



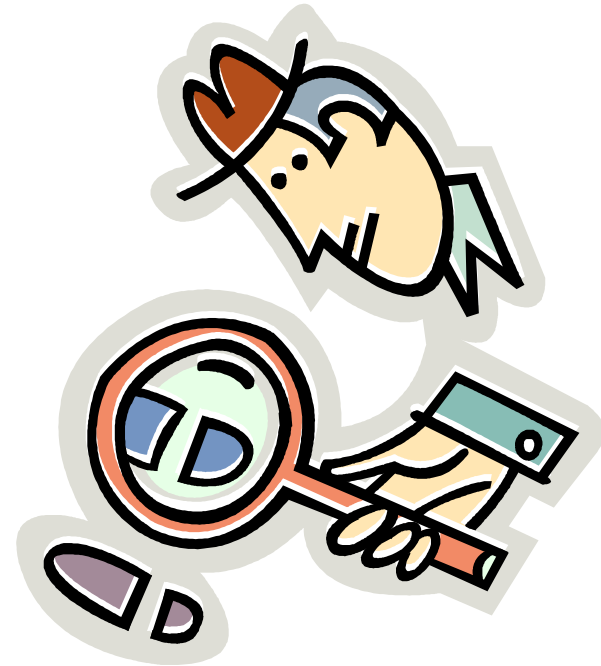
Protecting Your Testing Program The Answer's in the Data

Dennis Maynes
Chief Scientist

November 17, 2010

Webinar Agenda

- Overview of science
- Security breaches
- The legal case & policy
- Case studies with methods
- Conclusion



Data Forensics

- The science and discipline of gathering and evaluating statistical evidence of test fraud.
- One of at least three important security pillars for your certification and licensure program.
- Top Three Objectives
 - Data Forensics evidence can be acted upon if it is credible, defensible, objective and strong.
 - You can protect your program using Data Forensics if you establish fair and reasonable policies.
 - With our Data Forensics capability, we have the expertise and ability to help strengthen your exam security.

Overview of Science

- Test scores are trusted
 - Life-changing decisions
- The data are the best record of the testing session
 - The data can be relied upon
 - Other statistics can be trusted
- Risky behaviors leave forensic tracks
 - Collusion
 - Answer-sheet tampering
 - Exposure of content
 - Brain dump usage

What can Data Forensics Do?

- Infer existence of potentially fraudulent behaviors
 - Individual behaviors
 - Group behaviors
 - Exam security risks
- Data Forensics modes
 - Investigation
 - Monitoring

If you are willing to trust the test score data to make high stakes decisions, then shouldn't you be willing to use the data to determine whether scores are trustworthy?

Security Breaches

- Pre-knowledge
 - Access to forbidden content
 - Improper assistance by session administrators
 - Disclosure of content by instructors
- Collusion and/or Collaboration
 - Answer copying
 - Impersonators and proxy test takers
 - Cheat rings – “Flying V”
- Tampering
 - Changing answer sheets
 - Changing scores

Detection of Behaviors

<p>Pre-Knowledge</p> <ul style="list-style-type: none">–Use of Braindump content–Rogue Review courses–Disclosure of content–Answer key theft	<p>Aberrance / Person-fit</p> <p>Gain scores</p> <p>High pass rates</p> <p>Perfect tests</p> <p>Score differencing</p> <p>Trojan Horse items</p>
<p>Collusion</p> <ul style="list-style-type: none">–Impersonators–Proxy test takers–Sharing content	<p>Similarity</p> <p>Identical tests</p> <p>Large clusters – similarity counts</p> <p>Source-copier analysis</p>
<p>Tampering</p> <ul style="list-style-type: none">–Answer sheet falsification–Score report falsification	<p>Erasure Counts</p> <p>Associated gains</p> <p>Inconsistent score data</p>

The Legal Case

- Disclaimer: I am not a lawyer
- Do you have to prove “cheating”?
- Contract law – “good faith”
 - Language of agreements
 - Documented policies and procedures
 - Taking all steps to show “good faith”

Great Deference to Exam Administrators

- State Actors: Due process
- No guarantee, but ...
 - Establish “good faith”
 - Present compelling defense
 - No “victories” by cheaters
- Benchmark examples from case law

Criteria for Using DF Evidence

- **Factual** –something that happened
- **Defensible** –scientific methods
- **Objective** – statement of probability
- **Credible** – inferred risky behavior
- **Strong** – very unlikely (<1 in 1 million)

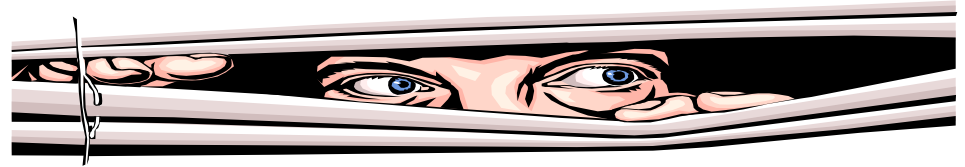
Nature of Circumstantial Evidence

- Credibility of Evidence
 - Are the data credible?
 - Do plausible explanations exist?
- Deduction
 - Are any of the statistics extreme?
 - Are all of the statistics extreme?
- Pass rate spreads (evidence of an advantage)
 - “unfair advantage”
 - No harm, no foul?

Questions You MUST Answer

- Which behaviors pose the greatest risk to your exam security?
- How can you prevent item theft?
- What can you do when your items are stolen?
- Do your test use agreements support your actions?
- At what level of extremeness will you take action? At a level of DNA matching (1 in 40 million)?

Case #1: Copying



- Two test takers at same table
- Proctor observed “roving eyes”
- Witnessed by another proctor
- Mid-way through, “neighbor” covered answer sheet
- “If you check both answer sheets you may be able to determine at which point there was a difference in the answers, if any.”
- “My neighbor cheated! I didn’t cheat!”

Response Data

Key - Q1-25	ACCBCAADDCCDABDDBBBCBAAAD
Source	ACCBCCBADDCBDABDBBCBBAAADD
Copier	ACCACCBADCBDABDDBCBCAAADD
Key - Q26-50	BCACBBAADDDCBABCCBBADACCB
Source	BAACBAACADDCAABCCDBADACCB
Copier	CCCACCDDBDACBABCBADAAACCA



Tally of Results

	Source	Copier	Matches	Wrong	Right
Score 1-25	16	17	22	7	15
Score 26-50	19	11	8	0	8
Total	35	28	30	7	23
Test Score	70%	56%			
1 st Half Score	64%	68%			
2 nd Half Score	76%	44%			
Match Score	77%	77%			
Disagreed Score	60%	25%			

Three measurements to consider

- Probability of matches
- Probability of score changes (match versus mismatch)
- Probability of split-half randomness
- Combine probabilities to strengthen evidence
 - http://www.caveon.com/articles/dennis_ncme_09.htm

Analysis of Case #1

- Assume
 - items of equal difficulty
 - Probability of a match = .392 (about 40%)
- Probabilities
 - 30 matches or more in 50 questions: 1 in fifty
 - 22 matches or more in 25 questions: 1 in 100,000
 - randomness between halves: 1 in 20,000
 - observed advantage copier 77% versus 25%: 1 in 2,500
 - all three (30 matches, randomness, difference): 1 in 125,000
 - all three (22 of 25 matches, randomness, difference): 1 in 10 billion

Survey

Should the proctors have intervened or should they have let the copying continue?

Case #2: Rogue Review Course

- Detection through monitoring
 - Compute many statistics
 - Search for many behaviors
 - Analyze individual tests
 - Aggregate marginal flags
 - Use small sample techniques
- Puzzled: “rapidly answered tests and high pass rates”
- Confirmation by investigation



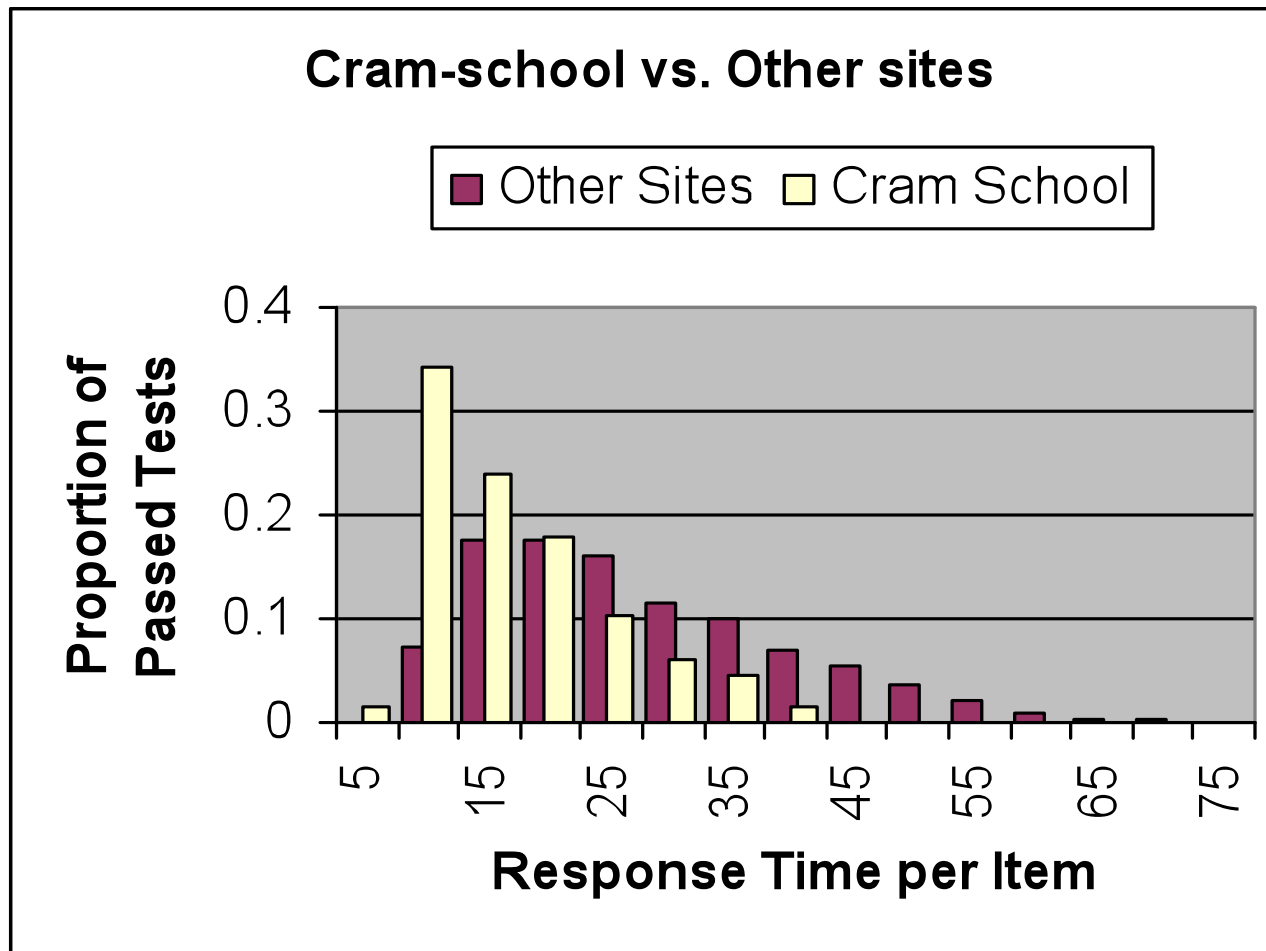
Computed Statistics

- Guttman's G Statistic
- Latency Aberrance
- Fast-High Statistic
- Rapid Response Statistic
- M4 Similarity
- Volatile Scores
- Perfect Tests
- Identical Tests
- Retake Violations

Tally of Salient Results

Statistic	Baseline	Site of Interest
Number of Tests	1152	78
Pass Rate	0.69	0.86
Pass Index	0	3.51
Rapid Response Rate	0.17	0.51
Pass Rate With Rapid Response	0.73	0.98
Pass Rate Without Rapid Response	0.68	0.74
Rapid Response Difference Index	1.08	2.59
Rapid Response Rate Index		12.45

Fast Tests that Passed!



Analysis of Case #2

- Index value: probability = $10^{-\text{index}}$
- Probability of 40 or more “rapid tests” from sample size of 78 < 1 in 1 trillion (index = 12.45)
- Probability of pass rate change (.98 versus .74) = .00257 (index = 2.59)

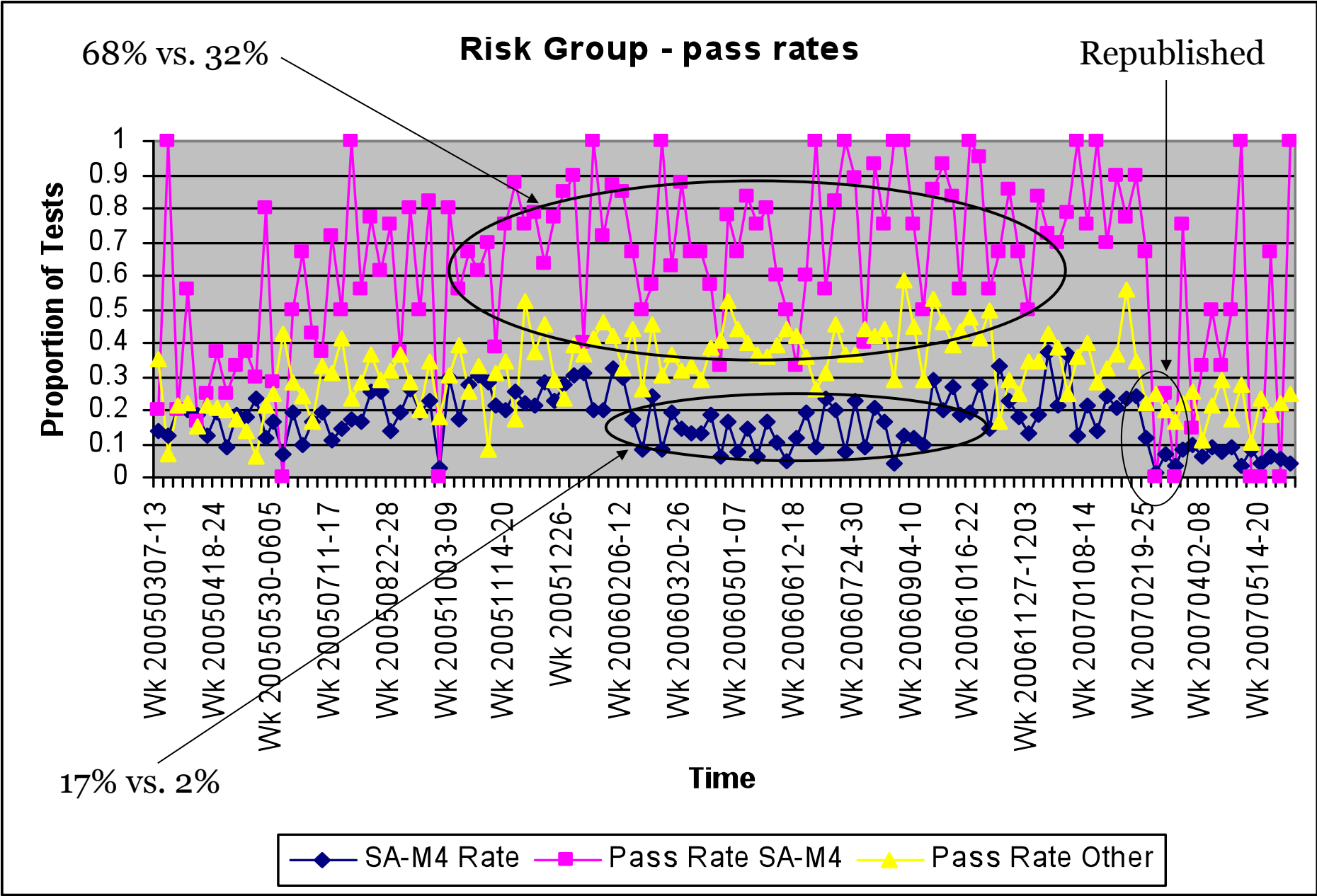
Case #3: Item Harvesting



- Internet rumor: “excellent” review course
- Course participants were “recalling” items
- Acquired content from review course
- Raided by local law enforcement
- Risk group compromise
- Climbing pass rates for risk group
- Forensic analysis confirmed “collusion”
- Pass rates plummeted with republication

Statistical Detector: SA-M4

- Created a combination statistic
 - Score aberrance AND
 - M4 Similarity
- Optimized to detect the behavior
 - Coaching on content AND
 - Score inconsistency (Theta lower than expected)



Analysis of Case #3

- Pass Rate Difference (68% vs 32%)
 - Probability < 1 in 10^{86} !!!
- Participation Detection (17% vs 2%)
 - Probability < 1 in 10^{300} !!!!
- Pass rate drop after republication
 - Overall spread no longer significant: 51.4% vs. 48.0%
- Twenty scores were invalidated.

Survey Question

- If your program had a breach similar to this, would you prefer to publicize what you did or would you keep it quiet?

Conclusion

- The SCIENCE behind Data Forensics is based on statistics and probability theory
 - Use of Data Forensics identifies behaviors and results that reflect elements of RISK
 - Use of Data Forensics improves the SECURITY and INTEGRITY of your testing program
- Data Forensics evidence can be acted upon if it is credible, defensible, objective and strong.
- You can protect your program using Data Forensics if you establish fair and reasonable policies.
- With our Data Forensics capability, we have the expertise and ability to help strengthen your exam security.



Thanks for attending!

And, yes!

The answer is in the data.