

State Board of Education
CERTIFICATE OF MASTERY STUDY COMMITTEE

DRAFT MINUTES

of

May 1, 2001 Meeting

Members Present: Gary Gainer, Barbara Clausen, Lynn Fielding, Marc Frazer, Greg Hall, Gary Kipp, Bob McMullen, Bill Moore, Steve Mullin, Marv Sather, Sue Shannon, Dennis Wallace, Ron Woldeit

Members Unable To Attend:

Guests: Tom Haladyna, Duncan MacQuarrie, Chris Thompson, Laura Fuhrman, Carolyn Tolas, Paul Dugger, Terri Cassidy,

Staff: Larry Davis, Pat Eirish

▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼

Chairman Gainer called the meeting to order at 3:10^{pm} and introduced the guest presenter, Dr. Thomas M. Haladyna, Professor of Educational Psychology, Arizona State University—West (Phoenix).

Dr. Haladyna's organized his remarks around a handout (Attachment A) that included the following questions and issues:

- What is accountability?
 - What are high-stakes test score uses?
 - What is the value of high-stakes test score use?
 - What content constitutes student learning today?
1. Legal Defensibility and Validity (nine considerations ruled in favor of the State of Texas in G.I. Forum v. Texas)
 2. Opportunity to Learn (short list of five issues)
 3. Activities in Which the State Board Might Engage (five suggestions)

Minutes of COM Study Committee
Meeting of May 1, 2001
Page Two

Points and observations made during the presentation included:

- Evidence relating to validity and reliability should be collected by an outside, neutral, third party who should be charged with evaluating the evidence and making a judgment as to its adequacy. Publish all evidence and put on a web site. There are different categories of evidence, including: statistical evidence, study evidence, procedural evidence, and product evidence.
- Setting/changing cut-scores cannot be separated from validity and reliability. The A+ Commission and the State Board of Education have to have a seamless approach. Cut-scores have nothing to do with testing. They are a value judgment apart from testing. Setting cut-scores is easy.....dealing with the consequences is what is not easy. The consequence of not earning a diploma is big - the correlation between lifetime earnings with and without a diploma is very significant. Consider letting districts decide where to set the cut-score for their students. Nebraska allows districts to choose the test and set the cut-scores.
- The WASL is fine. What needs an examination is systemic reform. If classroom assessment doesn't change and improve, the WASL won't change things. Toolkits should contain test items that teachers can use. Train teachers to score the WASL. Teachers can learn assessment and can and should receive professional development toward that end. Use the Advanced Placement model for professional development.
- How do you monitor if systemic reform is reaching into classrooms? Some factors exist outside schools and affect scores. You can't just look at test scores and make a judgment. The critical question is, "What are you doing?" If you're not testing all the EALRs, you're in trouble.
- Random error is a fact of life in the assessment arena. All standardized test scores are corrupted by certain factors. There are 20-30 sources of systematic error in testing.
 1. Cheating: by students, teachers, administrators (changing scores, reading the answers, extending the test-taking time, etc.)
 2. Excluding students who are expected to score low or fail the test.
 3. Rater effect. Two raters for a writing assessment. Hard rater means it's tougher to get a passing score. Easy rater means it's easier to get a passing score.
 4. Plodders or slow workers (i.e., test takers). May be marked down even if they know all the answers.

**Minutes of COM Study Committee
Meeting of May 1, 2001
Page Three**

- There is a national backlash against graduation testing that needs to be studied further. It will take twenty years, at least, to pull off education reform.

▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼

The suggestion was made to begin developing an Opportunity to Learn Inventory Checklist that would identify for each element: What's in place to meet the "sufficiently reliable and valid" threshold?, What's not in place to meet to meet the "sufficiently reliable and valid" threshold and what is needed to close the gap? Also, identify which elements are the responsibility of the state and which ones are school district responsibilities.

▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼

Following a dinner break, Dr. Haladyna spoke briefly about the Oregon experience. Oregon has not yet solved issues relating to Opportunity to Learn, alternate assessments, or consequences. The Oregon Roundtable has developed a model that costs out for any school or district what it will cost to get students to standard.

▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼

It was agreed there would be no July meeting and members will be polled for meeting dates in August/September and November/January.

▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼ ▼▼▼▼▼

The meeting adjourned at 6:54^{pm}.

ATTACHMENTS

Attachment A - State of Washington Board of Education, Washington Assessment of Student Learning, Dr. Thomas M. Haladyna, Professor of Educational Psychology, Arizona State University—West, 4701 West Thunderbird Road, Phoenix, AZ, 85069-7100, tmh@asu.edu, 602-543-6319.

Attachment B - Excerpts from STANDARDS for educational and psychological testing [American Educational Research Association-AERA, American Psychological Association-APA, National Council on Measurement in Education-NCME]

Attachment C - Draft starting point for Opportunity to Learn inventory checklist

Appendix C

AERA/APA/NCME Standards for Educational Testing	Source	Primary Obligator				
Validity	AERA	Leg	SPI	SBE	A+	Dist
Reliability and Errors of Measurement	AERA					
Test Development and Revision	AERA					
Scores, Norms, and Score Comparability	AERA					
Test Administration, Scoring and Reporting	AERA					
Supporting Documentation for Test	AERA					
Rights and Responsibilities of Test Takers	AERA					
Testing Individuals of Diverse Linguistic Backgrounds	AERA					
Testing Individuals with Disabilities	AERA					
The Responsibilities of Test Users	AERA					
Educational Testing and Assessment	AERA					
Testing in Program Evaluation and Public Policy	AERA					
AERA Position Statement Concerning High-Stakes Testing in Pre K-12 Education	Source	Primary Obligator				
Protection Against High-Stakes Decisions Based On A Single Test	AERA	Leg	SPI	SBE	A+	Dist
Adequate Resources and Opportunity to Learn	AERA					
Validation for Each Separate Intended Use	AERA					
Full Disclosure of Likely Negative Consequences of High-Stakes Testing Programs	AERA					
Validity of Passing Scores and Achievement Levels	AERA					
Appropriate Attention to Language Differences Among Examinees	AERA					
Appropriate Attention to Students with Disabilities	AERA					
Careful Adherence to Explicit Rules for Determining Which Students Are to be Tested	AERA					
Sufficient Reliability for Each Intended Use	AERA					
Ongoing Evaluation of Intended and Unintended Effects of High-Stakes Testing	AERA					
Standards Relating to System Readiness - Opportunity To Learn	Source	Primary Obligator				
Adequate Resources Aligned to the SLGs and EALRs		Leg	SPI	SBE	A+	Dist
Adequate Access to Resources						
Staff Instructional Training Aligned to the SLGs and EALRs						
Staff Assessment Training Aligned to the SLGs and EALRs						
Curriculum Alignment with the SLGs and EALRs						
Adequate Time for Students to Learn the SLGs and EALRs Prior to Testing						
Language Ability of Students Vis-a-Vis the WASL						
"Differently Gifted" Ability of Students Vis-a-Vis the WASL						
Ongoing Evaluation of Intended and Unintended Effects of High-Stakes Testing						
Adequate Public Notice About the SLGs, EALRs, WASL, and COM Graduation Requirement						

Appendix C -- continued

G.I. Forum vs. Texas Key Considerations	Source	Primary Obligator				
Professional Testing Standards	GI Forum	Leg	SPI	SBE	A+	Dist
Validity	GI Forum					
Reliability	GI Forum					
Clear Content Standards	GI Forum					
Opportunity to Learn	GI Forum					
Opportunities for Retesting after Remediation	GI Forum					
Passing Standards	GI Forum					
Adverse Impact	GI Forum					
Technical Issues	GI Forum					
Robert Linn Seven Points	Source	Primary Obligator				
Provide safeguards against selective exclusion of students from assessments. This would reduce distortions such as those found for Title I in the fall-spring testing cycle. One way of doing this is to include all students in accountability calculations.	Robert Linn	Leg	SPI	SBE	A+	Dist
Make the case that high-stakes accountability requires new high-quality assessments each year that are equated to those of previous years. Getting by on the cheap will likely lead to both distorted results (e.g., inflated, non-generalizable gains) and distortions in education (e.g., the narrow teaching to the test).	Robert Linn					
Don't put all of the weight on a single test. Instead, seek multiple indicators. The choice of construct matters and the use of multiple indicators increases the validity of inferences based upon observed gains in achievement.	Robert Linn					
Place more emphasis on comparisons of performance from year to year than from school to school. This allows for differences in starting points while maintaining an expectation of improvement for all.	Robert Linn					
Consider both value added and status in the system. Value added provides schools that start out far from the mark a reasonable chance to show improvement while status guards against "institutionalizing low expectations" for those same students and schools.	Robert Linn					
Recognize, evaluate, and report the degree of uncertainty in the reported results.	Robert Linn					
Put in place a system for evaluating both the intended positive effects and the more likely unintended negative effects of the system.	Robert Linn					