Reducing Bias in the Admissions Process

Brent Bridgeman
Educational Testing Service

Overview

- What is test bias?

- Writing fair tests
  - Expert fairness review
  - Statistical question review
    - Differential item functioning (DIF)
  - Statistical test review
    - Differential prediction

- Can fair tests contribute to unfair selection?

Bias in Test Scores
Score Differences NOT Proof of Bias

- Mean height of men does not equal mean height of women. Does **NOT** prove tape measures are biased.

- Mean blood pressure of Black men exceeds mean blood pressure of White men. Does **NOT** prove blood pressure monitors are biased against Black men.

- Mean GRE quantitative scores of White men exceed mean scores of Hispanic men. Does **NOT** prove GRE quantitative test is biased.

Real Group Differences

- Education is **NOT** yet equal for all groups in the USA.

- Family & community resources related to education are **NOT** yet equal for all groups in the USA.

- These differences lead to real differences in educational outcomes among groups in the USA.

- Nevertheless, test bias can exist and we need to be on guard.

Some Ways ETS Evaluates Test Fairness

- Fairness Review

  – For example, why would fairness review reject:

  – "The pioneers and their wives…"

  – "Read the graph and determine the percentage increase in cancer deaths from smoking from 1990 to 2005…"

The Same Question May be Biased in One Context and Fair in Another

How many meters in 18.2 kilometers?

- Not fair as measure of multiplication skill

- Fair as measure of ability to convert within the metric system

More Ways ETS Evaluates Test Fairness

- Prediction-related bias—Do students from some groups do better than predicted from their test scores while students from other groups do worse?

- Differential Item Functioning (DIF)--Do students from majority and minority groups who have identical total scores differ in success on specific items?

- Understanding DIF by talking to students

Why might this item be harder for Hispanic students?

2 To win a game, Tamika must spin an even number on a spinner identical to the one shown below.
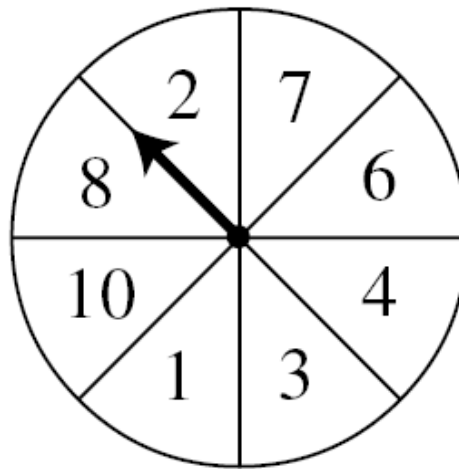
Are Tamika's chances of spinning an even number certain, likely, unlikely, or impossible?

A. certain

B. likely

C. unlikely

D. impossible

ITEM 2:   To win a game, Tamika must spin an even number on a spinner identical to the one shown below

Child 2

To win the game Tamika needs

to have the number on the spinner

identical to the number one shown below,

equal to 1

Can unbiased tests produced biased selection?
Unbiased Selection of a Basketball Team

- Base selection on height as measured with a tape measure
  - Tape measures are unbiased
  - Most basketball players are well above the population average in height (i.e., there is some validity to height as a selection variable)
  - Do top-down selection

- Conclusion: Selection of team is unbiased and valid

Who Is and Is Not on Team?

Sultan Kosen 8'3"

Nate Robinson 5'9"

Test Bias, Bias in Other Indicators, and Evaluation Process Bias

- **Lack of bias in *all* selection indicators does not necessarily ensure an unbiased evaluation process**

- **Bias can come from what is excluded as well as what is included**

Bias in Granting AP Credit?

Correlation of AP History Tests with College History Grades

Test/n	'84 American History	'85 American History
Essay	.29	.23
Multiple-Choice	.28	.24
Composite	.35	.29
n -- students	991	342
n -- colleges	18	11

Bias in Granting AP Credit?
(continued)

Sex Difference for AP Scores and Grades

AP Examination	Advanced Placement Scores			Course Grade
	Essay	Multiple-Choice	Composite	
'84 American History	.05	.48	.32	-.09
'85 American History	.11	.46	.34	-.06
European History	-.17	.43	.17	-.15
English Language	-.23	.05	-.11	-.04
Biology	.08	.26	.19	.03

Note. Entries are expressed in $\underline{d}$ units, averaged over colleges; positive values indicate higher scores for men. $\underline{N}$s are the same as in Tables 2 and 4.

Mean GRE Scores for Intended Physics Majors

	White Male (n=3381)		W F (n=837)	B M (n=99)	B F (n=27)	H M (n=246)	HF (n=66)
	Mean	SD	*d*	*d*	*d*	*d*	*d*
Verbal	159	6.1	0.0	1.1	1.5	0.7	1.1
Quant	161	5.2	0.4	1.3	1.9	0.7	1.3
AW	4.1	.70	-0.1	0.9	0.7	0.4	0.4

d = (White Male Mean-Minority Group Mean)/White Male SD

But is Analytical Writing a Good Predictor for Graduate Programs in the Physical Sciences?

GRE Quartile Comparisons: Doctorate-Seekers

General Program Area (With Its 2-Digit CIP Code)		Parks, Recreation, Leisure & Fitness Studies (31)	Philosophy & Religious Studies (38)	Physical Sciences (40)	Psychology (42)	Security & Protective Services (43)
k (Total # of Universities Contributing Data)		1	3	**7**	6	1
N (Total # of Students Contributing Data)		25	106	**508**	417	26
Probability of Grade of C+ or Lower						
GRE Verbal	Low Quartile	33%	17%	**23%**	10%	33%
	High Quartile	14%	4%	**25%**	8%	0%
	Low/High	2.33	3.75	0.90	1.34	N/A
GRE Quantitative	Low Quartile	40%	12%	**33%**	11%	33%
	High Quartile	17%	16%	**24%**	6%	0%
	Low/High	2.40	0.74	1.40	1.95	N/A
GRE Analytical Writing	Low Quartile	25%	11%	**34%**	18%	40%
	High Quartile	0%	8%	**17%**	9%	29%
	Low/High	N/A	1.38	2.04	1.94	1.40
Probability of Cumulative Graduate GPA ≥ 3.8						
GRE Verbal	High Quartile	86%	71%	**38%**	76%	50%
	Low Quartile	67%	34%	**26%**	56%	0%
	High/Low	1.29	2.06	1.48	1.36	N/A
GRE Quantitative	High Quartile	83%	34%	**40%**	75%	50%
	Low Quartile	20%	59%	**16%**	64%	33%
	High/Low	4.17	0.58	2.57	1.17	1.50
GRE Analytical Writing	High Quartile	100%	55%	**43%**	76%	14%
	Low Quartile	63%	44%	**22%**	52%	0%
	High/Low	1.60	1.23	1.96	1.45	N/A

Impact of a Strict Cutoff Requiring GRE-Q Above 160

- Number of eligible intended physics majors in each category
  - White = 1734
  - Domestic Asian = 219
  - Hispanic = 63
  - Black = 20 (including only 3 females)

Are Score Differences on Cognitive and Non-Cognitive Measures Comparable?

- PPI provides ratings from undergraduate faculty on six scales:

- Knowledge and Creativity
  - Has a broad perspective on the field
  - Is among the brightest persons I know
  - Produces novel ideas
  - Is intensely curious about the field

- Communication Skills
  - Speaks in a clear, organized and logical manner
  - Writes with precision and style
  - Speaks in a way that is interesting
  - Organizes writing well

- Teamwork
  - Supports the efforts of others
  - Behaves in an open and friendly manner
  - Works well in group settings
  - Gives criticism/feedback to others in a helpful way

PPI Scales (continued)

- Resilience
  - Accepts feedback without getting defensive
  - Works well under stress
  - Can overcome challenges and setbacks
  - Works extremely hard

- Planning and Organization
  - Sets realistic goals
  - Organizes work and time effectively
  - Meets deadlines
  - Makes plans and sticks to them

- Ethics and Integrity
  - Is among the most honest persons I know
  - Maintains high ethical standards
 - Is worthy of trust from others
  - Demonstrates sincerity

Score Differences on Cognitive Measures and Behavioral Ratings

Measure	Minority Subgroup	Cohen's d
Verbal Ability	Asian	0.20
	Black	-0.91
	Hispanic	-0.40
Quantitative Ability	Asian	0.39
	Black	-1.12
	Hispanic	-0.40
Analytical Writing Ability	Asian	0.15
	Black	-0.79
	Hispanic	-0.34
Knowledge & Creativity	Asian	0.03
	Black	-0.14
	Hispanic	-0.07
Communication Skills	Asian	0.02
	Black	-0.08
	Hispanic	-0.06
Teamwork	Asian	0.18
	Black	0.07
	Hispanic	0.01
Resilience	Asian	0.13
	Black	0.03
	Hispanic	0.08
Planning & Organization	Asian	0.10
	Black	-0.10
	Hispanic	-0.04
Ethics & Integrity	Asian	0.00
	Black	-0.04
	Hispanic	-0.02
Overall Evaluation	Asian	0.03
	Black	-0.04
	Hispanic	-0.10

Minority Admission Rates for Selective Institutions – Selection of Top 50% of Applicant Pool
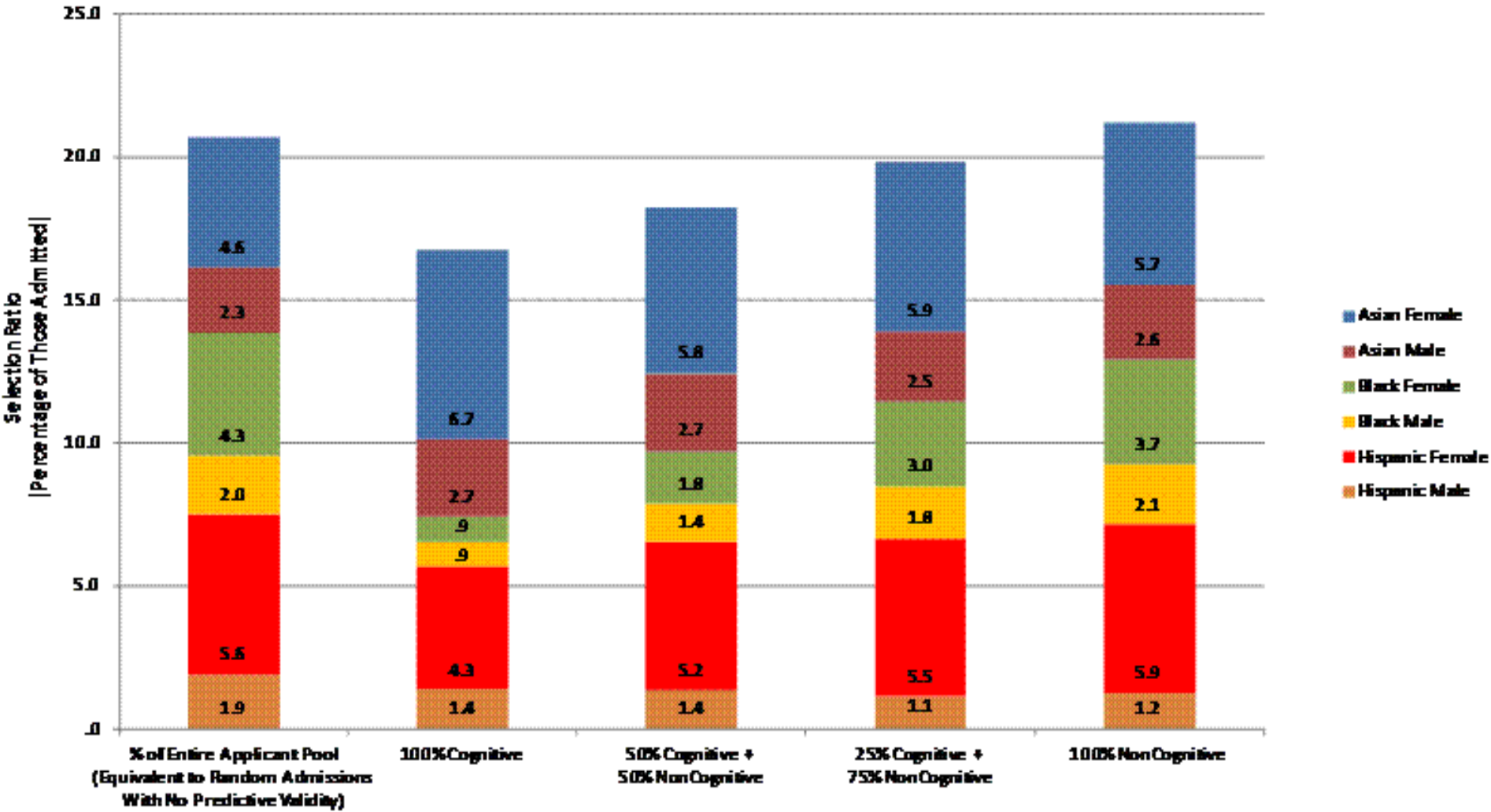
Bias in Other Non-Test Indicators

- Applications with pictures
- Interviews
- Undergraduate institution

What do These Colleges Have in Common?

- Juniata College
- Augustana College (Illinois)
- Virginia Tech
- Stevens Institute of Technology
- University of Arizona
- Brooklyn Polytechnic
- University of South Dakota

What do These Colleges Have in Common?—None is in the top 70 in US News rankings

	US News Nat'l Liberal Arts Rank
Juniata College	96
Augustana College (Illinois)	100
	US News Nat'l University Rank
Virginia Tech	72
Stevens Institute of Technology	75
University of Arizona	119
Brooklyn Polytechnic	139
University of South Dakota	199

What Else do These Colleges Have in Common?

	Graduates
Juniata College	William Phillips
Augustana College (Illinois)	Daniel Tsui
Virginia Tech	Robert Richardson
Stevens Institute of Technology	Frederick Reines
University of Arizona	Brian Schmidt
Brooklyn Polytechnic	Martin Perl
University of South Dakota	Ernest Lawrence

Conclusion

- Understanding bias in individual measures is necessary but not sufficient to evaluate bias in admissions system

- What you choose to measure or not measure is critical