Possible New Approaches to Reduce Adverse Impact

Joel P. Wiesen, Ph.D.
jwiesen@appliedpersonnelresearch.com

31th Annual IPMAAC Conference

St. Louis, MO

June 11, 2007

Adverse Impact Remains

- **Greatest single challenge to testing today**
- 1 standard deviation b-w mean score difference on many cognitive ability tests
- Much effort to reduce adverse impact
- Much anguish, modest progress
- Even well developed tests have adverse impact

Greatest Challenge to Testing

- Adverse impact results in:
  - Pressure to abandon testing
  - Pressure to lower standards
  - Legal exposure

One S.D. Mean Score Difference

- 1 S.D. difference in mean test score is huge
- Even .25 S.D. difference drives ranking

Reasons for Adverse Impact

- Tests may be the thermometer
- Many possible reasons/sources
  - educational
  - financial
  - medical
  - See previous IPMAAC presentation (web site)
- Tests may be a source of adverse impact

If Tests Are Fair

- Why strive to reduce adverse impact?
  - Tests are fair
  - No slope or intercept bias
  - No bias in prediction of job performance

Criteria May Be Biased

- Consider pay as one universal criterion
  - short people paid less than tall
  - attractive people (male and female) paid more
  - women paid less than men
- Does pay reflect job performance?
- If criteria are biased, then tests may be too

What Can We Do?

- Maintain the status quo
- Try new test item types and test areas
  - Difficult to do
  - Will meet with resistance
  - May increase validity
- Try new ways to use test data
- Strive to reduce adverse impact

One New Test Item Type

- Face Recognition Questions
  - Used for years in police officer exams
  - Room for improvement

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Content Analysis/Critique

- Measures:
  - Perceptual speed
    - Adam's apple missing in choice B
    - Lips different in choice C
    - Ear different in D
  - Analytical ability or attention to detail
    - Look for small changes in lips, eyes, chin, etc.
    - May become a one item test

Content Analysis/Critique

- Unrealistic task
  - Line drawings
  - Cloned (almost identical) faces
  - Same expression for all choices
  - Black and white images
  - Choose one face from a few choices
  - No long-term memory involved

Content Analysis/Critique

- Confounded with reading ability
- Consider these words in the question:
  - represents the face
  - alleged
  - appearance
  - assume no surgery

Face Recognition Origins

- Police line-up artists
  - Choose basic face shape
  - Change hairdo, jewelry, hats
- Use convenient, available technology
- Mug books
- Wanted poster approach
  - But missing the side view!

Possible Improvements

- Use photos rather than line drawings
- Choose a face from a large group of faces
- Use different perspectives/expressions
- Require aging of person
- Require long-term memory

Two Possible New Face Items

- Use many faces
- Vary perspectives/expressions

Question 1

- Study the mug shots on page one
- Identify corresponding people on page two

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Question 2

- Study the photos on page one
- Identify corresponding people on page two

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

Test Item Content Omitted

- Test item images omitted from the public version, for the sake of security.
- Professionals and jurisdictions may contact the author for copies.

New Face Items Are Feasible

- Challenges:
  - High quality photos
  - Printing photos (b&w vs. color)
  - Measure memory not comparison
- Potential to measure important ability
  - Currently untapped
  - Face memory

Research Needed

- Content validity
- Criterion-related validity
- Relationship to memory for names
- Relationship to other cognitive abilities

Call for Research Partners

- Offer use of items in exchange for data

What Can We Do?

- Maintain the status quo
- Try new test item types and test areas
- Try new ways to use test data
- Strive to reduce adverse impact

One New Way to Use Test Data

- Greatest Strength Model (GSM)
  - Use a battery of tests
  - Screen on all tests
  - Rank on highest test

Team Orientation

- Team members not all equal
- Some can rebound (e.g., Marcus Camby)
- Some can score, but have some weaknesses
- Build on strengths
- Compensate for weaknesses

Greatest Strength Model

- Step 1. Give several tests
- Step 2. Put tests on common metric
- Step 3. Determine highest score
- Step 4. Fail candidates with any low score
- Step 5. Rank candidates based on their highest scores

Firefighter Example

- Written (M/C) test of cognitive ability
- Structured oral interview
- Physical performance test (PPT)

Greatest Strength Model: Firefighter Example

Person	Written	Oral	PPT	Highest Grade
A	80	90	95	95
B	90	70	80	90
C	75	65	80	80

Evaluating the GSM

- Adverse Impact
- Validity

Simulation Study

- Create imaginary applicants
- Create test/job data with known correlations
- Evaluate data two ways:
  - Greatest Strength Model (GSM)
  - Conventional uses of test data

Simulation Study Methodology

- Specify intercorrelations
- Generate data with these intercorrelations
- Create gender and EEO groups
- Create mean score differences
- Evaluate adverse impact in appointments
- Evaluate validity

Firefighter Example

- Written (M/C) test of cognitive ability
- Structured oral interview
- Physical performance test (PPT)

Specify Intercorrelations

	Oral	PPT	Gender	EEO Gp.	Job Perf.
M/C Cog	.2	0	0	0	.35
Oral		0	0	0	.35
PPT			0	0	.35
Gender				0	0
EEO Gp.					0

Key: PPT stands for Physical Performance Test

Create Mean Score Differences

- Gender: 1.25 s.d. on PPT
- EEO Group: 1 s.d. on written cognitive test
- EEO Group: .5 s.d. on job performance

Adverse Impact in Appointments

- Will vary by selection ratio
- Lower selection ratios yield higher impact
- Assume we hire top 20% of applicants
  - Realistic for some jurisdictions

Adverse Impact in Appointments

M/C Cog Only					
EEO	Gender				
.29	.84				

Note: Based on 50,000 cases, hiring top 20%.

Adverse Impact in Appointments

M/C Cog Only		Composite			
EEO	Gender	EEO	Gender		
.29	.84	.49	.39		

Note: Based on 50,000 cases, hiring top 20%

Adverse Impact in Appointments

M/C Cog Only		Composite		GSM	
EEO	Gender	EEO	Gender	EEO	Gender
.29	.84	.49	.39	.65	.58

Note: Based on 50,000 cases, hiring top 20%.

Key: GSM stands for Greatest Strength Model.

Evaluate Adverse Impact

- Much lower adverse impact with GSM
- Potential for even greater reduction
  - Use more tests (e.g., work-style)

Evaluate Validity

- We have job performance data!
- We have GSM grade
- We can calculate a composite score based on M/C cognitive, oral, and PPT
- Can compute criterion-related validity

Evaluate Validity

M/C Cog Only		
.41		

Note: Based on 50,000 cases, hiring top 20%

Evaluate Validity

M/C Cog Only		GSM
.41		.45

Note: Based on 50,000 cases, hiring top 20%

Key: GSM stands for Greatest Strength Model.

Evaluate Validity

M/C Cog Only	Composite	GSM
.41	.56	.45

Note: Based on 50,000 cases, hiring top 20%

Key: GSM stands for Greatest Strength Model.

Evaluate Validity

- Higher validity for GSM than M/C cognitive
- M/C cognitive was the standard for generations

Greatest Challenge to Testing

- Adverse impact results in:
  - Pressure to abandon testing
  - Pressure to lower standards
  - Legal exposure
- So reduce adverse impact!
  - Innovative approaches show great promise

What Can We Do?

- Try new test item types and test areas
- Try new ways to use test data
- Strive to reduce adverse impact

Summary

- New item types:
  - Feasible
  - May increase validity
- New ways of using data:
  - Feasible
  - Will reduce adverse impact
  - Will maintain validity

Final Thoughts

- Our field is still young
  - Continue to look for new approaches
- Call for research partners
  - Use of items in exchange for data

Copies of this presentation are available at:
http://appliedpersonnelresearch.com/papers