Alternative Proctoring Methods: Considerations and Future Challenges

Presenters: John Weiner, PSI Services LLC
            Greg Pope, Yardstick
Session Overview

• Introduction
• Example Applications
• Considerations & Research
• Future Challenges
• Open Discussion
Introduction

Advances in technology + economic forces have spurred new assessment models
- Design, Delivery, Use
Alternative Delivery Models

- **Low Stakes**
  - Unproctored Internet Testing (UIT)
- **High Stakes**
  - Remote Proctored Testing
  - Proctored Testing Center
Early Adopters of UIT

• Distance Learning

• Talent Assessment
How is UIT being Used?

- Increasingly common in employment arena
  - Typically first stage of a multi-stage selection process
  - Most often used for “non-cognitive” measures (e.g., self-report descriptions of experience or personality) that do not have fact-based scoring protocols
  - Verification testing often used as a “fail-safe”

- Primarily used for practice tests in credentialing arena
Who is using UIT?

TA Examples
- Proctor & Gamble
- CSX
- IBM
- Bank of America
- Frito Lay
- Wachovia
- Marriott
- Wells Fargo
- Dept of Defense research for Career Testing

Education Examples
- Association of Accounting Technicians
- AQA - General Cert of Secondary Ed.
- Scottish Qualifications Authority
- EAL - engineering, manufac. Exams

Corporate Training Examples
- Booster Juice
- Moen
- Enterprise
- Snap On Tools
- Hill's Pet
- Tumi
- NEC
Example Applications
Example: Certification/Licensure - Ethics and Jurisprudence Exams

- Ethics and Jurisprudence exams are being administered more often online by regulators.
- The stakes level is not usually “certification” but in some contexts candidates have to pass the JP exam in order to maintain their professional standing.
Example: JP exams model 1

1. Candidate purchases JP exam
2. Has a certain time period (24 hours) to take exam at home/work in UIT
3. Testing interface has added security features to limit cheating:
   - Simple clean interface
   - Candidate can only see one item at a time and can’t go back/forward
   - Printing is disabled
   - Selecting for copy/paste is disabled
Example: JP exams model 2

1. Candidate is selected by regulator to take JP exam (fees included in annual membership)

2. Candidate can take exam at any time within a certain period of months

3. Candidate agreement is in place but exam is “open book” with resources provided

4. Candidate can take the exam multiple times until they pass, obtaining feedback each time on where they are strong/weak
Remote Proctoring: Existing techniques

- Recording of candidate assessment session:
  - Audio and video recordings can be collected for candidates and reviewed (e.g., spot checked) post-exam administration to ensure no incidents occurred
  - Software Secure provides this model

- Real time statistical collusion detection:
  - Many techniques exist documented in the academic literature to identify candidates copying answers from one another:
    - Wollack Omega
    - Frary $G_2$
    - Angoff B
    - Bellezza and Bellezza ESA
  - Apply these techniques as candidates are taking the exams with ‘suspend’ rules
Considerations
Threats to Measurement Quality

- Test scores may be affected by factors that are unrelated to the purpose of the test
  - *May impact reliability and validity*

- Concerns for UIT and Remote Proctored Delivery:
  1. Cheating
  2. *Intellectual property loss*
  3. Content exposure
  4. *Non-standard environment – technology*
  5. *Differential access (digital divide)*
  6. Candidate acceptance
  7. Corporate brand, reputation
### What’s at Stake?

<table>
<thead>
<tr>
<th>Stakes</th>
<th>Context</th>
<th>IP Costs</th>
<th>Cheating Costs</th>
<th>Safety/Legal Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Training &amp; Development; Low impact decisions</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>Moderate impact decisions</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>High</td>
<td>Public safety; High impact decisions</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
Risk Management

Low Stakes
- Open Access
- Remote

High Stakes
- Controlled Access
- Managed Access

Supervised Access
- Onsite
## What do Test Standards say?

<table>
<thead>
<tr>
<th>GUIDELINES</th>
<th>ISSUES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standards for Educational and Psychological Testing</strong> (AERA, APA, NCME, 1999)</td>
<td>Cheating, stand-ins, equal treatment, testing environment (Standard 5.6, 8.7, 7.12, 5.4)</td>
</tr>
<tr>
<td><strong>Principles for the Validation and Use of Employee Selection Procedures</strong> (SIOP, 2003)</td>
<td>Security, standardization of testing conditions, candidate authentication, and accessibility of testing (pp. 40, 55, 56)</td>
</tr>
<tr>
<td><strong>International Guidelines on Computer-Based and Internet Delivered Testing</strong> (ITC, 2005)</td>
<td>Technology, psychometric quality, levels of testing supervision, and security</td>
</tr>
<tr>
<td><strong>Guidelines for Computer-Based Testing</strong> (ATP, 2002)</td>
<td>Technology, development, test administration, security</td>
</tr>
</tbody>
</table>

[www.intestcom.org/guidelines](http://www.intestcom.org/guidelines)
Research
Cheating
Impact on Validity & Decisions

### Type 1 Decision Error (False Positive)

<table>
<thead>
<tr>
<th>Cheat</th>
<th>No Cheat</th>
<th>Fail</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>True Negative</td>
<td>False Positive</td>
<td>Pass</td>
</tr>
<tr>
<td>Pass</td>
<td>False Negative</td>
<td>True Positive</td>
<td></td>
</tr>
</tbody>
</table>
Cheating Simulation Study

- *What is the potential impact of cheating?*
  - Validity
  - Decision-making: Type-1 errors (“false positives”)

- *How do different factors moderate impact?*
  - Rate of cheating among examinees
  - Success in cheating - score inflation
  - Who cheats (r.xc)
  - Cut Score Level - where decisions are made
Cheating
Impact on Predictive validity

Simulation Study
(Weiner & Ruch, 2006)
Bivariate normal distributions:
N=100,000 Xc = X + e (inflated score)

Weiner & Ruch, 2006
Cheating Impact on Decisions

Type-1 Decision Errors
Low cut score (-1SD) \( r_{xc}=0 \)

Additional passers

Up to 9% error

Score Inflation (SDs)

Percent error
Cheating
Impact on Decisions

Type-1 Decision Errors
High Cut score (+1SD) \( r_{xc}=0 \)

Percent error

Score Inflation (SDs)

Up to 42% error
Cheating Study Conclusions

- **Impact of cheating on validity and decisions appears to be negligible**, when...
  - Cheat rate is low (< 10% cheat rate)
  - Score inflation is low to moderate (< 1SD)
  - Passing score is low to moderate passing score

- **Impact of cheating can be dramatic under extreme conditions**
  - Validity & decision accuracy reduced by half or more
  - Substantial Type-1 errors (40% false positives)
  - **We don’t know the cheat rate on any given exam**
Verification Strategies

Model 1: Consistency Testing

- UIT $\rightarrow$ initial screen
  - PT $\rightarrow$ check consistency for passers (cheating detection)
    - If verified, UIT results are retained
    - If not consistent, PT replaces UIT for decision making

Model 2: Successive Hurdles

- UIT $\rightarrow$ initial screen
  - PT $\rightarrow$ secondary screen
    - PT measures same and/or additional traits for decision making
    - No detection of cheating
Verification Value

- Economic value
  - Valid tests enable better decisions
    
    $ value from increased productivity, reduced staffing costs
  - Cheating reduces validity, results in selecting less qualified candidates

    Verification value is in avoiding losses due to cheating

- Other value
  - Integrity, culture, sets the tone
  - Organizational credibility, brand
Verification Value

Example: Cost of Cheating per 100 Hires
Four Job Scenarios, Four Levels of Cheating
Assumptions: \( r=.30, \) \( \text{SR}=50\% \)

The classic utility model (Brogden, 1946)

\[ U\Delta = (r \text{ SDy Xs}) - C \]
Verification Conclusions

- Verification offers value in risk assurance vs. potential loss due to cheating
- Beneficial especially for high volume UIT programs
- Advantages of alternative verification models warrant further exploration
- Further research needed to develop guidelines for practice
Future Challenges
Remote Proctoring: Near future techniques

- **Apple “Guided Access” on iPads (iOS 6):**
  - Developed by Apple to address both candidate accessibility as well as security issues by disabling the home button and restricting touch input on certain areas of the screen.
  - Online testing applications can leverage this feature to provide more secure (locked down), mobile, testing environment for candidates that is built right into the devices.

- **Virtual proctoring with no live human proctor:**
  - Using web cams to track eye movement, facial action coding interpretation, face tracking, facial recognition.
  - Audio analysis techniques to identify external unauthorized sources.
  - Virtual proctor can not only document incidents but can suspend exam, communicate with candidate.
Remote proctoring: Distant future techniques

• Assessment systems that decode more directly what candidates know and can do:
  - Research in the area of decoding mental states from brain activity is early but promising
  - “Scientists have developed a computer program that predicts the mental patterns a picture will elicit, and thus can tell what somebody saw simply by looking at their brain’s activity.” Wired, 2008
  - Candidates ‘think’ of the right answer: E.g., How to safely operate a crane in a certain context
Future Challenges

• Rapid technology advances will enable multiple delivery devices
• Strategies to neutralize cheating concerns
• Further research needed to develop guidelines for practice
Conclusions

- UIT is now widely used in low stakes exams for TA and distance education
- Verification offers value in risk assurance vs. potential loss due to cheating for moderate stakes exams
- Proctored delivery remains method of choice for high stakes exams (today)
Discussion

- What do you see as the opportunities or threats related to UIT that deserve more focus within your testing program?

- Can you envision a future in which UIT would become the norm for testing? What would make this possible or impossible?

- Have your candidates or other stakeholders started asking about UIT?

- Will technology replace human proctors?
Thank You

If any questions, contact

jweiner@psionline.com

gregp@getyardstick.com