Educational data forensics

Sebastiaan de Klerk
Conference on Test Security, Park City, UT
October 11 – 3:00 PM – White Pine II

xquiry
educational
data forensics

Who am I?

- Sebastiaan de Klerk
  - 2011 – 2015 PhD research project at eX:plain and University of Twente
  - 2016 – Involved in start-up Xquiry as an independent researcher and consultant
  - 2017 – Xquiry founded as a subsidiary of eX:plain
  - 2018 – Research on educational data forensics and consultant at Xquiry

Content

- Educational data forensics (EDF)
- EDF Monitor
- Empirical research – EDF Monitor
- Empirical research – EDF Protocol

Educational data forensics (EDF)

Educational Data Forensics

What: The use of (statistical) algorithms to detect potential test fraud in the response data of test takers.

Goal: Detection of potential test fraud (but also support prevention).

Why: Methods of test fraud keep advancing. This fact requires organizations to handle this proactively and on a continuous basis. Although test fraud cannot be 'exterminated', it is essential to fight it in the best way possible. EDF helps in this process.

EDF Monitor

EDF and aberrant patterns

- EDF is a set of methods and algorithms to detect aberrant patterns in test takers' response data. These aberrant patterns may indicate potential test fraud.
- We have developed a webbased application that can automatically analyze the response data of test takers. The algorithms are so-called person-fit indices.
- The analyses can be done on both the individual and the group level and on fixed and random tests.
- All item types can be analyzed for the detection of potential test fraud.

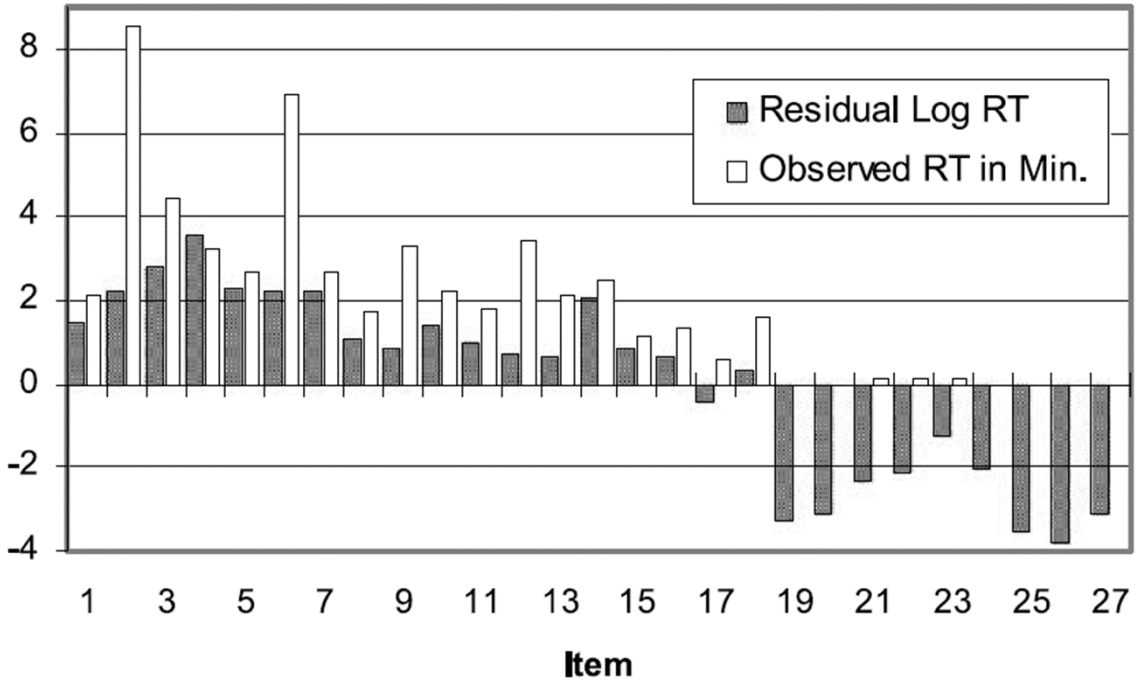
Indices aberrant patterns

- Individual
  - Guttman error analysis
  - Response time analysis
  - Combinations and small adaptations
- Group (in addition to above)
  - Similarity analysis
  - P-value analysis

Guttman error analysis (Meijer, 1994)

- Sort items from easy to difficult
- It's more likely that test takers will answer an easier item correct than a more difficult one
- Aberrations to this pattern are called Guttman errors
- We also incorporate several adaptations to the Guttman error model
  - These adaptations incorporate the 'distance' in difficulty (p-value or another difficulty parameter) between two items

Lognormal response time analysis (Van der Linden, 2006)

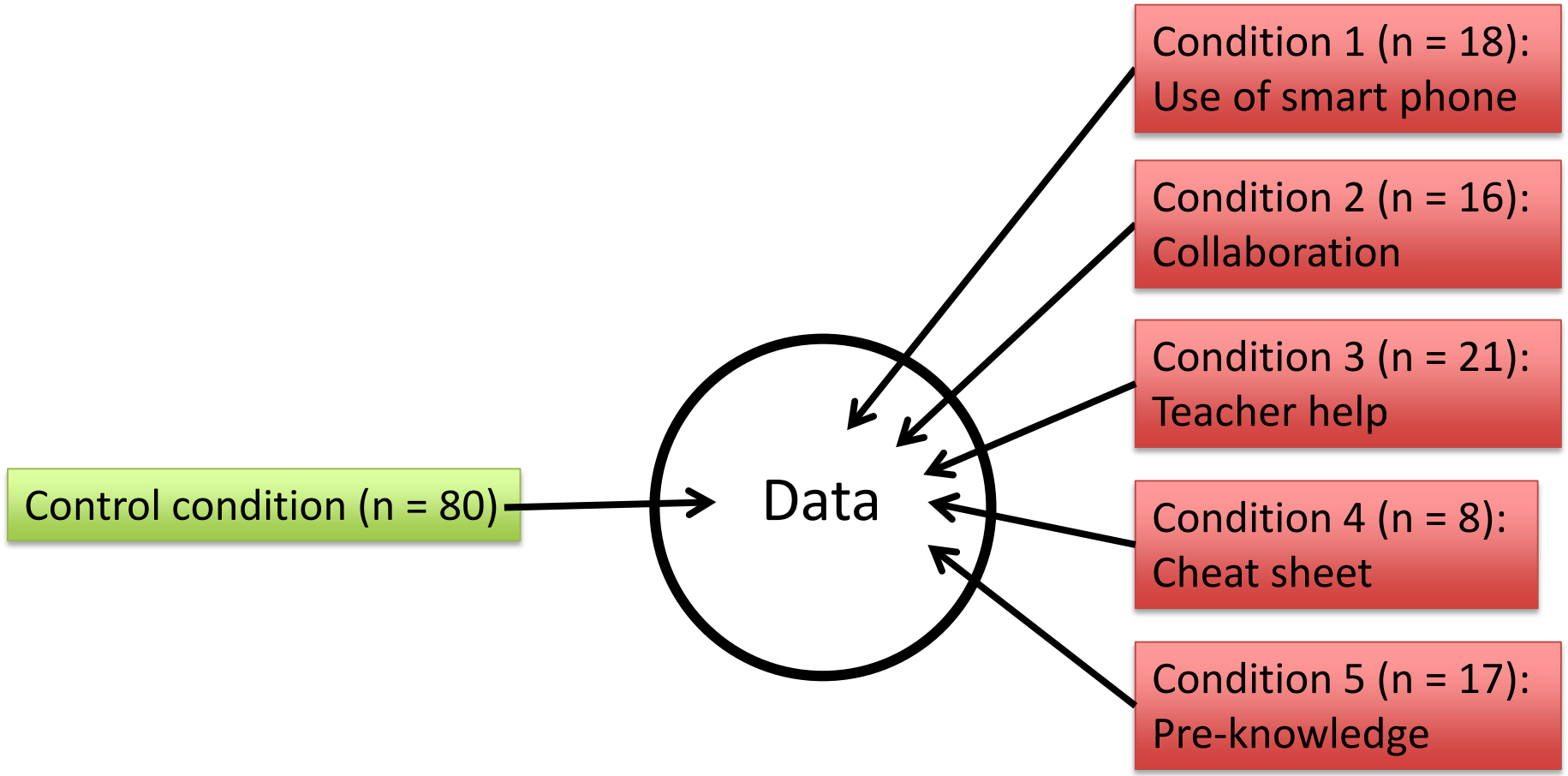


Empirical Research – EDF Monitor

Research questions

- Which manual setting(s) of the EDF algorithms yield most information about potential test fraud (i.e., the highest detection rate and reliability)?
- To what extent can a combination of algorithms (i.e., Guttman error analysis and response time analysis) detect instructed fraudulent test takers (detection rate) and what is the reliability (i.e., true positive ratio)?
- To what extent can various methods of cheating be detected through the EDF algorithms?

Data collection and procedure



Results

- We tested three indices: log-normal response time analysis (Van der Linden, 2006), standard Guttman error analysis (Meijer, 1994), and p-value (CTT) adjusted Guttman error analysis (Van Noord, 2018)
- All indices were significantly higher in the combined experimental cheating condition (n=80) than in the control condition (n=80)

Results

- A combination of the log-normal response time analysis *together* with p-value adjusted Guttman error analysis yielded the highest reliability or true positive ratio:
  - 96.8% of test takers who were labeled ‘fraudulent’ by the software were indeed in one of the experimental cheating conditions
  - The detection ratio was 37.5% (i.e., 37.5% of students in the cheating conditions were indeed detected)

Results

- The true positive and false negative ratio differed over the different conditions
  - The true positive ratio was significantly higher, and the false negative ratio significantly lower in the pre-knowledge and teacher help conditions

- The indices are best in detecting these types of test fraud

Discussion

- EDF algorithms do work:
  - Significantly more fraudulent students were detected
  - When you are detected it is rather likely that you have cheated
- Number of false negatives still too high (i.e., the detection ratio is still rather low), but on a higher abstraction (e.g., proctor, test center) level the likelihood of being detected is higher
- Increase number of indices in the software
- Improve logging methods
- Expand research

Empirical Research – EDF Protocol

EDF Protocol

- A set of 10 scientifically established evidence-based standards for the prevention and detection of test fraud
- Practice oriented
- Test security ‘audit’ for testing companies
- Measures the current level of security breaches in the testing program

Method

- Step 1: Literature search
- Step 2: Developing an EDF Protocol prototype
- Step 3: Validating the EDF Protocol prototype standards
  - 7 semi-structured expert interviews
- Step 4: Adjusting the EDF-protocol prototype and final protocol development
- Step 5: Validation of the final EDF Protocol

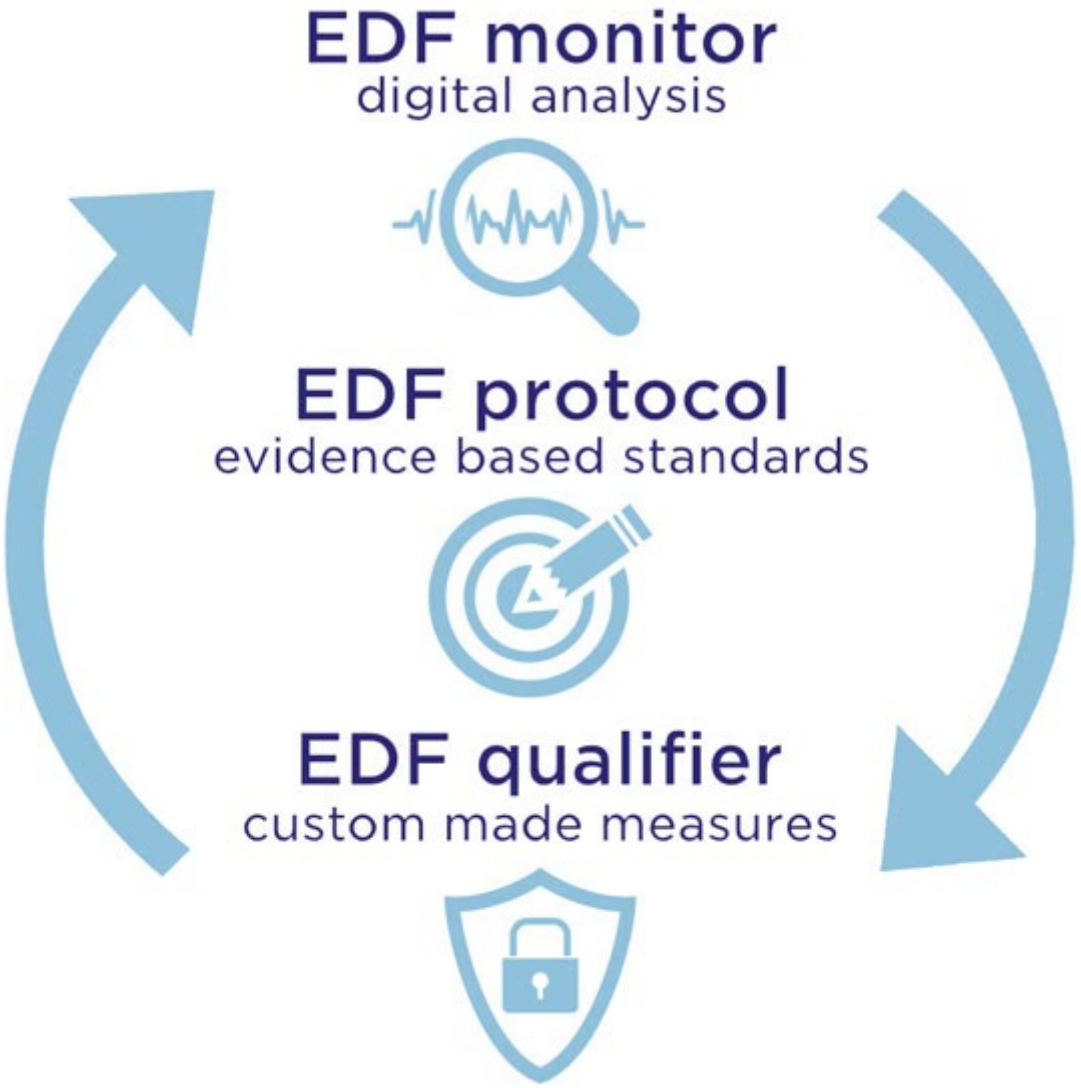
EDF Protocol – Example of one standard

Part A – Standards for Fraud Prevention

Standard 1: Security plan

	Insufficient (0)	Sufficient (1)	Good (2)	Score
Security Plan	Security practices exist without a formal security plan	Security plan exists as an internal document, approved by the management	The security plan is available to all involved personnel	
Security Goals	Provides minimal direction and oversight on security issues	A mission statement on security goals is present Goals include at least: an aim towards preventing disclosure of exam content as much as possible	A clear mission statement on security is present and integrated with practice	
Security Policy	Policy governing security efforts is limited to general statements that may be challenging to translate into measures	Policy governing security efforts provide adequate directions for security measures Policy includes at least: Everyone who has access to the content of the exam signed an agreement which prohibits the disclosure of exam content	Policy governing security efforts provide effective directions with sufficient clarity to ensure appropriate implementation	
Actuality	The security plan has not been reviewed / revised within the past 24 months	The security plan is reviewed/revised within the past 12 months	The security plan is reviewed/revised within the past 12 months, and is discussed internally in the past 12 months	
Financial Resources	There is no sufficient budget to be able to implement the security plan and/or to solve security incidents	The security costs are included in the budget for maintenance and development of the exam. Budget are in accordance with the security plan	The budget is checked according to a yearly set timetable and adjusted if necessary	
	Total score on standard:			

Determining security risk for Standard 1	
The total score on this standard is '10'	Low security risk
The total score on this standard is '5' or 'higher', without an 'insufficient' score ***Advise***: Although all criteria score at least 'sufficient', it is advised to improve your security measures to meet 'Good' rubric descriptions, to reduce the security risk	Medium security risk
One or more 'Insufficient' score(s) on one of the criteria ***Advise***: Direct your resources towards the criteria with the 'Insufficient' score, as it forms a high security risk for your exam	High security risk


EDF monitor
digital analysis
EDF protocol
evidence based standards
EDF qualifier
custom made measures

Thank you!

- The EDF protocol can be downloaded from our website for free:

www.xquiry.com

Thank you for your attention! Questions?