Plans for the ITEMS module on test security and a comprehensive framework for test security

Steve Ferrara

Presentation at the annual Conference on Test Security

October 17, 2019

Overview

- Design and contents of the ITEMS module
- A comprehensive framework for test security programs: PDIR

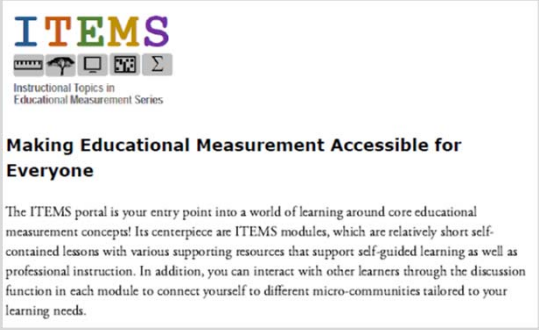
- Acronyms
  - Instructional Topics in Educational Measurement Series
  - Prevention, detection, investigation, and resolution

ITEMS modules

- Short (1-2 hours), self-contained lessons
- Supported self-guided learning
- Easy navigation through modules and lessons

- Reliability, IRT, diagnostic models
- Subscores, operational item analysis, socio-cognitive assessment

- https://www.ncme.org/resources/items



ITEMS module on test security

1. PDIR framework
2. Prevention
3. Investigation and resolution
4. Detection: Forensics methods
5. Year-to-year performance spikes
6. Inordinately similar response strings
7. Wrong-to-right answer changes
8. Another method (TBD)
9. Prevention and forensics for unique item designs
10. Protecting test content via web patrolling
11. Elements of a comprehensive security program

Topic developers

- Michelle Croft, ACT
- Steve Ferrara, Cognia
- Will Lorie, independent consultant
- Frank Padellaro, Cognia
- Sarah Toton, Caveon
- Xi Wang, Cognia
- Cengiz Zopluoglu, University of Miami
- Others, TBD
- Others who want to contribute are invited to propose topics!

PDIR framework

- Prevention
  - Threats, countermeasures, roles and responsibilities
- Detection
  - Empirically supported and promising detection methods (statistical and other), including evidence of efficacy and effectiveness

Table 4. *(Continued)*

Supported or Promising Detection Methods[b]	Research Findings on Efficacy or Effectiveness[c]
(15) Person fit indices	Zopluoglu (2017; table 2.1) lists person fit indices and concludes that, although person fit indices generally have limited use in detecting answer copying, H^T and $D_{(6)}$ appear to perform well enough in some contexts (p. 33); a review by Karabatsos supported H^T and *U3* (cited in Kim, Woo, & Dickison, 2017, p. 77), who also support the l_z and l_z^* indices (p. 73)
Caveats for use: Type I error rates and power always must be considered; most evaluations are based on data simulations (Zopluoglu, 2017, p. 33), rather than real test data; person fit flags alone is not evidence of item preknowledge (Eckerly, 2017, p. 104)	
(16) Clustering algorithms to detect group collusion (Maynes, 2013, p. 190), deterministic gated IRT model (see Eckerly, 2017), hierarchical growth models, and factor analysis (Maynes, 2013, p. 196)	Analysis of group based collusion has "potential to be…powerful" but "much work is needed to understand the nature of group-based collusion, how to measure it, and how to investigate it" (Maynes, 2013, p. 191)
Caveats for use: Cluster analysis looks "quite promising" though more research is indicated (Wollack & Maynes, 2017, pp. 147, 148)	

Ferrara, 2017

PDIR framework: Investigation

- Who should conduct them?
- Who is well suited and qualified to conduct them?
- *School Cheating Thrives While Investigations Languish* (Judd, 2012)

PDIR framework: Resolution

- Difficult, complicated set of issues
  - Staff sanctions
  - Recouping the value of lost assets
  - Re-releasing test results
  - Calibrating sanctions to the extent and violation of the security violation
- Limited evidence on following up says
  - We're not completing the process (e.g., Office of the Inspector General, 2014; Ferrara, 2017)
- Peer review: *Remediation*

PDIR framework

- What do we know about state assessment program test security programs and practices?
  - Information is not widely shared
- Why propose a comprehensive framework?
- Evidence on the structure, content, and operation of state test security programs is scattered and often inaccessible
  - Proposals to remediate that…

Thanks!

Steve Ferrara
Ferrara.steve@cognia.org

References

Ferrara, S. (2017). A comprehensive framework for policies and practices to improve test security programs: Prevention, detection, investigation, and resolution (PDIR*). Educational Measurement: Issues and Practice*, *36*(3), 5-23.

Judd, A. (2012, September 22). School cheating thrives while investigations languish. *Atlanta Journal-Constitution*.

Office of the Inspector General. (2014). *Final audit report*. Final Audit Report ED-OIG/A07M0001. Retrieved April 11, 2017 from http://www2.ed.gov/about/offices/list/oig/auditreports/fy2014/a07m0001.pdf