In-Basket Criterion-Related Validity: A Meta-Analysis

Deborah L. Whetzel, Ph.D.

International Personnel Assessment Council
Capitol Ideas for Assessment

July 17-20, 2011
Agenda

- Introduction
- How reliable are in-baskets?
- How accurate are in-baskets in predicting performance?
  - Meta-analysis of validity of in-baskets
- What attributes of in-baskets affect validity?
- What attributes of validity studies affect validity?
- To what extent are in-baskets correlated with $g$?
- Are results influenced by publication bias?
Criterion-Related Validity Evidence

• Validity of in-baskets
  – Shippman, Prien, Katz (1990) narrative review

• Validity of work samples
  – Roth, Bobko, and McFarland (2005); $\rho = 0.33$
  – Hunter and Hunter (1984); $\rho = 0.54$

• Meta-analysis of the validity of in-baskets
  – Whetzel and Rotenberry (2010)
Validity Evidence

• Conducted a meta-analysis to assess validity of in-baskets

• Literature review
  – Computerized databases (PsycInfo)
  – Listservs (SIOP, PTC/NC, PTC/MW, Academy of Management, IPAC, I/O Practitioners network)
Validity Evidence

• Decision Rules
  – Used job and training performance and salary criteria (not starting salary or personal temperament)
  – Did not include studies that reported only an Overall Assessment Rating (OAR) across all exercises
  – Did not include studies that reported only statistically significant validities

• Number of validity coefficients for each criterion
  – Job performance \((k = 32; \ N = 3,986)\)
  – Training performance \((k = 8; \ N = 1,563)\)
  – Salary \((k = 14; \ N = 1,624)\)
Validity Evidence

• Inter-rater agreement
  – 2 independent coders
  – 190 data points; 18 “disagreements”
  – 90.5% agreement

• Meta-analysis method
  – Corrections for criterion unreliability
    • Job performance distribution (Pearlman, Schmidt, & Hunter, 1980); average = .60
    • Training performance distribution (Pearlman, Schmidt, & Hunter, 1980); average = .80
    • Salary was assumed to be perfectly reliable at 1.0
Moderators of validity

• Characteristics of the in-basket
  – Scoring (objective vs. subjective)
  – Content (job-specific vs. generic)

• Characteristics of the study
  – Design (predictive vs. concurrent)
  – Source (published vs. unpublished)
Reliability of In-Baskets

• Two methods for computing reliability
  – Inter-rater reliability (agreement across raters). This is good for methods of multiple constructs
  – Coefficient alpha (internal consistency). This is good for unidimensional measures, such as cognitive ability, or conscientiousness

• In-basket is a method that can measure any number of constructs
## Reliability of in-baskets

<table>
<thead>
<tr>
<th></th>
<th>Bare Bones Meta-Analysis</th>
<th>80% Credibility Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>k</td>
</tr>
<tr>
<td>Reliability (interrater)</td>
<td>3,159</td>
<td>28</td>
</tr>
<tr>
<td>Reliability (alpha)</td>
<td>2,410</td>
<td>18</td>
</tr>
</tbody>
</table>
## Estimated Population Validity of In-Baskets

<table>
<thead>
<tr>
<th>Criterion</th>
<th>N</th>
<th>k</th>
<th>$\bar{r}$</th>
<th>SD $\bar{r}$</th>
<th>$\rho$</th>
<th>SD $\rho$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Performance</td>
<td>3,986</td>
<td>32</td>
<td>.18</td>
<td>.09</td>
<td>.36</td>
<td>.14</td>
<td>.19</td>
<td>.54</td>
</tr>
<tr>
<td>Training Performance</td>
<td>1,563</td>
<td>8</td>
<td>.17</td>
<td>.05</td>
<td>.31</td>
<td>.06</td>
<td>.22</td>
<td>.39</td>
</tr>
<tr>
<td>Salary</td>
<td>1,624</td>
<td>14</td>
<td>.14</td>
<td>.11</td>
<td>.23</td>
<td>.16</td>
<td>.03</td>
<td>.44</td>
</tr>
</tbody>
</table>
Publication Bias

- Exists when the research that appears in the published literature is systematically unrepresentative of the population of completed studies

- The funnel plot
  - X axis displays the magnitude of the effect size
  - Y axis displays precision (highly correlated with sample size)
  - Distribution will be symmetrical if sampling error is only cause of variance
Publication Bias Results

Funnel Plot of Precision by Fisher's Z

Precision (1/Std Err)

Fisher's Z

-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0

30 20 10 0
Job Performance criterion—9 studies imputed; change in validity .06—not much evidence of publication bias
## Publication Bias Effects on Observed Validity

<table>
<thead>
<tr>
<th>Criterion</th>
<th>N</th>
<th>k</th>
<th>$\frac{r}{\Delta r}$</th>
<th>Studies imputed</th>
<th>$\Delta r$</th>
<th>Adjusted observed validity</th>
<th>Higgins $I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Performance</td>
<td>3,986</td>
<td>32</td>
<td>.20</td>
<td>9</td>
<td>.06</td>
<td>.14</td>
<td>51</td>
</tr>
<tr>
<td>Training Performance</td>
<td>1,563</td>
<td>8</td>
<td>.18</td>
<td>2</td>
<td>.02</td>
<td>.16</td>
<td>46</td>
</tr>
<tr>
<td>Salary</td>
<td>1,624</td>
<td>14</td>
<td>.19</td>
<td>4</td>
<td>.09</td>
<td>.10</td>
<td>63</td>
</tr>
</tbody>
</table>
### Estimated Population Validity of In-Baskets

<table>
<thead>
<tr>
<th>In-Basket Moderator</th>
<th>Bare Bones Meta-Analysis</th>
<th>Corrected for Criterion Unreliability</th>
<th>80% Credibility Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>k</td>
<td>$\bar{r}$</td>
</tr>
<tr>
<td>Objective</td>
<td>1,125</td>
<td>12</td>
<td>.15</td>
</tr>
<tr>
<td>Subjective</td>
<td>2,230</td>
<td>16</td>
<td>.18</td>
</tr>
<tr>
<td>Job-specific</td>
<td>1,916</td>
<td>18</td>
<td>.19</td>
</tr>
<tr>
<td>Generic</td>
<td>2,070</td>
<td>14</td>
<td>.16</td>
</tr>
</tbody>
</table>
## Publication Bias Effects on Observed Validity

<table>
<thead>
<tr>
<th>In-Basket Moderator</th>
<th>N</th>
<th>k</th>
<th>( \bar{r} )</th>
<th>Studies imputed</th>
<th>( \Delta r )</th>
<th>Adjusted observed validity</th>
<th>Higgins ( I^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>1,125</td>
<td>12</td>
<td>.16</td>
<td>2</td>
<td>.04</td>
<td>.12</td>
<td>50</td>
</tr>
<tr>
<td>Subjective</td>
<td>2,230</td>
<td>16</td>
<td>.23</td>
<td>7</td>
<td>.11</td>
<td>.12</td>
<td>60</td>
</tr>
<tr>
<td>Job-specific</td>
<td>1,916</td>
<td>18</td>
<td>.22</td>
<td>6</td>
<td>.04</td>
<td>.18</td>
<td>55</td>
</tr>
<tr>
<td>Generic</td>
<td>2,070</td>
<td>14</td>
<td>.19</td>
<td>2</td>
<td>.03</td>
<td>.16</td>
<td>49</td>
</tr>
</tbody>
</table>
## Estimated Population Validity of In-Baskets

<table>
<thead>
<tr>
<th>Study Feature Moderator</th>
<th>N</th>
<th>k</th>
<th>$\bar{r}$</th>
<th>SD $\bar{r}$</th>
<th>$\rho$</th>
<th>SD $\rho$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive</td>
<td>897</td>
<td>10</td>
<td>.11</td>
<td>.12</td>
<td>.23</td>
<td>.22</td>
<td>-.06</td>
<td>.51</td>
</tr>
<tr>
<td>Concurrent</td>
<td>3,089</td>
<td>22</td>
<td>.20</td>
<td>.06</td>
<td>.41</td>
<td>.07</td>
<td>.32</td>
<td>.49</td>
</tr>
<tr>
<td>Published</td>
<td>2,547</td>
<td>18</td>
<td>.17</td>
<td>.09</td>
<td>.35</td>
<td>.14</td>
<td>.17</td>
<td>.53</td>
</tr>
<tr>
<td>Unpublished</td>
<td>1,974</td>
<td>14</td>
<td>.15</td>
<td>.10</td>
<td>.31</td>
<td>.17</td>
<td>.08</td>
<td>.53</td>
</tr>
</tbody>
</table>
## Publication Bias Effects on Observed Validity

<table>
<thead>
<tr>
<th>Study Feature Moderator</th>
<th>N</th>
<th>k</th>
<th>( \bar{r} )</th>
<th>Studies imputed</th>
<th>( \Delta r )</th>
<th>Adjusted observed validity</th>
<th>Higgins ( I^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive</td>
<td>897</td>
<td>10</td>
<td>.16</td>
<td>1</td>
<td>.04</td>
<td>.12</td>
<td>62</td>
</tr>
<tr>
<td>Concurrent</td>
<td>3,089</td>
<td>22</td>
<td>.22</td>
<td>8</td>
<td>.06</td>
<td>.16</td>
<td>40</td>
</tr>
<tr>
<td>Published</td>
<td>2,547</td>
<td>18</td>
<td>.21</td>
<td>7</td>
<td>.09</td>
<td>.12</td>
<td>56</td>
</tr>
<tr>
<td>Unpublished</td>
<td>1,974</td>
<td>14</td>
<td>.20</td>
<td>0</td>
<td>.00</td>
<td>.20</td>
<td>47</td>
</tr>
</tbody>
</table>
## Correlation with $g$

<table>
<thead>
<tr>
<th></th>
<th>Bare Bones Meta-Analysis</th>
<th>Corrected for in-basket unreliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>k</td>
<td>$\bar{r}$</td>
</tr>
<tr>
<td>Correlation with $g$</td>
<td>2,906</td>
<td>18</td>
</tr>
</tbody>
</table>
## Effect of Range Restriction

<table>
<thead>
<tr>
<th>Range Restriction</th>
<th>N</th>
<th>k</th>
<th>$\bar{r}$</th>
<th>SD $\bar{r}$</th>
<th>$\rho$</th>
<th>SD $\rho$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concurrent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = .405$</td>
<td>3,089</td>
<td>22</td>
<td>.20</td>
<td>.06</td>
<td>.55</td>
<td>.00</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>$u = .500$</td>
<td>3,089</td>
<td>22</td>
<td>.20</td>
<td>.06</td>
<td>.47</td>
<td>.03</td>
<td>.42</td>
<td>.51</td>
</tr>
<tr>
<td>$u = .595$</td>
<td>3,089</td>
<td>22</td>
<td>.20</td>
<td>.06</td>
<td>.41</td>
<td>.07</td>
<td>.32</td>
<td>.49</td>
</tr>
<tr>
<td><strong>Predictive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u = .405$</td>
<td>897</td>
<td>10</td>
<td>.11</td>
<td>.12</td>
<td>.31</td>
<td>.28</td>
<td>-.05</td>
<td>.66</td>
</tr>
<tr>
<td>$u = .500$</td>
<td>897</td>
<td>10</td>
<td>.11</td>
<td>.12</td>
<td>.26</td>
<td>.25</td>
<td>-.06</td>
<td>.58</td>
</tr>
<tr>
<td>$u = .595$</td>
<td>897</td>
<td>10</td>
<td>.11</td>
<td>.12</td>
<td>.23</td>
<td>.22</td>
<td>-.06</td>
<td>.51</td>
</tr>
</tbody>
</table>
Limitations

• Low k and low N
  – Companies may be concerned about risk of doing a criterion-related validation study
  – Results are often proprietary
  – In-baskets are often part of an assessment center and the data are often reported by dimension/competency
Conclusions

• Objective vs. Subjective—not much difference
• Job-specific vs. Concurrent—minimal difference (but in expected direction)
• Predictive vs. Concurrent—pretty large difference; lots of variance around mean validity for predictive
• Published vs. Unpublished—not much difference for population estimates, but pretty large difference for observed after pub bias (but unexpected direction)
• Correlation with $g$—likely subgroup differences
Utility of the In-Basket
Utility analysis is a method for determining the dollar value of a selection method. It answers the question, “How much money is saved or earned using a valid selection method?”

The formula for calculating utility (Brogden, 1949; Cronbach & Gleser, 1965) is:

\[ U = (T N_s r_{xy} SD_y Z_x) - C \]
**Utility Formula**

- \( U = (T N_s r_{xy} SD_y Z_x) - C \) where:
  - \( U \) = the dollar value (utility) of the selection procedure
  - \( T \) = number of years that an employee remains on the job (tenure)
  - \( N_s \) = the number of people hired each year
  - \( r_{xy} \) = the correlation between the assessment and job performance; the validity of the assessment
  - \( SD_y \) = the difference between high and low levels of job performance (Research shows 40% of salary)
  - \( Z_x \) = the score of people above the “cutoff”; ratio of the number of selected applicants to total applicants
  - \( C \) = cost of developing, validating, and administering the assessment to applicants
Utility Example: HR Manager

- $T = 10$ years (assume HR Manager tenure in an organization is about 10 years)
- $N_s = 2$ (assume the average number of HR Managers hired per year in an organization)
- $r_{xy} = .23$ (predictive validity of in-baskets)
- $SD_y = 36,000$ (assume the average salary for HR Managers is $90,000; underestimate not including benefits)
- $Z_x = .80$ (mean of 0 and SD of 1).
- $C =$ Development/Validation Study and administration costs = $10,000.$
Utility Results and Implications

• The value to an organization of using an in-basket for the first year is $122,480, assuming
  – 2 HR Managers are hired each year
  – Each one stays for 10 years
  – They make an average of ~$90,000 per year (median salary; O*NET, 2009)

• The difference between good and bad HR Managers is about 40% of their annual salary.

• While a savings of $122,480 may seem high, think of the critical hire/fire decisions an HR manager makes and the advice they provide regarding legal HR issues
References


Questions?
Thank you!