









# **How to Defend the Integrity of a Testing Program Against Cheating?**

**Wim J. van der Linden**  
**CTB/McGraw-Hill**



# Outline

- Practice of cheating 
- Prevention and detection
  - Security policy 
  - Security audits 
  - Web patrol 
  - Item/testing format 
  - Data forensics 


# Practice of Cheating

- Cheating is no longer the old-fashioned battle between teachers and students
- Facilitating factors
  - Higher stakes for examinees
  - Strong sense of entitlement
  - Walk-in testing
  - Anonymity of test takers
  - Miniaturization of electronic devices  

# Practice of Cheating *Cont'd*

- Not only students cheat
  - Organized crime; e.g., [www.actualtests.com](http://www.actualtests.com) 
  - Test proctors
  - Admission officers
  - Teachers/school administrators 
    - Levitt & Rubner (2005), *Freakonomics: A rogue economist explores the hidden side of everything*.
  - For weekly updates, see [www.caveon.com/cheating\\_news.php](http://www.caveon.com/cheating_news.php)

# Practice of Cheating *Cont'd*

- Frequency of student cheating
  - Fox & Meijer (APM, 2008) 
- Common types of cheating
  - Answer copying
  - Collusion
  - Crib sheets
  - Item harvesting
  - Answer erasures

# Practice of Cheating *Cont'd*

- Common types of cheating
  - Impersonation
  - Plagiarism
  - Etc.



# Security Policy

- Testing companies are no longer naïve about cheating
  - Validity of the test scores
  - Fairness to other test takers
- One of the best lines of defense is a clear, accessible policy against cheating
  - Rules regarding confidentiality, proper handling, storage and transmission of materials, instruction to proctors and test takers, etc.

# Security Policy *Cont'd*

- Clear, accessible policy *cont'd*
  - Should include definition of cheating
  - Procedures and sanctions (“zero tolerance”)
  - Publishing of actions
- Example of GMAT





# Security Audit

- Have experts periodically check all procedures, systems, materials, etc., in your testing program
  - Policy
  - Storage, handling of materials
  - Checks on staff
  - Item types, testing format
  - Psychometric procedures
  - Computer systems



# Web Patrol

- Checks on websites, chat rooms, social media
  - Compromised test items?
  - Collusion?
  - Any other security breaches?
  - Etc.



# Item/Testing Format

- Some item types are more sensitive to cheating than others
  - Multiple-choice tests
  - Performance assessments
- Item/test re-use
  - Testing-window size
  - Item-pool size
  - Rotating item pools

# Item/Testing Format *Cont'd*

- Rule-based item generation
  - Randomization of minor elements (e.g., answer keys, context descriptions, data)
- Testing format
  - Adaptive testing
  - Item-exposure control methods
- Published item pools

# Item/Testing Format *Cont'd*




- Use of verification tests
  - e.g., ASVAB



# Data Forensics

- Statistical tests to detect fraudulent behavior
  - Post-hoc flagging of test takers
  - Real-time checks
- Detection of cheaters vs. compromised items
- Types of data
  - Responses
  - Response times

# Data Forensics *Cont'd*

- Statistical tests for the detection of cheating should be used in combination with other sources of evidence
- Three examples
  - Answer copying 
  - Preknowledge of items 
  - Erasure analysis 



# Spyer™ Agent Watch

## Camcorder

This stylish watch has a secret. It contains a hidden color camcorder. Ideal as both a spy camera & a hidden camera small enough to wear anywhere! Wherever you go with your mini camcorder on your wrist, you'll know for sure that you won't miss anything! No one will know you are spying or using it for your own personal security. With built-in 2GB flash memory and rechargeable lithium battery, it lets you record both video and voice. This is one of the most useful spy gadgets we have ever tried. The built-in DVR makes this truly portable and spy worthy.



Spyer™ Agent Watch  
KC257SG \$149.95

USB cable and  
traveler charger



USB Slot





# USA TODAY

A GANNETT COMPANY

## Newsline

MONDAY, MARCH 7, 2011

SPECIAL REPORT TESTING THE SYSTEM

# When scores seem too good to be true

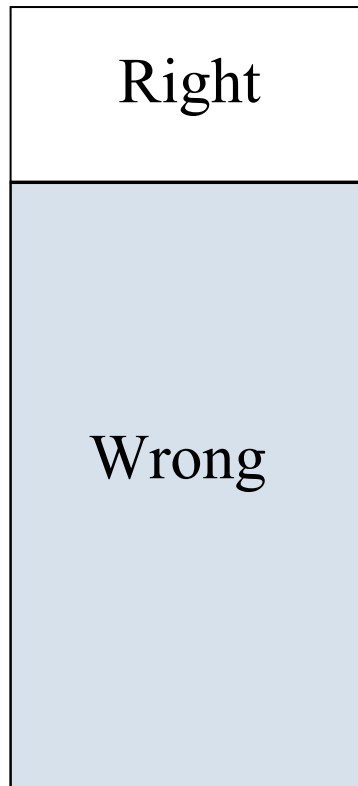
To help students and schools  
make the grade, are some  
educators crossing the line?

# Answer Copying

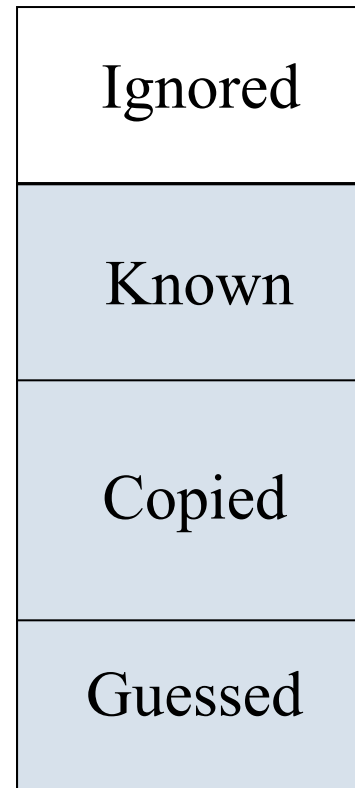
- Example of shifted binomial test
- Assumption of three-stage response process:
  - Test taker thinks he knows answer and gives it
  - Test taker does not know answer and tries to copy from neighbor
  - Test taker cannot copy and guesses randomly
- Focus on agreement of matching wrong answers between suspected copier and source

# Answer Copying *Cont'd*

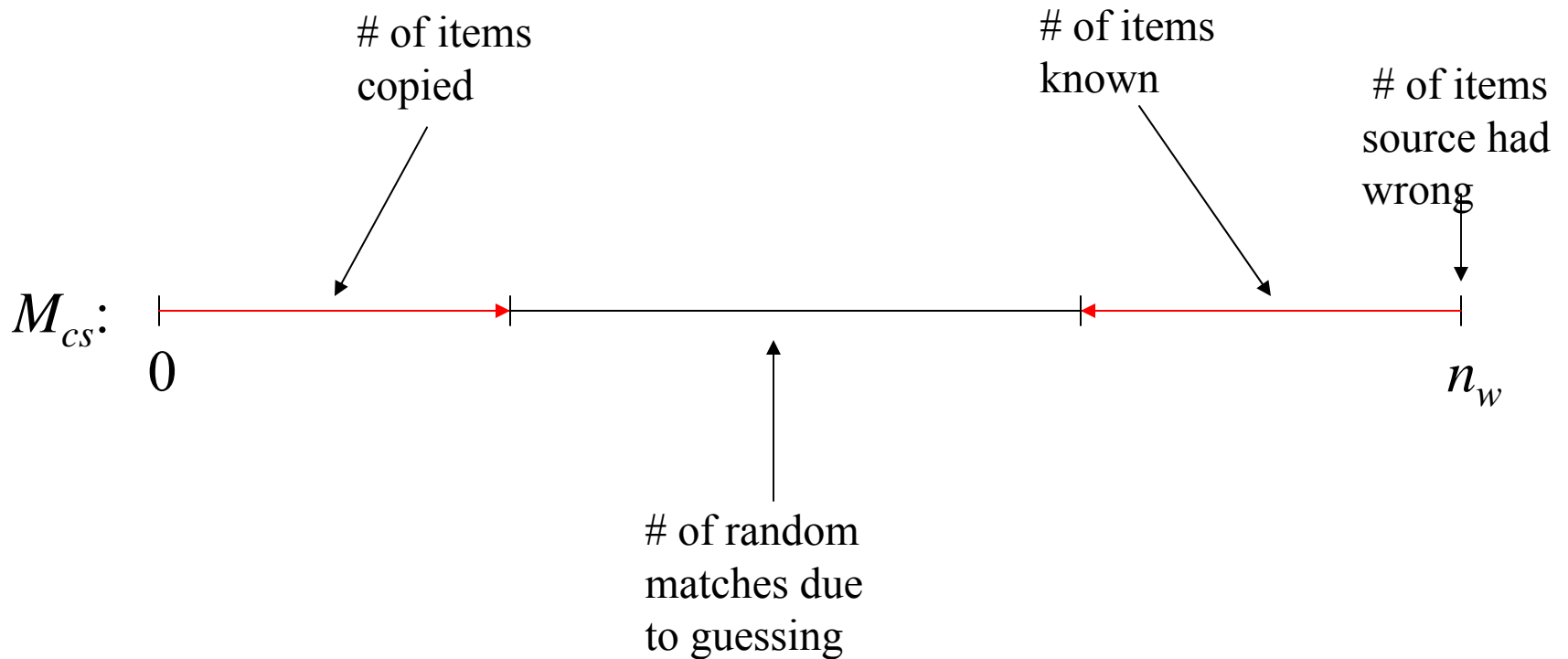
Source




Copier



# Answer Copying *Cont'd*



# Answer Copying *Cont'd*

- Null distribution
  - Binomial distribution on range of number of random matches
  - Critical value follows from choice of significance level  $\alpha$
- Example of power curves 



# Preknowledge of Items *Cont'd*

- Use of responses and response times
- Advantages of response times
  - More informative
  - Can be used for adaptive tests
  - Hard to fake realistic response times
- Lognormal response-time model
  - Item and person parameters (as in IRT model)

# Preknowledge of Items *Cont'd*

- Lognormal RT model:

$$f(t_{ij}) = \frac{\alpha_i}{t_{ij} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} [\alpha_i (\ln t_{ij} - (\beta_i - \tau_j))]^2 \right\}$$

Speed  
↓  
Time intensity  
↑  
Discrimination

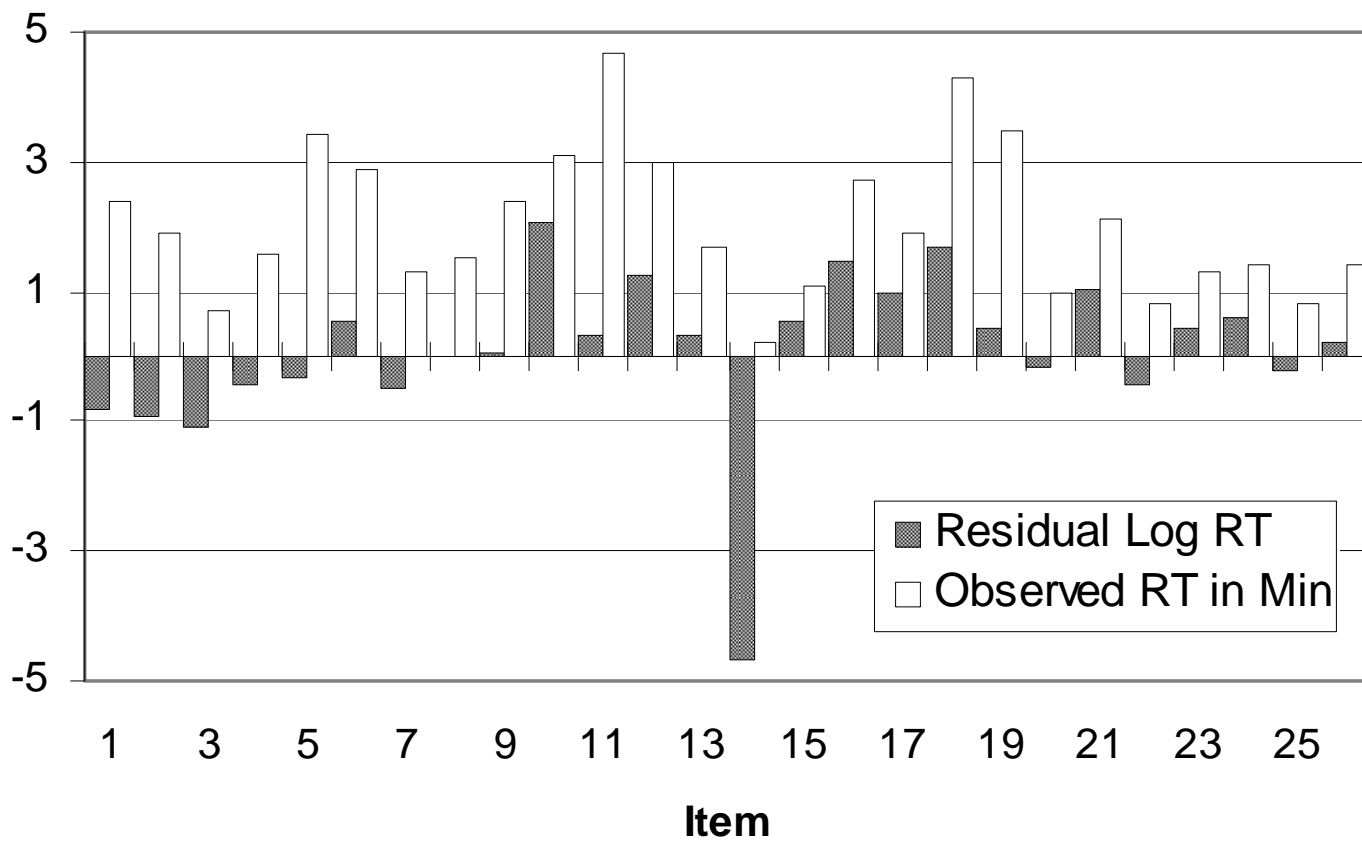
# Preknowledge of Items *Cont'd*

- Example of suspicious item from GMAT study
  - Large negative RT residual (-4.66)
  - RT of 12.3 seconds (expected RT under the model was 88.9 seconds!)
  - Test taker had correct response but very low estimated ability relative to item difficulty
  - Four other test takers with same behavior on same item





# Suspicious Item (Nr. 14)



# Erasure Analysis

- Optical scanners can be set to detect erased answers with changes from:
  - right to wrong (RW)
  - wrong to wrong (WW)
  - wrong to right (RW)

# Erasure Analysis *Cont'd*

- Two-stage modeling
  - Regular IRT model for probability of correct initial response
    - $\Pr\{U^{(1)}=1\}$
  - Two *conditional* models for probability of a correct response upon review
    - $\Pr\{U^{(2)}=1|U^{(1)}=0\}$
    - $\Pr\{U^{(2)}=1|U^{(1)}=1\}$
- No local independence between  $U^{(1)}$  and  $U^{(2)}$

# Erasure Analysis *Cont'd*

- Statistical test of number of WR erasures
- Residual analysis
  - Example for large-scale assessment (Grade 3 math) 