



Policies and procedures for the Independent evaluation of educational assessment programs

Andrew Wiley

Paper presented at the annual meeting at the
National Council on Measurement in Education
April, 2015

Three categories to describe

- ▶ Pre review activities
- ▶ Essential program areas
- ▶ Themes for findings



Pre-review activities

- ▶ Purpose
- ▶ External Sources
- ▶ Outcome
- ▶ Intended audiences
- ▶ Authorship



Pre review - Purpose

- Formal review required for a program accreditation
- Contractual requirements with certain clients
- Internal process improvement



Pre review - External Sources

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association.

Association of Test Publishers (2001). Operational Best practice for Statewide large Scale Assessments, DC: Author.

International Test Commission. (2001). ITC test adaptation guidelines. Retrieved July 12, 2004 from http://www.intestcom.org/test_adapt.htm.

Society for Industrial and Organizational Psychology (2003). Principles for the validation and use of personnel selection procedures. College Park, MD: Author.

Pre review - Outcome

- ▶ Written reports
- ▶ Presentations
- ▶ Project plans to address any concerns that were raised
- ▶ Implementation assistance



Pre review - Intended Audience

- ▶ Technical or research staff
- ▶ Operational staff responsible for implementation
- ▶ Executive team/Leadership
- ▶ Members of the public
- ▶ Issue of confidentiality



Pre review - Authorship

- ▶ Two conflicting perspectives
 - ▶ The organization that has been audited would like the freedom to publicize any results (positive of course) that can demonstrate the utility of their program
 - ▶ The auditor views their reputation as something that can not be compromised, and cannot allow their words to be misrepresented to support a program
- ▶ Essential that rules and regulations for the release and use of the report are specified prior to the beginning of the review

Program areas

- ▶ Purpose
- ▶ Developmental processes
- ▶ Technical information
- ▶ Validity
- ▶ Reporting



Program Areas - Purpose

- ▶ Standard 1.1, requires that “the test developer should set forth clearly how test scores are intended to be interpreted” (p. 23).
- ▶ Have the programs gone through the process of connecting their purpose or mission statement to each critical decision point?
- ▶ Mission creep can occur over years.



Program Areas - Developmental Processes

- ▶ Standard 4.0 states that test developers “should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses” (p. 85).
- ▶ Should also include information regarding the people (SMEs) involved in the development process and how well they represent the intended population for the assessment



Program Areas - Technical Information

Scoring

- ▶ Standard 6.9 requires that test sponsors document quality control procedures and the training that test scores are provided.
- ▶ Standard 3.5 requires that test sponsors work to ensure that during the test administration there are no barriers to student performance across different student subgroups.
- ▶ Procedures developed to ensure the reliability of the scores assigned to the performance tasks

Program Areas - Technical Information

Equating/Scaling

- ▶ Match between the data available and the methodology used
- ▶ Replication of equating results
- ▶ Quality control procedures for the transfer of data files
- ▶ Quality control procedures for the implementation of equating results

Program Areas - Technical Information

Standard Setting

- ▶ Match between the intended purpose of the test scores and the methodology used
- ▶ Documentation for the implementation of the standard setting and the data and information collected at each step
- ▶ Documentation of the population of standard setting panelists and how well they represented the intended users of the assessment



Program Areas - Validity

- ▶ Almost all programs have some statements on the purpose and use of the test scores
- ▶ Almost all programs have some data and evidence to support the validity of their assessment
 - ▶ Documentation of TD procedures
 - ▶ Correlation with external measures
- ▶ Does the program have an overall validation framework that connects the evidence to the purpose of the test scores?



Program Areas - Reporting

- ▶ Has sufficient information been included in the score reports to aid in the appropriate interpretation of the test scores?
- ▶ Has the organization ever completed any work to evaluate how well their score reports are understood?
- ▶ Has the organization consider the full range of their test users and whether the reports provided are appropriate for each intended user?

Themes

- ▶ Alignment
- ▶ Documentation
- ▶ Communication
- ▶ External contractors
- ▶ Test Security



Themes - Alignment

- ▶ Mission creep
- ▶ Unintended or unknown consequences



Themes - Documentation

- ▶ More consistent and thorough documentation
- ▶ One critical feature is providing information that highlights how appropriate documentation
 - ▶ Requires time and commitment on the part of employees AND their boss
 - ▶ Can actually be a time saver in the 2nd round, and is critical to the long term flexibility of the organization



Themes - Communication

- ▶ Clear communication protocols between employees
- ▶ Clear communication of mission and goals from executive staff to the rest of the organization
- ▶ Understanding who is responsible for communication lines and which groups/divisions need to perform sign-off
- ▶ Over-communication; respecting everyone's Inbox

Themes - External contractors

- ▶ All too easy to become over-reliant upon the external contractors
- ▶ Seems to be an even greater occurrence in some of the more technical areas (i.e. equating/scaling)
- ▶ Need to communicate how the lack of knowledge can harm the long-term flexibility of the organization



Themes - Test Security

- ▶ Procedures focused on the prevention of test security breaches
 - ▶ Role of external vendors
 - ▶ Proper communication to testing centers
- ▶ Detection of security breaches
- ▶ Enforcement
 - ▶ Does the organization have a coherent action plan for when a security breach is detected.
 - ▶ Are roles and responsibilities clearly defined
 - ▶ Are means of communication defined, especially for emergency breaches that occur during off hours

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Association of Test Publishers (2001). *Operational Best practice for Statewide large Scale Assessments*, DC: Author.

Brennan, R. L. (2006). *Educational Measurement*, 4th ed. Westport, CT: American Council on Education and Praeger.

Buckendahl, C. W. & Plake, B. S. (2006). Evaluating tests. In S. M. Downing & T. M. Haladyna (eds.) *Handbook of Test Development*. pp. 725-738.

Downing, S. M. & Haladyna, T. M. (1996). A model for evaluating high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practices*, 15(1), 5-12.

Haladyna, T. M. (2002). Supporting documentation: Assuring more valid test score interpretations and uses. In G. Tindal & T. Haladyna (Eds.), *Large-scale assessment programs for all students: Development, implementation, and analysis* (pp. 89-198). Mahwah, NJ: Lawrence Erlbaum Associates.

Thank you!

- ▶ Questions/comments?
- ▶ Awiley999@gmail.com

