Thinking Outside the Box in Merit Selection

Joel P. Wiesen, Ph.D.
wiesen@appliedpersonnelresearch.com
MAPAC Fall Conference
Princeton, NJ
11/16/06
Two Most Vexing Problems

• Adverse Impact
• Ceiling on Validity
Adverse Impact

- Seen with cognitive ability tests
- DOJ sometimes favors random selection
  - Perhaps with a low cut point
- Search for alternative selection procedures has led to innovations in personnel selection
Ceiling on Validity

• Rarely observe validity over .50
New Ideas

• New ways to use test scores
  - Reduce adverse impact
  - Maintain validity

• New selection tools/approaches

• Based on 2006 IPMAAC presentation
New Ways To Use Test Scores

• Propose: Occupational Diversity Models
  - A family of models
  - Depart from weighted component model
• Greatest Strength Model (SGM)
• Many variations of ODM
  - Greatest Two Strengths Model
  - Drop Lowest Score Model
Why Look for New Models?

• Pesky facts about adverse impact
  – Expectation of a sum = Sum of the expectations
    • Mean of sum will equal the sum of the means
    • Mean difference between groups will be no less than mean difference for the component with the greatest adverse impact
  • Adverse impact is additive (in terms of means)
Occupational Diversity Models

• Proposed new ways to use test data
• Employees contribute based on strengths
• Cookie-cutter model may be wrong for some jobs
• Team orientation
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

<table>
<thead>
<tr>
<th>Person</th>
<th>Written</th>
<th>Oral</th>
<th>PPT</th>
<th>Highest Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>80</td>
<td>90</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>70</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>C</td>
<td>75</td>
<td>65</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>
Evaluating the GSM

• Adverse Impact
• Validity
Approaches to Evaluation

• Data from the real world
  – Now seeking such data

• Simulation study
  – Easier to do
  – Faster and more comprehensive evaluation
Simulation Study

• Create imaginary applicants
• Create test/job data with known correlations
• Evaluate data two ways:
  - GSM
  - Conventional approaches
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
Specify Intercorrelations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M/C Cog</td>
<td>.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.35</td>
</tr>
<tr>
<td>Oral</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.35</td>
</tr>
<tr>
<td>PPT</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>.35</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EEO Gp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Create Mean Score Differences

• Gender: 1.25 s.d. on PPT
• EEO Group: 1 s.d. on written cognitive test
Adverse Impact in Appointments

- Will vary by selection ratio
- Lower selection ratios yield higher impact
- Assume we hire top 3% of applicants
  - Extreme example
  - Realistic for Massachusetts firefighters
Adverse Impact in Appointments

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
<th>EEO</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.08</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.

GSM stands for Greatest Strength Model.
Adverse Impact in Appointments

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEO</td>
<td>Gender</td>
</tr>
<tr>
<td>.08</td>
<td>1.04</td>
</tr>
<tr>
<td>.30</td>
<td>.21</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.

GSM stands for Greatest Strength Model.
Adverse Impact in Appointments

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
<th>Composite</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEO</td>
<td>Gender</td>
<td>EEO</td>
</tr>
<tr>
<td>0.08</td>
<td>1.04</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.

GSM stands for Greatest Strength Model.
## Adverse Impact in Appointments

<table>
<thead>
<tr>
<th>Sample of 1,000</th>
<th>M/C Cognitive Only</th>
<th>Composite</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EEO</td>
<td>Gender</td>
<td>EEO</td>
</tr>
<tr>
<td>1</td>
<td>.03</td>
<td>1.00</td>
<td>.15</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.88</td>
<td>.43</td>
</tr>
<tr>
<td>3</td>
<td>.07</td>
<td>.76</td>
<td>.11</td>
</tr>
<tr>
<td>4</td>
<td>.03</td>
<td>1.14</td>
<td>.25</td>
</tr>
<tr>
<td>5</td>
<td>.11</td>
<td>1.50</td>
<td>.25</td>
</tr>
<tr>
<td>6</td>
<td>.00</td>
<td>.88</td>
<td>.50</td>
</tr>
<tr>
<td>7</td>
<td>.03</td>
<td>1.31</td>
<td>.15</td>
</tr>
<tr>
<td>8</td>
<td>.03</td>
<td>1.00</td>
<td>.50</td>
</tr>
<tr>
<td>9</td>
<td>.15</td>
<td>.76</td>
<td>.30</td>
</tr>
<tr>
<td>10</td>
<td>.15</td>
<td>.76</td>
<td>.30</td>
</tr>
<tr>
<td>Average</td>
<td>.07</td>
<td>1.00</td>
<td>.29</td>
</tr>
</tbody>
</table>
Evaluate Adverse Impact

• Much lower adverse impact with GSM
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

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>.31</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.
Evaluate Validity

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>.31</td>
<td>.40</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.

GSM stands for Greatest Strength Model.
Evaluate Validity

<table>
<thead>
<tr>
<th>M/C Cog Only</th>
<th>Composite</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>.31</td>
<td>.52</td>
<td>.40</td>
</tr>
</tbody>
</table>

Note: Based on 50,000 cases.

Key: EEO stands for EEO group.

GSM stands for Greatest Strength Model.
## Evaluate Validity

<table>
<thead>
<tr>
<th>Sample of 1,000</th>
<th>M/C Cog. Only</th>
<th>Composite</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.29</td>
<td>.49</td>
<td>.41</td>
</tr>
<tr>
<td>2</td>
<td>.33</td>
<td>.52</td>
<td>.42</td>
</tr>
<tr>
<td>3</td>
<td>.30</td>
<td>.55</td>
<td>.43</td>
</tr>
<tr>
<td>4</td>
<td>.25</td>
<td>.46</td>
<td>.36</td>
</tr>
<tr>
<td>5</td>
<td>.34</td>
<td>.53</td>
<td>.43</td>
</tr>
<tr>
<td>6</td>
<td>.32</td>
<td>.55</td>
<td>.40</td>
</tr>
<tr>
<td>Average</td>
<td>.30</td>
<td>.51</td>
<td>.39</td>
</tr>
</tbody>
</table>
Evaluate Validity

• Higher validity for GSM than M/C cognitive
• M/C cognitive was the standard for generations
Two Most Vexing Problems

• Adverse Impact
• Ceiling on Validity
Ceiling on Validity

- Consider other models of job performance
- New ideas on tests and their uses
- “New” KSAPs may have unexpected relationships with criterion
Models of Job Performance

• Compensatory model
• Wiesen’s Occupational Diversity Models
  - Greatest Strength Model
  - Drop Lowest Score Model
  - Many other possible models
• Parse abilities more finely and look for non-linear solutions to regression equations
New Ideas on Tests

• New ways to use test scores
  - Reduce adverse impact
  - Maintain validity

• New selection tools/approaches  (Wiesen, 2004)
New Selection Tools/Approaches

• Alternate ways to pass the first hurdle
• More use of life/work experience
• Other types of tests
• Consider stability of personality traits
  - 75% of variation in weekly job performance is within person rather than between person. (Stewart and Nandkeolyar, 2006)
More Ways to Pass First Hurdle

• Written cognitive ability test
• High school rank
• Score on statewide HS graduation test
• College degree
• Honorable discharge from military
• Allow retaking test
• Several week course
Life and Work Experience

- Volunteer experience as Firefighter
- Paid experience as Firefighter
- Recommendations from teachers
Other Types of Tests

• Several week course on fire subjects
• Face recognition tests (esp. for police)
• Short term memory test
• Peripheral vision test
• Spatial orientation (esp. for firefighter)
• Balance
• Oral comprehension of various dialects
• Fine motor coordination (e.g., paramedics)
Other Types of Tests

- Mackworth Clock Test
  - Attentional capacity (e.g., Hollenbeck et al. 1995)
- Affect intensity (e.g., Larson, 1987)
- Education & experience evaluations
  - Citizenship behaviors
  - Altruistic behaviors
  - Ability to deal with interruptions
Stability of Personality Traits

• Cognitive ability is stable

• Within-person variability of personality
  - Sociability may vary from day to day
  - Responsibility may wax and wane
  - e.g., Beal et al. (2005), Fleeson et al. (2002)

• Can our current models handle this within person variability?
Summary

- New ways to combine test scores
  - Greatest Strength Model (GSM)
  - Wiesen Occupational Diversity Models
- Reduce adverse impact
- Maintain Validity
- New measurement tools and approaches
Final Thoughts

• Field is still young and developing
• Call for collaboration in simulations
  - Students
  - Researchers
  - Practitioners
• Call for real life applications
  - Police Officer
  - Firefighter

Copies of this presentation are available at:
http://appliedpersonnelresearch.com/pubs.html
References


References


References
