Ascertaining the Validity of Web-Based Personality Inventories

John A. Johnson
Penn State DuBois

http://www.personal.psu.edu/~j5j/Texas.ppt
Who Should Be Giving This Talk

- Chris Fraley, University of Illinois - Chicago
- How to Conduct Behavioral Research Over the Internet (Guilford, this month)
- http://www.web-research-design.net/
Equivalence of Usual & Web Formats

- Paper/pencil vs. Web format: Equivalent results?
- Can (should?) be addressed by brute empiricism
- Can compare
  - Descriptive norms (assuming samples comparable)
  - Reliabilities
  - Factor structure
  - Correlations with other variables
- Non-exhaustive survey: No differences
- More interesting: Common & special validity concerns
Common Validity Concerns

- Personality tests presented on the Web are still personality tests
- Presentation-independent threats to validity
  - Self-deception (unwitting + literally inaccurate)
  - Impression-management (deliberate + inaccurate)
  - Inconsistency/Incoherence
- Complete non-responsiveness
  - Random responding
  - Repeating the same response category
Special Validity Concerns of Web

- Reduced accountability due to anonymity
- Rushing due to ease of responding, submitting, and receiving feedback
- Reduced accountability + Rushing may cause
  - Repeat participation
  - Playful experimentation
  - Deliberate sabotage
  - Carelessness
  - Non-responsiveness (random or repeated responses)
Which Concerns Should Be Nonconcerns?

- Personality test doubters are worry-warts
- What respondents *might* do ≠ what they do
- When “misbehavior” does occur, it may not matter
  - Example 1: Providing socially desirable responses
  - Example 2: Endorsing literally untrue statements
- Unnecessary worry ⇐ failure to see that item responses are equivalent to performative speech acts from everyday social interaction
- Worthwhile worry: out-of-character presentations
Dealing With Special Concerns

- Repeat participants
  - Researcher-assigned ID/password
  - Comparing number of identical responses
Dealing With Special Concerns, 2

- Playful experimentation
  - Re-take with few changes = duplicate protocol
  - No way to detect someone pretending to be, say, Walter Cronkite or Daphne Deckers
Dealing With Special Concerns, 3

- Response repeating

- Costa & McCrae (in press?) report for sample of 983 cooperative volunteers, no person >
  - 6 consecutive “strongly disagree”
  - 9 consecutive “disagree”
  - 10 consecutive “neutral”
  - 14 consecutive “agree”
  - 9 consecutive “strongly disagree”

- Clearer picture from samples of 20,000+
<table>
<thead>
<tr>
<th>Consec</th>
<th>Zero</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21159</td>
<td>1801</td>
<td>355</td>
<td>2240</td>
<td>117</td>
<td>1653</td>
</tr>
<tr>
<td>2</td>
<td>911</td>
<td>5984</td>
<td>3275</td>
<td>7237</td>
<td>1046</td>
<td>6133</td>
</tr>
<tr>
<td>3</td>
<td>272</td>
<td>6819</td>
<td>7194</td>
<td>7138</td>
<td>4888</td>
<td>6982</td>
</tr>
<tr>
<td>4</td>
<td>136</td>
<td>4983</td>
<td>6415</td>
<td>3798</td>
<td>7591</td>
<td>4874</td>
</tr>
<tr>
<td>5</td>
<td>98</td>
<td>1735</td>
<td>3047</td>
<td>1420</td>
<td>4474</td>
<td>1556</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>740</td>
<td>1479</td>
<td>598</td>
<td>2463</td>
<td>838</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>406</td>
<td>662</td>
<td>279</td>
<td>1291</td>
<td>510</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>274</td>
<td>326</td>
<td>122</td>
<td>561</td>
<td>219</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>191</td>
<td>169</td>
<td>72</td>
<td>323</td>
<td>132</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>54</td>
<td>68</td>
<td>41</td>
<td>131</td>
<td>52</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>27</td>
<td>35</td>
<td>25</td>
<td>71</td>
<td>38</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>20</td>
<td>19</td>
<td>12</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>&gt;14</td>
<td>343</td>
<td>27</td>
<td>20</td>
<td>82</td>
<td>41</td>
<td>45</td>
</tr>
</tbody>
</table>
Dealing With Special Concerns, 4

- Deliberate sabotage, carelessness, or random responding
- Item Response Models considered, but too computationally intensive
- Two indices of coherence
  - Jackson intra-individual consistency
  - Goldberg semantic/response antonymy
Jackson (1976) Intraindividual Consistency

- Create half-scales from odd and even items for each of 30 facet scales
- Correlate odds with evens across 30 scales and correct with Spearman-Brown
- Mean intra-individual coefficients similar to Jackson’s (~.84, SD = 10)
- 96% of cases > .64
- Cutoff arbitrary, but probably want to eliminate 11 people with negative coefficients
Goldberg Antonymy

- Intercorrelate all 300 items
- Locate 30 unique pairs with highest negative r’s
  - #31 “Fear for the worst”
  - #154 “Think that all will be well”
- Correlate individuals’ responses across 30 pairs
- Reverse sign, mean consistency = .48 (SD = 20)
- Measure correlates $r = .49$ with Jackson measure
Graph of Goldberg Measure

- Std. Dev = .20
- Mean = .48
- N = 20698.00
Does Coherence Matter?

- Barely affects results of factor analysis
  - Predicted loadings from bottom quartile about .05-.08 lower than loadings from upper quartile
  - Loadings from groups split at median nearly identical
- Testing moderating effect on self-acquaintance correlations would be worthwhile
- Tentative conclusions
  - Jackson and Goldberg measures useful only for screening extremely incoherent cases
  - Coherence itself is an individual differences variable
## Personological Mapping of Coherence

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Facets</td>
<td>.70</td>
<td>-.21</td>
<td>-.06</td>
<td>-.15</td>
<td>.00</td>
</tr>
<tr>
<td>E Facets</td>
<td>-.11</td>
<td>.64</td>
<td>-.10</td>
<td>.20</td>
<td>.10</td>
</tr>
<tr>
<td>O Facets</td>
<td>.02</td>
<td>.13</td>
<td>.04</td>
<td>-.01</td>
<td>.61</td>
</tr>
<tr>
<td>A Facets</td>
<td>-.01</td>
<td>.10</td>
<td>.67</td>
<td>.02</td>
<td>.06</td>
</tr>
<tr>
<td>C Facets</td>
<td>-.23</td>
<td>-.06</td>
<td>.24</td>
<td>.64</td>
<td>-.03</td>
</tr>
<tr>
<td>Jackson</td>
<td>-.32</td>
<td>.01</td>
<td>.11</td>
<td>.10</td>
<td>.36</td>
</tr>
<tr>
<td>Goldberg</td>
<td>-.49</td>
<td>.02</td>
<td>-.04</td>
<td>.19</td>
<td>.21</td>
</tr>
</tbody>
</table>
Conclusions

- Psychologists worry too much about factors irrelevant to Web-based personality testing.
- Coherence and other validity measures useful for screening out maybe 2-3% of protocols.
- Future research should examine correlations with external criteria such as acquaintance ratings.
- An even more pressing and interesting problem is ascertaining validity of individual narrative reports used as feedback.