Filtering Rapid Responses for Low-Stakes Student Learning Outcomes Assessment: An Investigation of Two Approaches

Joseph A. Rios
Associate Research Scientist
Educational Testing Service
Context

• Increased use of student learning outcomes assessment in higher education
  – HEIghten™: critical thinking, written communication, quantitative literacy, intercultural competency and diversity, civic competency and engagement, digital information literacy

• Results important for institutions, but not for students
  – Critical to consider and control for motivation
Impact of Low Motivation on Test Results

• In terms of aggregated-score inferences, careless responding has been found to bias:
  – Treatment effects (Osborne & Blanchard, 2011)
  – Achievement gains (Wise & DeMars, 2010)
  – Teacher evaluation ratings (Wise et al., 2013)
  – Country-level comparisons (Debeer et al., 2014)

• Results from these studies may have been confounded by the procedures for:
  – Identifying careless responses
  – Purifying scores from careless responses
Rapid Guessing Filtering Procedures

Examinee-level Filtering

• Data are listwise deleted for any examinee rapid guessing on or more than a certain percent of items (e.g., 20%)

Response-level Filtering (IRT expected scores)

• Treat rapid responses as missing
• Estimate IRT item and person parameters
• Use estimated parameters to compute expected probabilities
• Sum expected probabilities within and across examinees
## Rapid Guessing Filtering Procedures

<table>
<thead>
<tr>
<th>Filtering Process</th>
<th>Examinee-Level</th>
<th>Response-Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Listwise deletes data from an examinee deemed to be “unmotivated”</td>
<td>• Treats a careless response as not administered</td>
</tr>
<tr>
<td>Advantage</td>
<td>• Can identify motivation using both response time and self-report</td>
<td>• Does not delete possible valid responses</td>
</tr>
<tr>
<td>Disadvantage</td>
<td>• Discards up to 25% of the sample data</td>
<td>• Assumes non-flagged responses are valid indicators of ability</td>
</tr>
<tr>
<td></td>
<td>• Assumes ability is unrelated to motivation</td>
<td></td>
</tr>
</tbody>
</table>
Comparison of Filtering Procedures when Careless Responding is UNRELATED to Ability

EXAMINEE-LEVEL FILTERING

Comparison between Rapid Guessing Filtering Procedures
Comparison of Filtering Procedures when Careless Responding is RELATED to Ability

EXAMINEE-LEVEL FILTERING

Comparison between Rapid Guessing Filtering Procedures
Comparison of Filtering Procedures when Careless Responding is UNRELATED to Ability

RESPONSE-LEVEL FILTERING

Proportion of Careless Responses in Total Sample

Comparison between Rapid Guessing Filtering Procedures
Research Questions

1. Is the assumption that rapid guessing is unrelated to ability (underlies examinee-level filtering) tenable in practice?

2. If not, does our perception of the impact of rapid guessing on the group mean differ when employing response-level filtering?
METHOD
Sample and Measures

• 1,422 first-year students from a large state university

• ETS Proficiency Profile
  – 108 multiple-choice items
  – Assesses reading, writing, math, and critical thinking ($\alpha = .95$)

• SAT verbal and quantitative scores
## Rapid Guessing Flagging Procedures

<table>
<thead>
<tr>
<th>Name</th>
<th>Procedure for Defining Flagging Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Seconds (3Sec)</td>
<td>&lt;3 second response</td>
</tr>
<tr>
<td>Normative Threshold (NT)</td>
<td>% of mean item response time (examined 15%, 20%, &amp; 25%)</td>
</tr>
<tr>
<td>Visual Inspection (VI)</td>
<td>Intersection of bimodal response time frequency distributions</td>
</tr>
</tbody>
</table>
Analyses

• Flag rapid guesses employing one of five procedures

• Evaluate impact of rapid guessing on mean scores when employing examinee-level filtering

• Compare performance differences by motivation groups on expected EPP and SAT scores

• Determine whether the impact of rapid guessing on mean scores is less for response-level when compared to examinee-level filtering
RESULTS
## Rapid Guessing in Sample

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>SD</th>
<th>Proportion Correct</th>
<th>Unmotivated Examinees (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SEC</td>
<td>4.61</td>
<td>9.75</td>
<td>.22</td>
<td>120 (9.08%)</td>
</tr>
<tr>
<td>NT15</td>
<td>5.64</td>
<td>10.85</td>
<td>.21</td>
<td>160 (12.10%)</td>
</tr>
<tr>
<td>NT20</td>
<td>6.81</td>
<td>11.74</td>
<td>.21</td>
<td>204 (15.43%)</td>
</tr>
<tr>
<td>NT25</td>
<td>7.82</td>
<td>12.39</td>
<td>.22</td>
<td>241 (18.23%)</td>
</tr>
<tr>
<td>VI</td>
<td>7.96</td>
<td>12.04</td>
<td>.22</td>
<td>244 (18.46%)</td>
</tr>
</tbody>
</table>

Comparison between Rapid Guessing Filtering Procedures
Impact of Rapid Responding on Mean Score
(Examinee-Level Filtering)

Comparison between Rapid Guessing Filtering Procedures

Filtered Mean (Examinee-Level Filtering)

Unfiltered Mean

Rapid Guessing Flagging Procedure

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Mean EPP Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SEC</td>
<td>30</td>
</tr>
<tr>
<td>NT15</td>
<td>31</td>
</tr>
<tr>
<td>NT20</td>
<td>32</td>
</tr>
<tr>
<td>NT25</td>
<td>33</td>
</tr>
<tr>
<td>VI</td>
<td>34</td>
</tr>
</tbody>
</table>
Discriminant Validity of Response Time Effort

\[ \text{skewness} = -4.44 \]
\[ r = .12 \]
\[ r_s = .19 \]

Comparison between Rapid Guessing Filtering Procedures
Performance Differences by Motivation Groups

Comparison between Rapid Guessing Filtering Procedures

<table>
<thead>
<tr>
<th>Rapid Guessing Flagging Procedure</th>
<th>Standardized Difference (Motivated - Unmotivated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SEC</td>
<td>1.5</td>
</tr>
<tr>
<td>NT15</td>
<td>1.46</td>
</tr>
<tr>
<td>NT20</td>
<td>1.45</td>
</tr>
<tr>
<td>NT25</td>
<td>1.45</td>
</tr>
<tr>
<td>VI</td>
<td>1.44</td>
</tr>
</tbody>
</table>

Copyright © 2015 by Educational Testing Service. All rights reserved. ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING is a trademark of ETS.
Performance Differences by Motivation Groups

Rapid Guessing Flagging Procedure

Comparison between Rapid Guessing Filtering Procedures

Standardized Difference (Motivated - Unmotivated)

Expected EPP 3SEC NT15 NT20 NT25 VI

SAT

Copyright © 2015 by Educational Testing Service. All rights reserved. ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING is a trademark of ETS. 30141
Impact of Rapid Responding on Mean Score (Response-Level Filtering)

Comparison between Rapid Guessing Filtering Procedures

- Unfiltered Mean
- Filtered Mean (Response-Level Filtering)

\[ d = 0 \quad d = .01 \quad d = .01 \quad d = .02 \quad d = .02 \]
DISCUSSION

Comparison between Rapid Guessing Filtering Procedures
Summary

• Filtered examinees were shown to possess significantly lower expected EPP and SAT scores
  – This may have resulted in artificially inflating filtered group means

• The unfiltered group mean was found to be equivalent to the response-level filtered expected group means
  – This suggests that the rapid responses in the dataset did not significantly impact the observed group mean
Implications

• If examinee-level filtering is employed, the assumption that ability is unrelated to careless responding should be evaluated beforehand.

• If this assumption cannot be evaluated or is found to be untenable, response-level filtering should be applied.
THANK YOU

Joseph A. Rios
Associate Research Scientist
Educational Testing Service

jrios@ets.org