing Bias Definitions

Bias is not necessarily a quality of a test that was intentional or obvious, and the test developer may not have created the test with any harmful intentions

Consequences of bias are nonetheless an important concern because they involve legal or social injustice

II. Testing Bias Definitions

- Group differences on a test do not necessarily reflect test bias
- Demonstrating bias requires demonstrating that the observed scores do not reflect true scores for one group compared with another group
- A test may be unbiased but may be used unfairly

II. Testing Bias Definitions

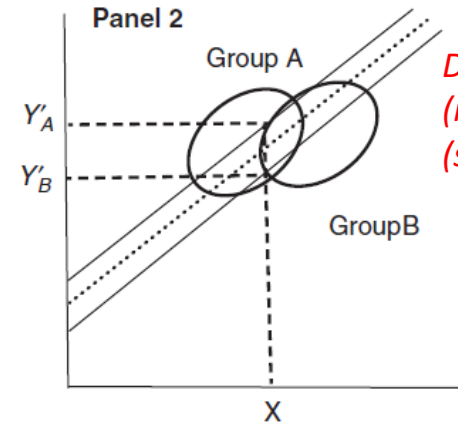
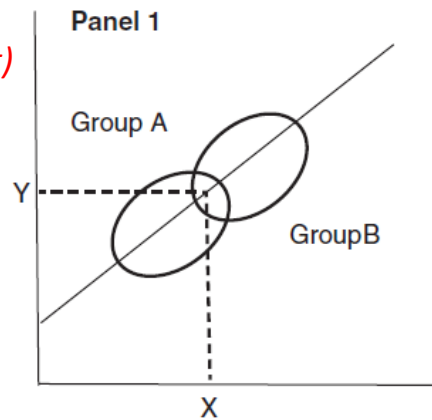
Construct bias – same average true score but observed scores differ across groups

Predictive bias – scores predict an outcome better for one group than another

e.g., career aptitude test predicts later income for boy but not girls

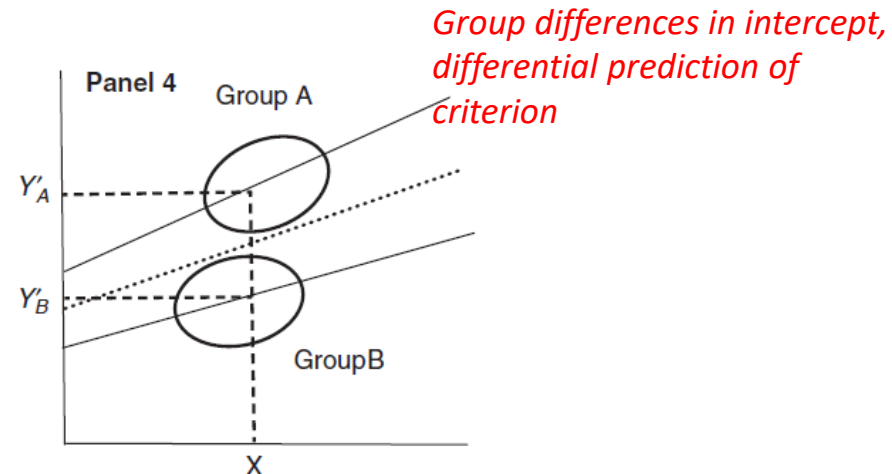
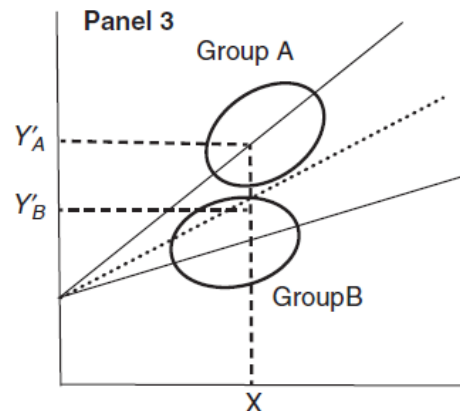
II. Testing Bias Definitions

No overall criterion (intercept) group differences or criterion prediction (slope), even though Group B has higher values on test and criterion than Group A



Difference in criterion (intercept), same prediction (slope)

No group differences in intercept, differential prediction of criterion



Group differences in intercept, differential prediction of criterion

p. 481. Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.

FIGURE 16.1. Scenarios for tests of differential prediction.

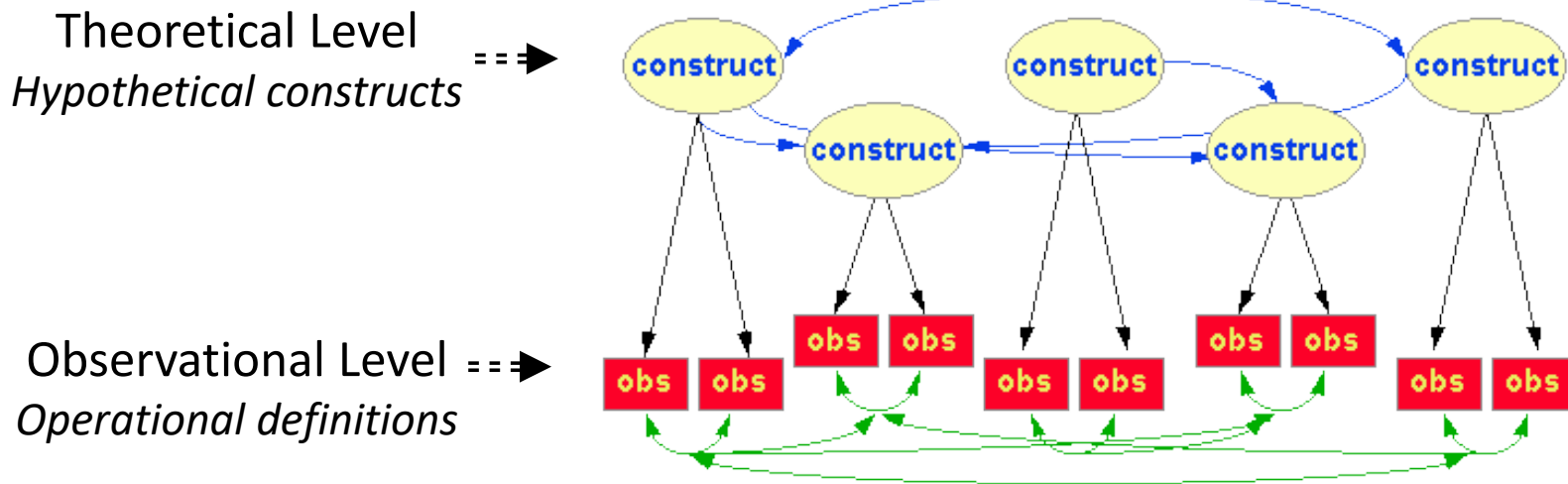
II. Testing Bias Definitions

Example

SAT over predicts college GPA for Black, Hispanic, and Native American students, and underpredicts for women compared with men (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Mattern, Patterson, Shaw, Kobrin, and Barbuti, 2008)

II. Testing Bias Definitions

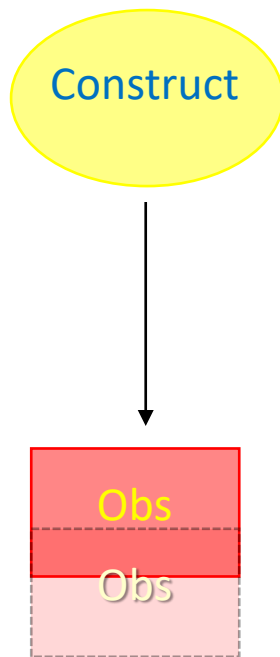
Recall nomological net



Source: William Trochim, <https://www.socialresearchmethods.net/kb/nomonet.php>

II. Testing Bias Definitions

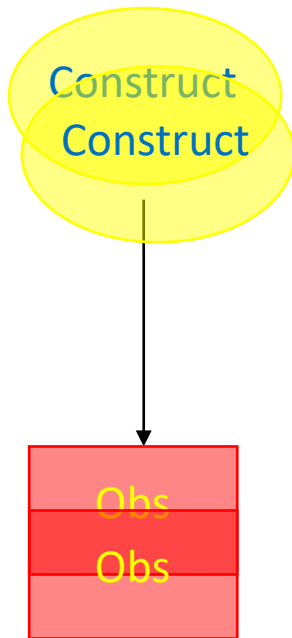
Test bias—groups differ on measure but not really on construct



Source: William Trochim, <https://www.socialresearchmethods.net/kb/nomonet.php>

II. Testing Bias Definitions

Group differences on construct



Source: William Trochim, <https://www.socialresearchmethods.net/kb/nomonet.php>

II. Testing Bias Definitions

Item bias refers to particular questions or test items that differ systematically across groups

Over and above any differences that may exist in the overall ability score

A variety of ways to investigate item bias that we will discuss later

II. Testing Bias Definitions

Biased questions may often be unintentional and not obvious

Think about words or phrases like:

On deck

(baseball term could be more familiar to boys)

Regatta

(likely more familiar to higher income students)

Cricket

(a sport in England or South Asia, but an insect in the US)

What color is a lemon?

(in Mexico limón is a lime)

III. Testing Bias Consequences

IQ tests have long legal history concerning racial biases, since the first tests that were developed (Binet & Simon, 1916)

1978 Supreme court case *Regents of the University of California v. Bakke* on use of race in medical school admissions sparked concern about testing biases in admissions tests

2013 boycott of Seattle's use of the Northwest Evaluation Association's Measures of Academic Progress (MAP) test because of concerns the test is biased against non-English speaking students

III. Testing Bias Consequences

Is racial bias in testing widespread?

Extensive reviews of evidence have come to somewhat different conclusions on either side (Brown, Reynolds, & Whitaker, 1999; Valencia & Suzuki, 2000)

We can say that standardized testing agencies have devoted considerable effort to try to eliminate bias, and that

Better tools for detecting bias exist now than at any point in the past

III. Testing Bias Consequences

One problem is that any small bias may have very serious consequences for individuals and have large cumulative societal effects

III. Testing Bias Consequences

Small hypothetical gender bias in math SAT scores

