



MO EOC ASSESSMENTS

Data Forensics Procedure

August 22, 2013



OVERVIEW

- Four planned data forensics procedures:
 - Aberration
 - Similarity
 - Answer Change
 - Response Time

- Each procedure will flag outlier, or unusual, values in test scores
 - Simply point to a potential problem
 - Not as “cheating” or “forgery”

- Each procedure will flag outlier, or unusual, values in level that is uniquely and consistently coded in the test data.
 - “School” as level of analysis
 - Flags aggregated to the district level

ABERRATION

- ❑ This identifies response patterns that are inconsistent with a student's ability level.
- ❑ Item responses are calibrated into IRT parameters that are then used in a Maximum Likelihood Estimation scoring algorithm to determine an ability, or theta, score.
- ❑ The Likelihood is then standardized to produce an individual aberration statistic, $L(z)$.
- ❑ An unusually low $L(z)$ value reflects a response pattern that is not consistent with the difficulties of the items. Paired with a high test score, this would be an aberration flag for a student.

SIMILARITY

- This requires comparing every student's responses to every other student's responses within a group.
- Three implementations will occur as follows:
 - Calculate the Euclidean distance between two vectors of item point values.
 - Compute the percentage of items between the two students that have identical answer options for multiple-choice items and point values for constructed-response items.
 - Determine the ratio of the number of identical correct answers to identical incorrect answers.

ANSWER CHANGE

- This examines the ratios of answer changes from right-to-wrong and from wrong-to-right based on the history of the captured responses.
 - The state-level distribution of answer changes will provide the baseline against which individual students are flagged.
- Answer changes from wrong-to-right will be of primary interest.
- Flagging threshold
 - Top 1% with most answer changes

RESPONSE TIME

- This refers to how long it takes students to answer each item on an online test.
- A response that is answered in “superhuman” time is flagged and indicates that the student was possibly working from an answer key or some other cue.

OUTLIERS AND FLAGS

- When a group receives multiple flags, the question that arises is “What is the probability that this occurred by chance?”
 - One approach to answer this question requires resampling - sample a test’s entire data set many times by randomly sampling the group size from students across the state and calculating the distribution of the data forensics statistic.
 - Very small probability indicates that the flag is not a random occurrence.

REPORTING DATA FORENSICS RESULTS

- ❑ Flags would be displayed graphically and compared to the norm or expected probability.
- ❑ Data forensics statistics
 - not direct evidence of wrongdoing
 - but multiple flags or combinations of certain flags should be considered for follow-up investigations.
- ❑ It is not in the best interest or expense of a state or contractor to investigate every oddity in data forensics, and instead the most egregious situations should be chosen for investigation (Fremer, 2013)

STATE POLICY AND COMMUNICATIONS

Test security policy

1. Will a test security policy be established?
2. What are the purposes of the test security policy?
3. Does the test security policy include preventative measures?
What are those measures?
4. Does the policy include clear consequences for violations?
Will they be enforced?
5. Is there a clear process in place to conduct investigations of test security breaches?
6. How will the test security policy be communicated to districts, schools, teachers, students, and parents?
7. Will test security training be provided to staff (e.g., test administrators)?

STATE POLICY AND COMMUNICATIONS (CONT.)

Data forensics policy

8. Will the state establish a data forensics policy? What purposes will it serve?
9. Will data forensics results trigger investigations or score holds?
10. What results will trigger such investigations or score holds?
11. What are the investigation procedures?
12. Who will be conducting investigations?
13. Who will have access to information? Under what circumstances?
14. How will the data forensics policy be communicated to stakeholders?