



Council on Licensure,
Enforcement and Regulation



2012 Annual Educational Conference · September 6-8

San Francisco, California

Alternative Proctoring Methods: *Considerations and Future Challenges*

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

Promoting Regulatory Excellence



Session Overview

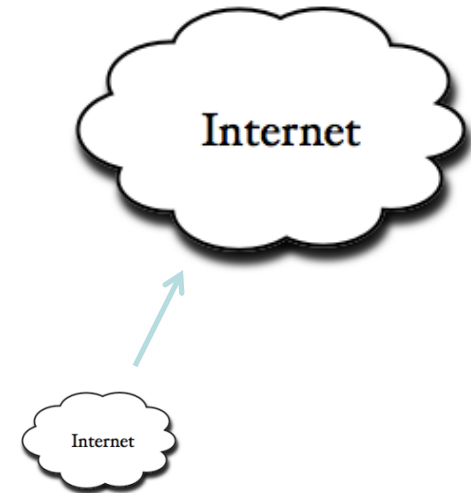
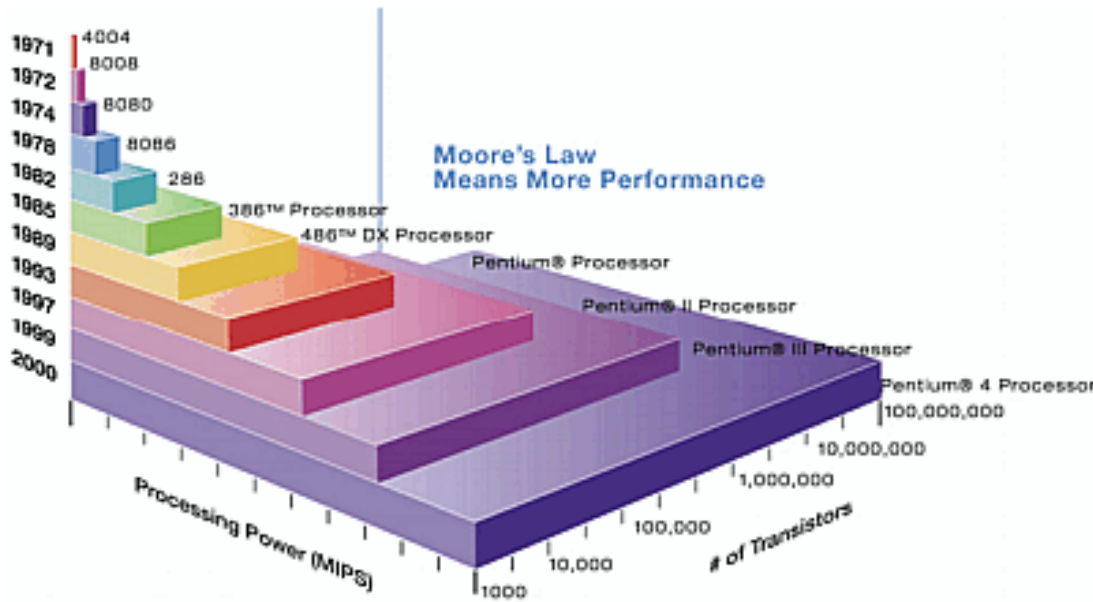
- Introduction
- Example Applications
- Considerations & Research
- Future Challenges
- Open Discussion



Introduction

*Advances in technology + economic forces
have spurred new assessment models*

- Design, Delivery, Use





Alternative Delivery Models

**Low
Stakes**

**High
Stakes**



**Unproctored
Internet Testing
(UIT)**

**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



2. KNOWLEDGE AND ASSESSMENT TOOL (KAT) TEST

Knowledge and Assessment Tool Agreement

The College of Licensed Practical Nurses of Alberta's (CLPNA) Knowledge and Assessment Tool (comprised of a Resource Guide & a Test) is designed to assess LPN knowledge of legislation, regulation, and other governing documents, which outline the key concepts set out in standards and code of ethics, and application of procedural knowledge essential to the safe and effective practice of an LPN in Alberta, Canada.

Although the CLPNA Knowledge and Assessment Test is an "open book" exam, registrants taking the Test should consider that a significant amount of expertise and investment have gone into the development of this Test. Therefore, it is considered very serious to compromise the security and confidentiality of the Test.

Please help protect the integrity of the CLPNA Knowledge and Assessment Test by carefully reading and abiding by the Registrant Statement of Understanding below. Report any concerns regarding the CLPNA Knowledge and Assessment Test security to the CLPNA immediately at info@clpna.com or 780-484-8886.

Thank you.

Registrant Statement of Understanding

Conduct occurring before, during or after participation in the CLPNA Knowledge and Assessment Test that violates these principles may result in invalidation of Test results.

I have read, understand and agree to abide by the above statement.

1. Registrants must complete this CLPNA Knowledge and Assessment Test individually and without help from any other person.
2. The CLPNA Knowledge and Assessment Test is confidential. It is strictly forbidden to copy, share, reproduce, or save any of the items or content from the CLPNA Knowledge and Assessment Test. The CLPNA Knowledge and Assessment Test questions are the property of the College of Licensed Practical Nurses of Alberta. Unauthorized disclosure of the CLPNA Knowledge and Assessment Test questions is prohibited under copyright laws. By declaring you agree to this statement of understanding, you will maintain the confidentiality of the CLPNA Knowledge and Assessment Test questions.

By taking the Test, you agree to the following:

1. Keep the CLPNA Knowledge and Assessment Test content confidential.
2. Realize that any recording or memorization of CLPNA Knowledge and Assessment Test questions is strictly forbidden.
3. Any cheating and/or breach of confidentiality/security or any attempt to subvert the CLPNA Knowledge and Assessment Test process by any registrant violates the purpose and principles of the CLPNA Knowledge and Assessment Test. Violations must be reported to the CLPNA immediately at info@clpna.com or 780-484-8886.
4. The CLPNA reserve the right to cancel or withhold any CLPNA Knowledge and Assessment Test results when, in their sole opinion, a testing irregularity occurs; cheating has occurred; there is an apparent discrepancy in, or falsification of, a registrant's identification; a registrant engages in misconduct or plagiarism; when aberrancies in performance are detected for which there is no reasonable and satisfactory explanation; or the results are believed to be invalid for any other reason.

By clicking the agreement button below, you agree to abide by all regulations, as well as oral and written instructions controlling the conduct of the CLPNA Knowledge and Assessment Test. These regulations are intended to preserve the integrity of the CLPNA Knowledge and Assessment Test process by providing standard conditions that yield valid and reliable results.

I agree to the above statements

You're about to begin **2. Knowledge and Assessment Tool (KAT) Test**.
Clicking the button below starts the exam.



Exam Time limit: 3 hours

The timer begins as soon as you click the button below. Your exam will automatically be submitted when the timer expires.

Start exam >



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*

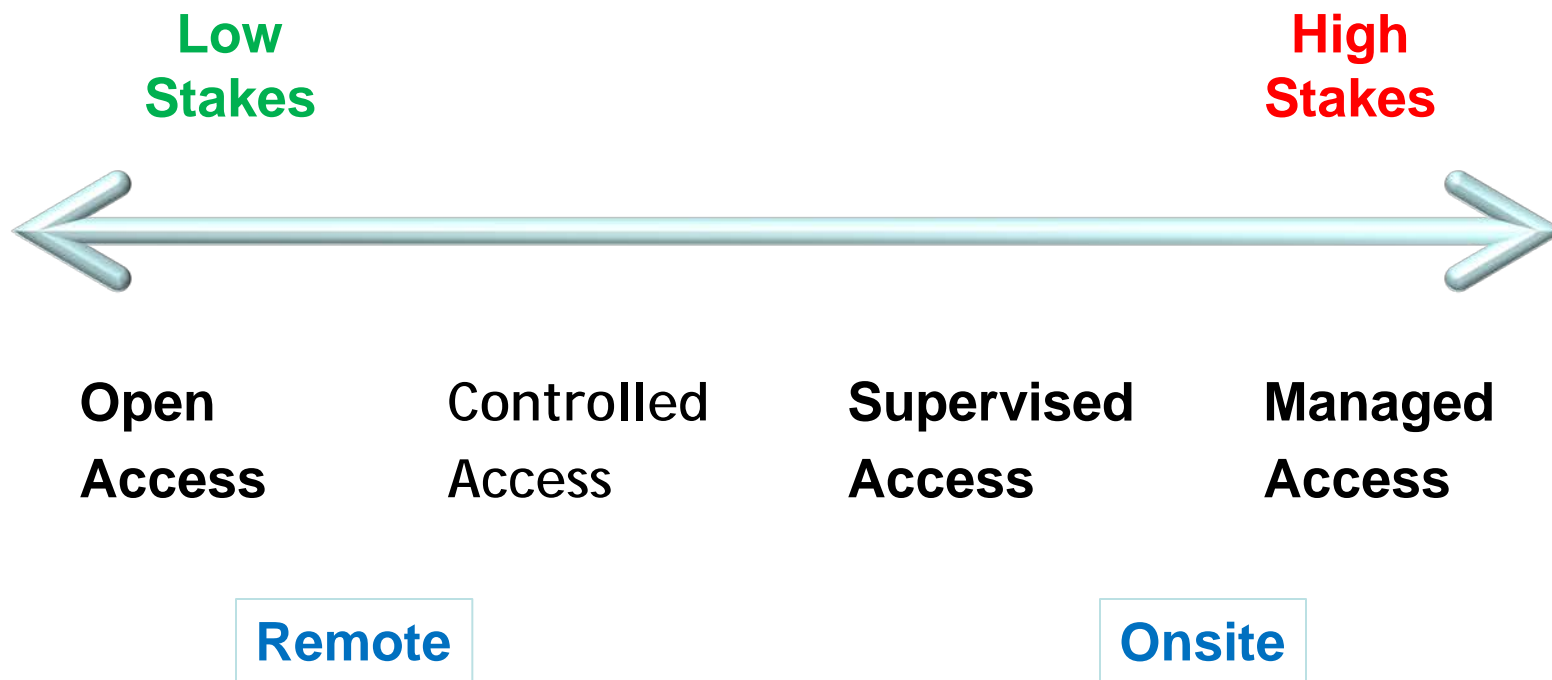


What's at Stake?

Stakes	Context	IP Costs	Cheating Costs	Safety/Legal Risks
Low	<i>Training & Development; Low impact decisions</i>			
Moderate	<i>Moderate impact decisions</i>			
High	<i>Public safety; High impact decisions</i>			



Risk Management





What do Test Standards say?

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



Research





Cheating

Impact on Validity & Decisions

Type 1 Decision Error (False Positive)

No Cheat	Cheat	
	Fail	Pass
Fail	True Negative	False Positive
Pass	False Negative	True Positive



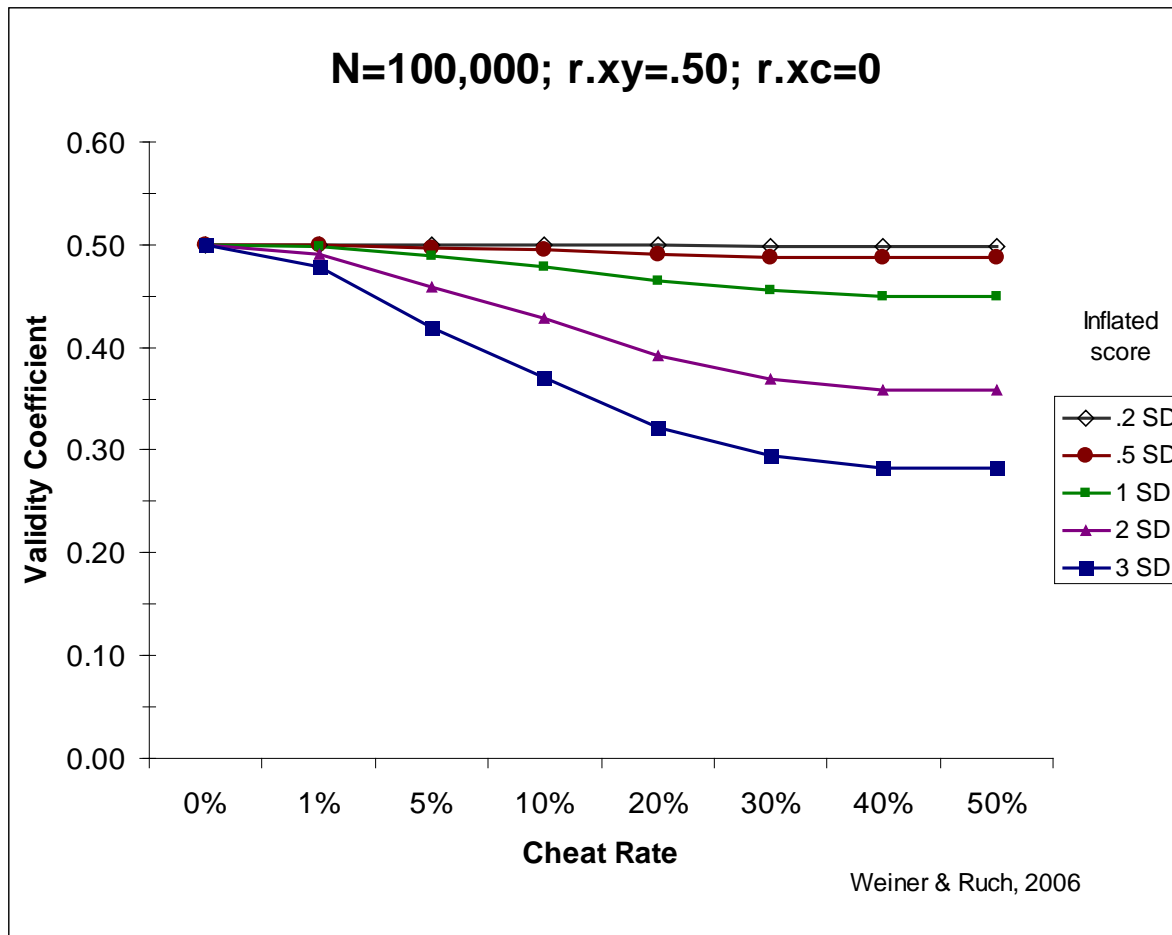
Cheating Simulation Study

- *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:
 $N=100,000$
 $X_c = X + e$ (inflated score)

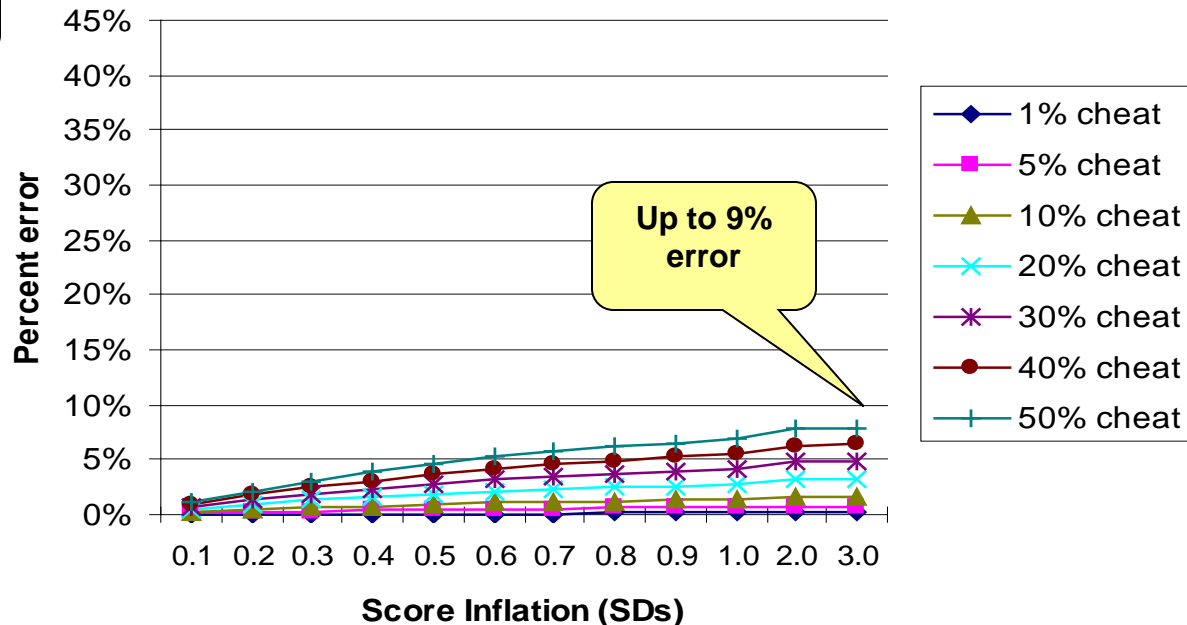


Cheating

Impact on Decisions

Additional
passers

Type-1 Decision Errors
Low cut score (-1SD) $r.xc=0$

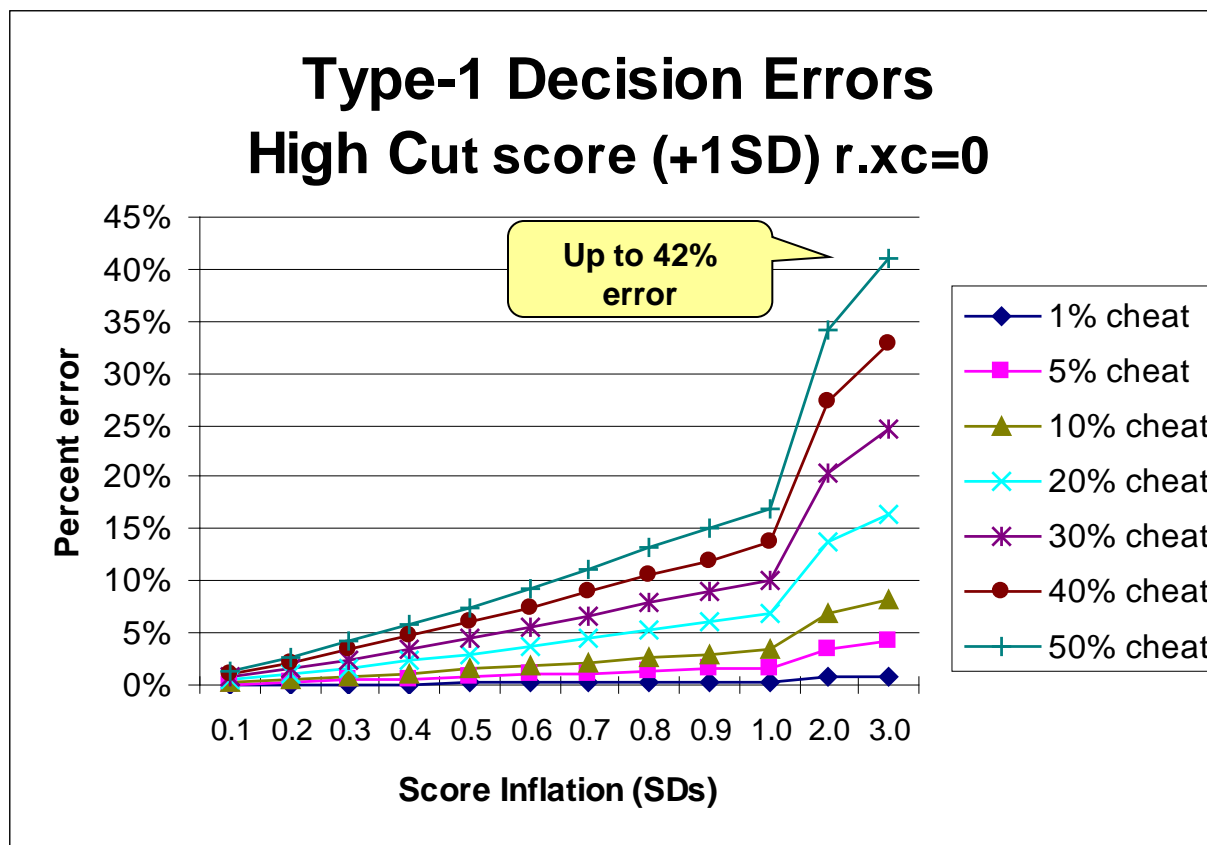


Up to 9%
error



Cheating

Impact on Decisions





Cheating Study Conclusions

- *Impact of cheating on validity and decisions appears to be negligible, when...*
 - Cheat rate is low ($\leq 10\%$ cheat rate)
 - Score inflation is low to moderate ($\leq 1SD$)
 - 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 → *initial screen*
 - PT → *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 → *initial screen*
 - PT → *secondary screen*
 - PT measures same and/or additional traits for decision making
 - No detection of cheating

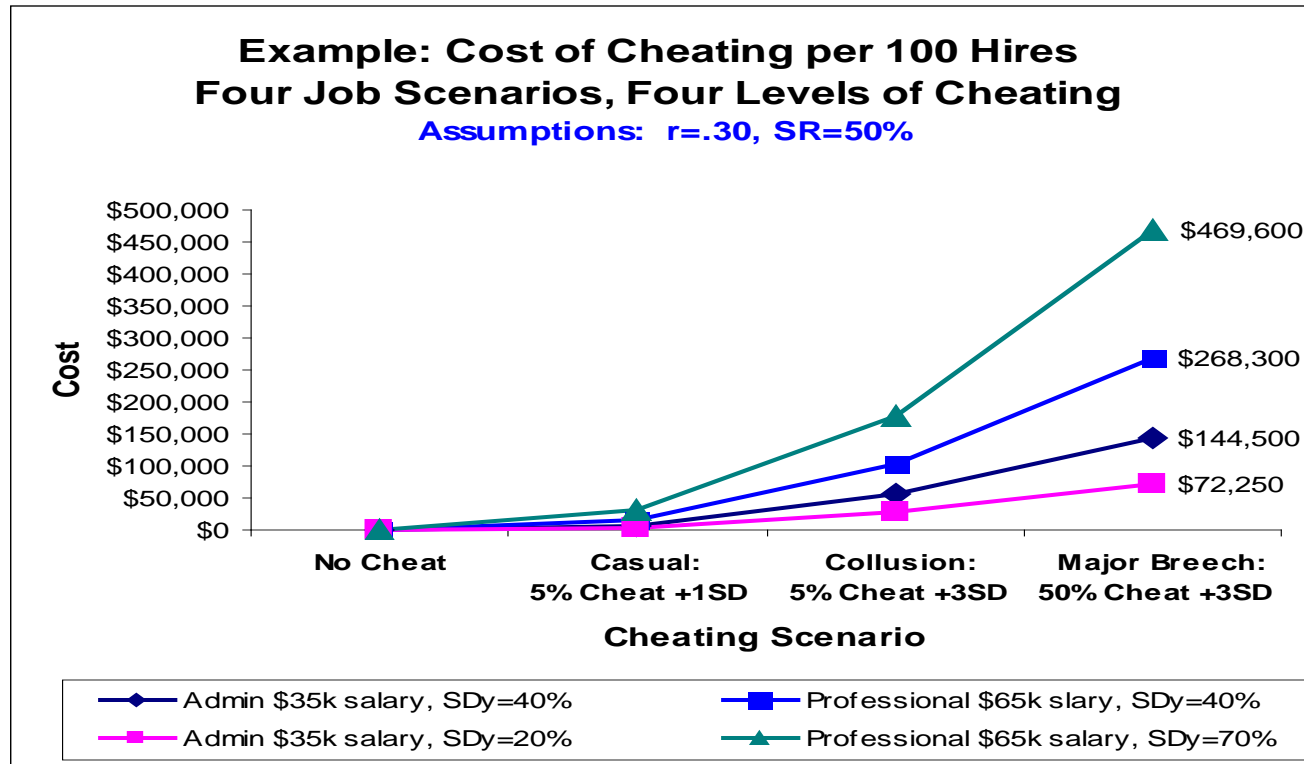


Verification Value

- 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



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](#)
 - Kay, Naselaris, Prenger & Gallant(2008) “Identifying natural images from human brain activity” Nature, v452.
 - 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