Comments on Assessment Consortia Test Security Presentations

Steve Ferrara

Presented in J. Steedle (Organizer), Test Security for Common Core Consortia Assessments, a session in the National Conference on Student Assessment
June 21, 2016
Overview

- Frameworks to guide my comments
  - PDIR
  - Threats to security

- Comments on each paper

- Concluding comments
Framework for comprehensive test security systems

- Prevention, Detection, Investigation, Resolution (PDIR)
- Progress on Prevention and Detection
- Investigation
  - (Mostly from media) Making progress on investigations
  - May be too much conflict of interest in local investigation
- Resolution
  - Much of the evidence is not accessible
  - Evidence we do have is…
## Threats to test security

<table>
<thead>
<tr>
<th>Before Test Administration</th>
<th>During Test Administration</th>
<th>After Test Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examinees</strong></td>
<td><strong>Test Administrators</strong></td>
<td><strong>Other Testing Site Staff</strong></td>
</tr>
<tr>
<td>Acquiring test content</td>
<td>Copying or supplying answers, impersonating</td>
<td>Divulging content, failing to report violations</td>
</tr>
<tr>
<td>Divulging, teaching, inappropriate test prep</td>
<td>Providing answers, allowing inappropriate conditions, inadequate protection</td>
<td>Changing answers, tampering with answer files, failing to report</td>
</tr>
<tr>
<td><strong>Program Managers and Operations Vendors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failing to publicize expectations, failing to train</td>
<td>Failing to observe administrations and protection</td>
<td>Failing to report, failing to account for materials</td>
</tr>
</tbody>
</table>

Some threats require human vigilance for prevention, detection, and investigation.

From Ferrara, 2016, Table 1; adapted from Fremer & Ferrara, 2013, Table 2.1
Jeff Steedle on PARCC forensics

- Statistical plus non-statistical approaches
- Addresses several threats
- Web crawling to detect sharing of content
  - PARCC states follow internal breach procedures
  - Is this adequate to protect current content and discourage future breaches via rigorous resolution procedures?
- Answer changing via points gained approach
  - Logically similar to WTR answer changing methods
  - Now, an automated system to detect changes for constructed response items!
PARCC forensics (cont.)

- Plagiarism for prose constructed response via latent semantic analysis
  - Great to see PARCC exploring this approach
  - Found very few flagged pairs; LSA adequately sensitive and specific?
  - Relatively high numbers of flags for Narrative Writing Task grades 5 (37) and 7 (127)
    - Hypothesis: *Retelling from a character’s perspective*—Plausible?
      Conditional on PCR score or theta?
- If think of plagiarism within schools, numbers of pairings may be manageable
PARCC forensics (cont.)

- Aberrant response patterns
- Modified caution index (MCI)
  - Findings consistent with the literature?
- Standardized log-likelihood person fit index
  - More sensitive in simulated data
  - Specificity/false negatives? Its performance conditional on theta/theta ranges?
PARCC forensics (cont.)

- Longitudinal performance modeling
  - Detect unusual performance changes via cumulative logit regression
- Clark et al. found it detects “test misconduct” with good power and conservative false positive flags
  - Two consecutive years—Could find downward spikes as well as expected upward spikes
  - Have to determine practical significance awa statistical
- Not yet used operationally
  - Look forward to hearing results from 2015-2016 investigation of cumulative logit regression
PARCC lessons learned

- Will continue response change analyses—good
  - Considering rules for flagging score increases—good
- Plagiarism/copying on constructed responses—tough detection problem
  - Approach: States will request focus on worrisome schools ("Known unknowns")
  - What about schools that are new to cheating? ("Unknown unknowns")
- Add answer copying/plagiarism for short responses, where answer copying/dictating responses is easier
Brandt Redd on Smarter Balanced views on test security

- “Security goals should benefit students”
  - Of course
  - Test security systems also should serve our responsibilities to the public and our state and federal sponsors: data integrity (Ferrara, 2012; USDE, 2013)

- Open source secure browsers—and roadmap to a common industry solution
  - Great way to pursue our responsibilities to federal sponsors
Smarter Balanced views (cont.)

- Security and CAT
  - 20:1 ratio of available and presented items; suspend exposed items
  - Would like to see that ratio within standards x scale locations matrix—“redundant items”
- Efforts to protect security and enable assistive technology—laudable
- Test administration policies, training, monitoring, and statistical forensics—laudable
- What about human vigilance to protect and detect other security threats?
Mark Hansen on Smarter Balanced forensics for CAT

- Aberrant response patterns
- Project W from R, L, S
  - What were the conditions for the 3% significant differences? (e.g., $\text{Theta} = -2$)
  - Same question for second illustration
  - Rudner, Bracey, & Skaggs asked that question 20 years ago
- Glad to hear Mark say bigger risk may be with constructed response items; our field not doing enough here yet
In closing

• Glad to see this work going on in the consortia
  ◦ Next year: Why not invite WIDA, ELPA21, NCSC, and DLM?

• Would like to see big effort on forensics for security threats for constructed response items

• I’ll summarize “empty cells” in threats x detection methods matrix
  ◦ Source for considering next method studies
Thanks!

Steve Ferrara
sferrara1951@gmail.com
References


## PARCC responses to test security threats

<table>
<thead>
<tr>
<th></th>
<th>Before Test Administration</th>
<th>During Test Administration</th>
<th>After Test Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examinees</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acquiring test content</td>
<td></td>
<td>Copying or supplying answers, impersonating</td>
<td>Divulging content, failing to report violations</td>
</tr>
<tr>
<td><strong>Test Administrators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divulging, teaching,</td>
<td></td>
<td>Providing answers, allowing</td>
<td>Changing answers, tampering with answer files, failing to report</td>
</tr>
<tr>
<td>inappropriate test prep</td>
<td></td>
<td>inappropriate conditions, inadequate protection</td>
<td></td>
</tr>
<tr>
<td><strong>Other Testing Site Staff</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failing to train, failing to monitor security</td>
<td></td>
<td>Failing to monitor administrations</td>
<td>Failing to report, tampering with answers and files</td>
</tr>
<tr>
<td><strong>Program Managers and Operations Vendors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failing to publicize expectations, failing to observe performances</td>
<td></td>
<td>Failing to observe administrations and protection</td>
<td>Failing to report, failing to account for materials</td>
</tr>
</tbody>
</table>

Test Security for Common Core Assessment Consortia