

Auditing Statewide Assessment Systems: Lessons Learned From Two States

Chris Brandt & Matt Gaertner WestEd Standards Assessment and Accountability Services Program



Overview of Presentations

- Steps to Consider When Implementing a Statewide Audit
- Considerations for Assessing Technical Quality
- Experiences and Lessons Learned
 - Oregon
 - Nevada



Why Audit a Statewide Assessment System?

- Eliminate Redundancies/Improve Test Efficiency
- Address Gaps in Assessment Needs (e.g., assessment literacy)
- Improve Accountability and Opportunities for Learning
- Vertical Coherence
- Improve Individual Test Quality and Alignment
- Support Peer Review



Planning for the Audit

- Clear Purposes
- Questions to Be Addressed
- Alignment to a Theory of Action
- Point Persons (District and School Champions)



Questions

- What is the Current State?
- Does the System Support Evidence-Based Practice?
 - As Designed
 - As Implemented
- What are the Pain Points and Root Causes?
- What Potential Solutions Will Improve the System?
- How Will We Roll Out, Test, Refine, and Scale Improvements?



Data Collection

- Objective Reality
 - District and School Data Inventories (Formative, Interim, Summative)
 - Classroom Observations
 - Computer Records (Testing time, Usage, etc.)
 - Technical Manuals and Administration Manuals
- Perceived Reality
 - Surveys
 - Interviews
 - Focus Groups



Data Analysis

- Individual Assessment Quality
 - Validity and Reliability
 - Alignment and Coverage
 - Accessibility
 - Usefulness in Relation to the System of Assessments
- Systemic Balance
 - Redundancies and Gaps
 - Pain Points
 - Perception vs. Reality



Recommendations

- Prioritize
 - Quick wins
 - Consider cost, time, ease of implementation, and relative impact
- Resources and Capacity to Support Change
- Locus of Control (State vs. District)



Targeted Reporting

- Who is the Audience?
- What are the Key Takeaways?
- What are the Implications and Next Steps?
 - Policymakers
 - District and School Leadership
 - Families and Communities



Part 2 Technical Quality



What Is Technical Quality?

SECTION 3: TECHNICAL QUALITY - VALIDITY

Critical Element 3.1 - Overall Validity, Including Validity Based on Content

The State has documented adequate overall validity evidence for its assessments consistent with nationally recognized professional and technical	SECTION 4: TECHNICAL QUALITY – OTHER Critical Element 4.1 – Reliability	
		Examples of Evidence
 testing standards. The State's validity evidence includes evidence that: <u>The State's academic assessments</u> measure the knowledge and skills specified in the State's academic content standards, including: Documentation of adequate alignment between the State's assessments and the academic content standards the assessments are designed to measure in terms of 	The State has documented adequate reliability evidence for its assessments for the following measures of reliability for the State's student population overall and each student group consistent with nationally recognized professional and technical testing standards. If the State's assessments are implemented in multiple States, measures of reliability for the assessment overall and each student group consistent with nationally recognized professional and technical testing standards, including: • Test reliability of the State's	 Collectively, evidence for the State's general academic assessments, the general ELP assessments, the AA-AAAS and AELPA must document adequate reliability evidence generally consistent with nationally recognized professional and technical testing standards. <i>For ELP assessments</i>, such evidence should also be provided for any domain or component sub-tests, if applicable. Evidence to support this critical element for the State's academic content and ELP assessments includes documentation such as: A chapter on reliability in the technical report for the State's assessments that shows reliability evidence (including state-specific information if assessment is part of a multi-state consortium); Documentation of reliability evidence generally consistent with expectations of current professional standards, including: Results of analyses for alternate-form, test-retest, or internal consistency reliability statistics, as appropriate, for each assessment and for each domain or component sub-test, if applicable; Report of standard errors of measurement and conditional standard errors of measurement, for example, in



What Are Reliability and Validity?





Which types of technical quality evidence should we emphasize?

Relevant evidence supports the claims we intend to make about assessment results



- 3.1 Overall Validity, including Validity Based on Content
- 3.2 Validity Based on Cognitive Processes / Linguistic Processes
- 3.3 Validity Based on Internal Structure
- 3.4 Validity Based on Relations to Other Variables

- 4.1 Reliability
- 4.2 Fairness and Accessibility
- 4.3 Full Performance Continuum
- 4.4 Scoring
- 4.5 Multiple Assessment Forms
- 4.6 Multiple Versions of an Assessment
- 4.7 Technical Analysis and Ongoing Maintenance



- 3.1 Overall Validity, including Validity Based on Content
- 3.2 Validity Based on Cognitive Processes / Linguistic Processes
- 3.3 Validity Based on Internal Structure
- 3.4 Validity Based on Relations to Other Variables

4.1 – Reliability



- 4.5 Multiple Assessment Forms
- 4.6 Multiple Versions of an Assessment
- 4.7 Technical Analysis and Ongoing Maintenance



- 3.1 Overall Validity, including Validity Based on Content
- 3.2 Validity Based on Cognitive Processes / Linguistic Processes
- 3.3 Validity Based on Internal Structure
- 3.4 Validity Based on Relations to Other Variables

• 4.1 – Reliability

• 4.2 – Fairness and Accessibility



4.7 – Technical Analysis and Ongoing Maintenance



Validity

Equating Procedures

Trocesses / Linguistic Frocesse

- 3.3 Validity Based on Internal Structure
- 3.4 Validity Based on Relations to Other Variables

• 4.1 – Reliability

- 4.2 Fairness and Accessibility
- 4.3 Full Performance Continuum
- 4.4 Scoring
- 4.5 Multiple Assessment Forms
- 4.6 Multiple Versions of an Assessment
- 4.7 Technical Analysis and Ongoing Maintenance



- 3.1 Overall Validity, including Validity Based on Content
- Mode Comparability Study

Structure

 3.4 – Validity Based on Relations to Other Variables

- 4.1 Reliability
- 4.2 Fairness and Accessibility
- 4.3 Full Performance Continuum
- 4.4 Scoring
- 4.5 Multiple Assessment Forms
- 4.6 Multiple Versions of an Assessment
- 4.7 Technical Analysis and Ongoing Maintenance



Lessons Learned

Presenting Evidence for Multiple Audiences

WestEd.org

State Legislature

- 1. Rather than implying an assessment may be "invalid" or "unreliable," describe the evidence of technical quality, for example:
 - Well Documented
 - Partially Documented
 - Not Documented
- 2. Constructive recommendations highlight next steps for the assessment program
- 3. Eschew obfuscation



Peer Review

- 1. An assessment audit can serve two purposes (legislative report + peer-review submission)
- 2. Be specific about where the evidence can be found (table, figure, and page numbers)

Pre-trial discovery

- Clear & navigable
- 3. Embrace critical element 4.7!







3.1 – Overall Validity, including Validity
 Based on Content

vanuity based on interna

Validation is an ongoing process

Structure

 3.4 – Validity Based on Relations to Other Variables

- 4.1 Reliability
- 4.2 Fairness and Accessibility
- 4.3 Full Performance Continuum
- 4.4 Scoring
- 4.5 Multiple Assessment Forms
- 4.6 Multiple Versions of an Assessment
- 4.7 Technical Analysis and Ongoing Maintenance