Oregon Mathematics Standards Verification Technical Report

Grades 3-8 and High School

October 8, 2010

Prepared for the Oregon Department of Education
by the Educational Policy Improvement Center

epic
Educational Policy
Improvement Center

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

LIST OF APPENDICES

PREFACE (February 9, 2011)

Please note that the following Oregon Mathematics Standards Verification Technical Report was prepared on October 8, 2010 before the Oregon Department of Education (ODE) presented the recommended cut scores to the State Board of Education for adoption. Since the report's publication, the State Board of Education adopted the following cut scores on October 28, 2010:

	Cut Scores		
Grade	**Nearly Meets**	**Meets**	**Exceeds**
3	**205**	**212**	**219**
4	**212**	**219**	**227**
5	**219**	**225**	**234**
6	**222**	**227**	**237**
7	**228**	**232**	**242**
8	**230**	**234**	**245**
HS	**232**	**236**	**251**

Following adoption by the State Board of Education, these cut scores went into retroactive effect for all OAKS Mathematics assessments administered during the 2010-2011 test window.

EXECUTIVE SUMMARY

2010 Mathematics Achievement Standards Verification

The Oregon Department of Education (ODE) conducted an achievement standards verification process for the Oregon Assessment of Knowledge and Skills for Mathematics on August 9-11, 2010. The research-based Bookmark Procedure was used to recommend achievement standards (cut scores) for Grades 3–8 and high school. ODE contracted with the Educational Policy Improvement Center (EPIC) to serve as external evaluators who monitored the process and documented evidence of validity for both the process and the results. This document summarizes the resulting recommendations and EPIC's evaluation results.

The recommended cut scores will be reviewed through a public process in September. The State Board of Education will then consider the cut scores on October 28, 2010. Oregon Assessment of Knowledge and Skills (OAKS) in mathematics will be aligned to the new 2007/2009 content standards.

Why is ODE setting new cut scores (achievement standards) for Mathematics? Oregonians have legitimate concerns about changing cut scores now. Concerns include recent budget reductions, having resources needed to raise the rigor for all students, anticipating 2014-2015 assessment change to address the Common Core State Standards, and the impact that new cut scores and raising AYP targets could have on schools' federal ratings. On balance ODE will move forward with the request that the State Board of Education adopt the new cut scores because:

- Local Instruction - Math teachers are basing their instruction on the content standards adopted in 2007 and 2009 and the assessment needs to follow. In 2009 the majority of

districts responding to ODE's survey said they based 2009-10 mathematics instruction on the new content standards.

- Predictability - School districts are concerned that success at elementary and middle school does not closely align to success at high school. Students and parents may have a false impression of a student's math achievement based on current standards. Only 50% of students who met the existing 3rd grade standard were likely to be ready to meet the high school standard. The proposed cut scores will help ensure students who meet the standard at each grade level are on a path to success in meeting the high school cut scores.
- Federal Requirements - Federal law requires that Achievement Standards align with Content Standards. Given the scope of the 2007/2009 change in content standards it is very unlikely that the US Department of Education would approve ODE to assess these new content standards without an achievement standard setting review process.
- Preparation - A new common assessment of the Common Core State Standards will be ready in 2014-15. Without the proposed cut score changes, Oregon students would continue to fall behind their peers in preparing for the new assessment.
- National Competitiveness – States around the country are taking a hard look at what their students need to compete in the 21st century job market. For the sake of our students, and our future economy, we strive for higher standards and work to ensure that Oregon students are graduating ready to compete with their peers nationally and internationally. Based on ODE's analysis of national data (National Assessment of Educational Progress or NAEP), the proposed cut scores are consistent with the more rigorous achievement standards of other states in the nation and will place Oregon close to Washington State's achievement standards. This will make the transition to the Common Assessment achievement standards, scheduled in four years, more seamless for Oregon. A level consistent with the "NAEP Proficient" level is a good predictor for where the 2014-2015 Common Achievement Standards will likely land.

The extended assessment cut scores are also changing. On Monday, August 16, 2010 eighteen math content specialists and special education teachers from across Oregon's K-12 public school system met with ODE staff and members of the Behavioral Research and Teaching research group from the University of Oregon to set alternate achievement standards for students taking Oregon's alternate assessment the Extended Assessment. Oregon's Extended Assessment is developed for students with the most significant cognitive disabilities and, in parallel to the OAKS general assessment, underwent recent changes to the mathematics assessment to ensure alignment with the new 2007/2009 content standards. The panel's recommendations will be submitted to the State Board of Education for approval this fall. The recommendations are available on the ODE website at http://www.ode.state.or.us/search/results/?id=223

How were recommended achievement standards identified? Nearly fifty Oregon educators, parents, and business representatives met with ODE staff August 9 – 11 to review the math tests and establish cut scores. Their charge was to make recommendations to the State Board of Education for new mathematics achievement standards. Panelists included teachers and administrators from K-12 public schools, community colleges, colleges and universities, parents, and business and industry representatives from fields requiring strong math skills.

Participants were recruited from across Oregon to participate in grade-band groups at grades 3-4, 5-6, 7-8, and high school. Within each group ODE divided participants into two tables that were balanced in terms of relevant demographic characteristics (e.g., gender, geographic location). Participants used booklets that contained approximately 70 secure test items arranged from least to most difficult to verify the knowledge and skills that students should demonstrate at each assessed grade level. The current cut scores, as well as achievement standards for other states, and national and international assessments, such as the National Assessment of Educational Progress (NAEP) and Programme for International Student Assessment (PISA), were marked in the booklets. In addition, each booklet included a projected cut score for that grade level. These projected scores were derived by analyzing longitudinal student progression from grade 3 to the required high school "Meets" score. By working back from an international standard (PISA average for the Organization for Economic Cooperation and Development) grade by grade, ODE assessment staff was able to project the score level that would most likely result in the student meeting the high school standard.

Panelists participated in three review rounds in which they individually recommended three cut scores (*Nearly Meets, Meets,* and *Exceeds*) that defined four performance levels: *Does Not Yet Meet, Nearly Meets, Meets,* and *Exceeds* for grades 4, 6, 8, and high school. At the end of Round Three, each grade-band group also submitted group consensus cut scores for their target grade, which were then used to derive the cut score for the adjacent grade. Subsequently, adjacent grade level standards were derived for grades 3, 5, and 7 and reviewed by each grade-band panel as well as the whole group. All derived achievement standards were confirmed through panels' review of the Ordered Item Booklets. They also considered impact data, an analysis which forecasts the potential percentages of students meeting, not meeting, and exceeding standards at each grade based on prior year's test results.

What achievement standards did the group recommend for 2010-11 and beyond?
Table 1 summarizes the cut scores and associated impact data for the four target grade levels based on the final round of discussion and voting, the analysis of the impact data, and the cross-grade articulation discussion by the full panel. Participants reviewed these data at the workshop; impact data are based on 2009-10 test administration. Table 2 runs through the impact data, that is the percentage of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment.

Table 1. Participant-recommended Mathematics Cut Scores and Associated Impact Data for Target Grades*

	Cut Scores			Impact Data**			
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds
4	212	**219**	227	20%	33%	**27%**	19%
6	222	**227**	237	30%	23%	**29%**	18%
8	230	**234**	245	27%	19%	**37%**	16%

HS	232	**237**	251	30%	21%	**43%**	6%

* Percent totals may not equal 100 due to rounding.

**Impact data indicate percentage of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment.

Derived Cut Scores

Once all grade-band panels completed Round Three, ODE assessment staff derived the cut scores for the adjacent grades (Grades 3, 5, and 7) using linear interpolation (and extrapolation) of normative data. This method was used by CTB/McGraw-Hill in establishing the 2007 achievement standards. Table 2 shows the derived cut scores and impact data for Grades 3, 5, and 7 along with the target grade data.

Cross-grade Articulation (Smoothing)

The cut scores and associated impact data determined for the adjacent grades by interpolation were presented to the participants during the cross-grade articulation, or "smoothing," discussion on Day 3. The purpose of this smoothing discussion was to establish a set of cut scores that was well-articulated and, at the same time, considerate of the participants' original recommendations. As participants reviewed the derived scores and impact data, each grade-band panel and the group as a whole gave careful consideration to the final recommended scores. Tony Alpert, Director of Assessment, was present during these discussions to answer policy-related questions.

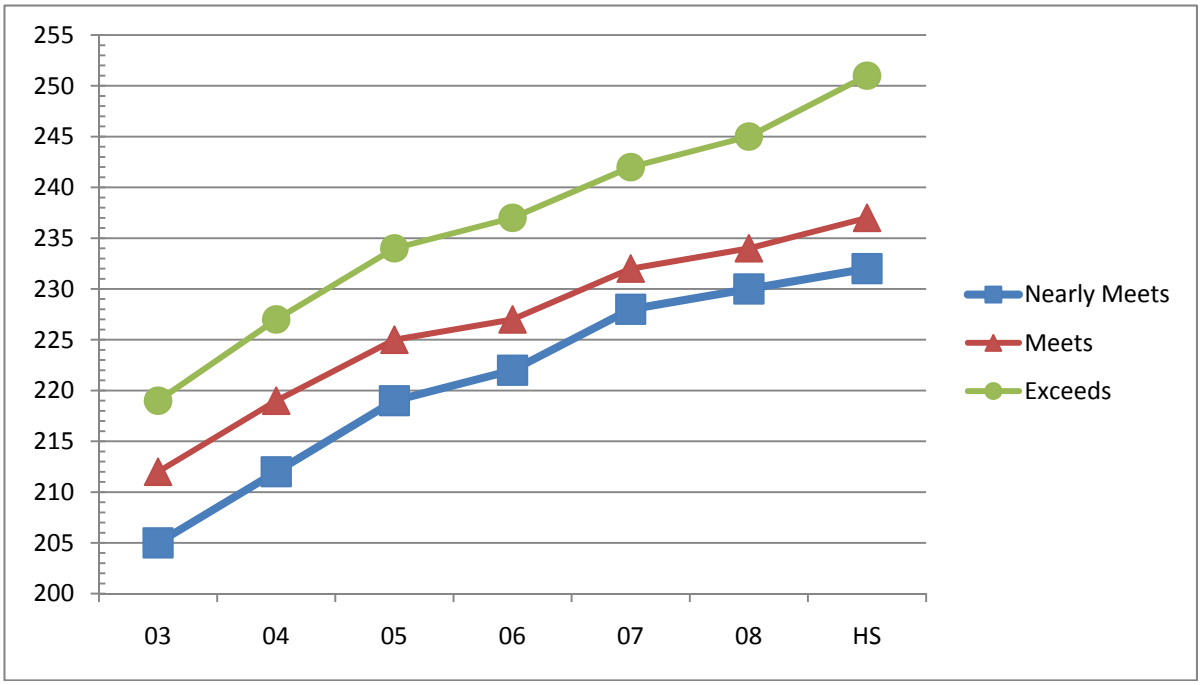
Table 2. Recommended and Derived Cut Scores and Impact Data for All Grades Showing Cross-Grade Articulation.

	Cut Scores			Impact Data**				
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds	Meets & Above
3*	205	212	219	25%	28%	29%	18%	47%
4	**212**	**219**	**227**	**20%**	**33%**	**27%**	**19%**	**46%**
5*	219	225	234	21%	30%	34%	16%	50%
6	**222**	**227**	**237**	**30%**	**23%**	**29%**	**18%**	**47%**
7*	228	232	242	29%	19%	34%	18%	52%
8	**230**	**234**	**245**	**27%**	**19%**	**37%**	**16%**	**53%**
HS	**232**	**237**	**251**	**30%**	**21%**	**43%**	**6%**	**49%**

*Derived data confirmed by Oregon panelists reviewing Ordered Item Booklets.
* *Impact data indicate percentage of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment

Summary:
Figure 1 below displays the recommended cut scores at the *Nearly Meets*, *Meets,* and *Exceeds* levels from grade 3 through high school.

Figure 1. Cross-Grade Progression of Recommended Cut Scores



The State Board of Education will consider adoption of the cut scores on October 28, 2010.

1. INTRODUCTION

In August 2010, staff from the Oregon Department of Education (ODE) conducted the Oregon Assessment of Knowledge and Skills (OAKS) mathematics standards setting using the bookmark procedure (Cizek & Bunch, 2007, Kane, 1994, Mitzel, Lewis, Patz, & Green, 2001).

A modification to standard bookmarking practice included the provision of non-test item data describing where Oregon's current mathematics achievement standards fall in relation to other states and countries for panelists to consider while reviewing and setting the new cut scores. Participants were provided reference cut points from studies linking the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA)to state standards (Bandeira de Mello, Blankenship, & McLaughlin, 2009, Phillips, 2009). Additional reference data were obtained by embedding items from the PISA in the OAKS, embedding released NAEP items in the OAKS, and through analysis of postsecondary student performance data obtained through ODE's partnership with Oregon's University System (OUS). ODE analyzed and integrated this data to provide panelists with projected cut scores, including Oregon's current cut scores, and the cut scores for other states and countries.

This information was provided for panelists to consider while they applied their expertise to determine what Oregon students should be able to know and do in terms of the content measured by the OAKS in mathematics at each grade level. The projected cut scores would, if adopted, raise expectations significantly and would place Oregon in the top five states in the nation in terms of NAEP equivalent cut scores for meeting mathematics standards in grades 4 and 8.

Cut scores were also determined by analyzing the longitudinal student progression from grade 3 to the required high school "Meets" score with the purpose of improving the ability to predict student success in high school and college mathematics based on OAKS performance in earlier grades. By working backwards grade by grade, ODE assessment staff were able to project the score level in each grade that would most likely result in the student meeting the high school standard with an average probability of around 0.6. Table 3 provides the projected cut scores.

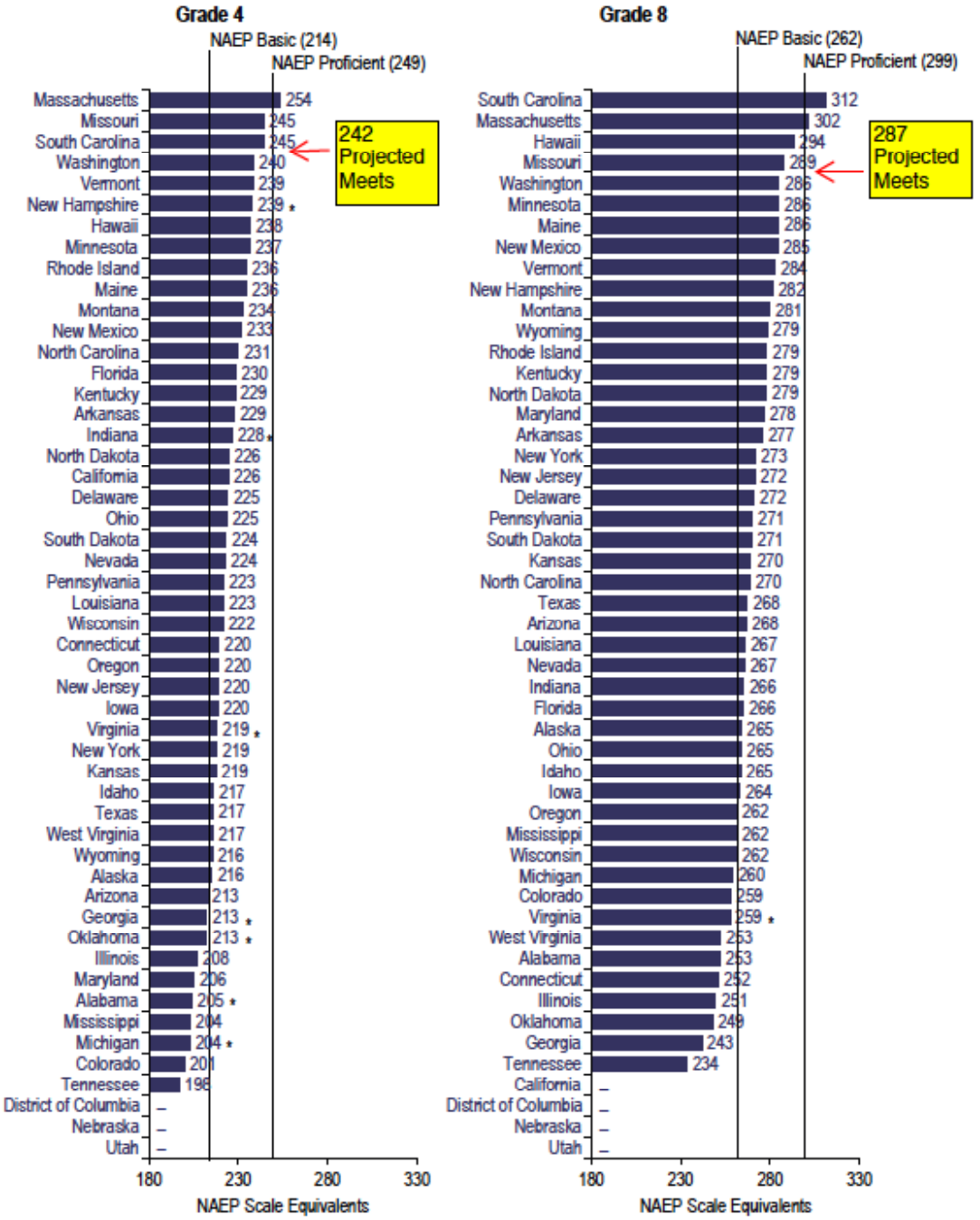
Table 3. NAEP Scale Scores at Projected Oregon Cut Scores

	Oregon Scale Score			NAEP Scale Score	
Achievement Level	Grade 4	Grade 8		Grade 4	Grade 8
Nearly Meets	215	231		232	274
Meets	218	234		242	287
Exceeds	227	243		264	313

Note: Projections made Prior to Standards Verification Meeting.

A graphic representation follows in figure 2, describing where the projected standards fall in relation to other states' standards on the NAEP scale (Bandeira de Mello, Blankenship, & McLaughlin, 2009).

Figure 2. NAEP Scale Equivalent Scores For The State Grades 4 And 8 Mathematics Standards For Proficient Performance, By State: 2007.



— State assessment data not available.
* Relative error greater than .5.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

The projected cut scores demonstrate ODE's commitment that Oregon students be taught and held to rigorous and high standards, ensuring preparation in an increasingly competitive world marketplace. The projections place Oregon at a similar level of expectations for student proficiency in mathematics as South Carolina and Washington (Oregon would be 4th highest in the nation) in grade 4 and between Missouri and Washington (Oregon would be 5th highest in the nation) in grade 8 (Bandeira de Mello, Blankenship, & McLaughlin, 2009). The projections also increase where Oregon's standards fall on the NAEP scale, moving up to just below the NAEP Proficiency standard (249) from just above the Basic standard (214). Note that currently only a single state (Massachusetts) rests above the NAEP Proficiency standard.

ODE sought stakeholder review of the current and projected achievement standards following changes to Oregon's mathematics content standards in 2007/2009 and in anticipation of participating in the upcoming common assessment of the Common Core State Standards. Reviewing the achievement standards was not only likely to increase expectations (and thus preparedness) for Oregon students but also to better position Oregon for the upcoming common assessment, anticipated in 2014-15, making the transition to the Common Assessment achievement standards more seamless. The 2014-2015 Oregon Achievement Standards will likely be set at or near the "NAEP Proficient" level. Thus, with more rigorous cut scores, Oregon students will be on par with or ahead of their peers in preparing for the new assessment. Until the new assessment is in place, the projected cut scores are also better predictors of success in high school and college and will help to ensure that students who meet the standard at each grade level are on a path to success in meeting high school cut scores.

To set the bookmarks, ODE recruited a diverse set of panelists. Participants were recruited from across the state of Oregon to represent stakeholder characteristics and to provide a range of educational and mathematical expertise. They were split into grade level groups and table teams within those groups. They then participated in four rounds of bookmarking and set three achievement standards defining four Achievement Levels, *Does Not Yet Meet, Nearly Meets, Meets,* and *Exceeds*, for grades 3-8 and high school. To a large extent, panelists confirmed and validated ODE's projected scores, in some cases recommending slightly higher standards than those projected.

The final recommendations from the panel are described in Table 4, which summarizes the standards recommended by the panel and the associated impact data.

Table 4. Recommended Cut Scores and Impact Data for All Grades Showing Cross-Grade Articulation.

	Cut Scores			Impact Data**				
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds	Meets & Above
3*	205	212	219	25%	28%	29%	18%	47%
4	**212**	**219**	**227**	**20%**	**33%**	**27%**	**19%**	**46%**
5*	219	225	234	21%	30%	34%	16%	50%
6	**222**	**227**	**237**	**30%**	**23%**	**29%**	**18%**	**47%**
7*	228	232	242	29%	19%	34%	18%	52%
8	**230**	**234**	**245**	**27%**	**19%**	**37%**	**16%**	**53%**
HS	**232**	**237**	**251**	**30%**	**21%**	**43%**	**6%**	**49%**

*Derived data confirmed by Oregon panelists reviewing Ordered Item Booklets.
* *Impact data indicate % of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment

Table 5. Change to Cut Score (+/- Resulting from Recommended Minus Current Cut Score)

	Grade 3	**Grade 4**	Grade 5	**Grade 6**	Grade 7	**Grade 8**	HS
Nearly Meets	+4	**+4**	+5	**+6**	+7	**+5**	**+1**
Meets	+7	**+7**	+7	**+6**	+6	**+4**	**+1**
Exceeds	+2	**+2**	+5	**+5**	+4	**+4**	**+5**

2. OVERVIEW

2.1. Oregon's Assessment System

Oregon's Statewide Assessment System, the Oregon Assessment of Knowledge and Skills (OAKS) provides instructionally useful information to educators about student mastery of the knowledge and skills described by the content standards. The OAKS is an online computer-adaptive test (CAT) designed to measure the grade-specific content described in the standards. Oregon's assessment is the first and only CAT in the nation to be approved by the Department of Education through the peer review process used for determining AYP and meeting NCLB requirements. This distinction firmly identifies Oregon as an innovator in developing and implementing high quality online adaptive NCLB testing programs, as even more stringent technical requirements and evidence of validity must be met for full approval (U.S. Department of Education, 2007)

OAKS assesses knowledge and skills using multiple-choice items and innovative technology enhanced constructed response items that assess higher order thinking skills; all items are aligned to grade-level content standards and are written to represent the state's content standards and the range of student proficiency. The emphasis of the tests matches the emphasis of the content standards such that the tests are representative and valid measures of the knowledge required by Oregon's Academic Content Standards. The content standards are grade leveled against national standards and are designed with stakeholder involvement to be rigorous, coherent, and demanding.

Additional information describing test results, development, and administration can be found in technical reports available for download from the Oregon Department of Education website at http://www.ode.state.or.us/search/page/?id=787.

2.2. Oregon's Mathematics Standards

Oregon's standards system consists of Oregon's Academic Content Standards and Academic Achievement Standards. Content standards define the knowledge and skills Oregon students are expected to demonstrate in each grade. Achievement Standards define four levels of performance (*Does Not Yet Meet, Nearly Meets, Meets,* or *Exceeds*) that students in each grade and content area can demonstrate on the OAKS.

2.2.1. Mathematics Academic Content Standards

All of the state tests are designed to measure the grade-level expectations for what students should know and be able to do as described in Oregon's Academic Content Standards. Oregon's content standards are updated regularly to ensure ongoing comprehension and rigor in content.

The mathematics content standards were most recently revised in 2009 for high school, and 2007 for K-8, and were subsequently adopted by the State Board of Education. The next anticipated revision of the mathematics content standards will occur when Oregon adopts the Common Core State Standards.

Oregon's Academic Content Standards are available on the Web site via the state's Searchable Standards Tool that allows you to locate, view, and export standards by subject,

grade level, and strand (Score Reporting Category (SRC)) at:

http://www.ode.state.or.us/teachlearn/real/standards/

2.2.2. Mathematics Academic Achievement Standards

Achievement standards define, in terms of performance on the OAKS, what students must do to meet or exceed Oregon's Academic Content Standards.

Achievement standards were originally set on September 19, 1996, and the Oregon State Board of Education adopted the achievement standards for grades 3, 5, 8, and high school in reading/literature and mathematics.

Oregon reviewed its achievement levels for all grades in the content areas of mathematics, reading/literature, and science in 2006–07. The State Board of Education reviewed the recommended achievement standards at its meeting on January 18 and 19, 2007 and received regular reports on the feedback from the field review and public input prior to adopting the standards in March 2007. Following adoption by the Board, these achievement levels were applied to all tests administered during the 2006–2007 school year. The current achievement levels for mathematics are provided in table 6.

Table 6. Current Cut Scores Mathematics 2009-2010

	Grade 3	**Grade 4**	Grade 5	**Grade 6**	Grade 7	**Grade 8**	**HS**
Nearly Meets	201	**208**	214	**216**	221	**225**	**231**
Meets	205	**212**	218	**221**	226	**230**	**236**
Exceeds	217	**225**	229	**232**	238	**241**	**246**

Starting with the ninth grade class in the fall of 2008, the State Board of Education required all students to take more rigorous coursework and higher levels of mathematics and science in order to receive a diploma. Additionally, all students were required to demonstrate their abilities in a variety of "essential skills"—initially reading, writing, applying mathematics, and speaking clearly. To meet these higher expectations for Oregon students, ODE reviewed and revised the mathematics standards for all grades and the achievement levels will be officially adopted in October 2010.

2.2.3. Mathematics Achievement Level Descriptors (ALDs)

Oregon's mathematics assessments use four levels of achievement –*Exceeds*, *Meets*, *Nearly Meets*, and *Does Not Yet Meet*. The grade and content-specific descriptors describe the knowledge and skills required by students at each level of performance. The preliminary ALDs are available in Appendix A and on the Department of Education website at http://www.ode.state.or.us/search/results/?id=223. The ALDs recommended by the Panel are on the ODE website at http://www.ode.state.or.us/search/page/?id=3007

Prior to the workshop, ODE drafted preliminary ALDs to include general and mathematics-specific policy definitions. The Policy Definitions provide an overarching definition (across grade) for each achievement level and describe how rigorous and challenging the Achievement Standards (cut scores) will be for the assessments. The general policy definitions are not linked directly to content but are more general statements that describe

rigor across grade levels and content areas.

ODE solicited initial feedback on the preliminary ALDs from members of the mathematics content panels. Panelists work closely with state standards and are familiar with the standards setting process; they are primarily educators with some business and industry partners. Through two surveys conducted in December, ODE received feedback from over 60 panelists from around the state. One survey was for members of the math content panel who were asked about the mathematics ALDs and the Policy Definitions. The second survey was for members of the other content panels (English, Science, Social Sciences, and ELPA) and just dealt with the Policy Definitions. Feedback from both surveys supported the direction of these drafts.

Based on feedback from the field, ODE staff made adjustments and improvements to the Policy Definitions and Mathematics Achievement Level Descriptors. ODE staff will continue to solicit feedback from the field throughout the standards setting process.

Suggested revisions based on the Standards Verification Workshop are provided in Appendix R.

3. THE 2010 MATHEMATICS STANDARDS VERIFICATION WORKSHOP

The ~~projected~~ achievement standards were verified ~~held~~ in August 2010 using a modified bookmarking standard setting procedure (Cizek & Bunch, 2007, Kane, 1994, Mitzel, Lewis, Patz, & Green, 2001). Forty-two Oregonians recommended achievement standards for grades 3-8 and high school in mathematics. ODE mathematics consultants and senior staff developed materials, planned the workshop, conducted the training, and led the participants through the workshop.

ODE contracted with the Educational Policy Improvement Center (EPIC) to review materials and the training process and to evaluate the validity of the recommended achievement standards resulting from the workshop. Expectations for evidence of validity were compiled from best practices prior to the evaluation, including NCLB peer review guidance, and existing standards (APA, AERA, NCME, 2008, Hambleton, 2001, NAGB, 2010, Perie, 2008, U.S. Department of Education, 2007). The extent to which the process met the expectations described for appropriate, high-quality achievement standards is summarized in Table 7.

Table 7. Evidence of Validity Documented During Evaluation

Standard	Evidence
Panels should be large enough and representative enough of the appropriate constituencies.	Grade Level Group Composition, in section 3.2.2.
Selection and qualification of participants should be documented.	Panel Participants, section 3.2.2.
Two panels or subpanels should be used to check the generalizability of the standards.	Grade Level Group Composition, in section 3.2.2; Placing the Bookmarks, section 3.2.4.
Background and demographic information about participants should be collected and documented.	Grade Level Group Composition, in section 3.2.2; Appendix D.
To ensure internal validity, the methods must be consistent so that ratings indicate increased internal consistency across rounds and panelists.	Training, section 3.2.3; Placing the Bookmarks, section 3.2.4; Variability, in section 3.2.4.
To ensure procedural validity, the procedures must be reasonable, carried out as intended and understood by panelists.	The 2010 Mathematics Standards Verification Workshop, Section 3; Training, Section 3.2.3; Placing the Bookmarks in Section 3.2.4; Training Evaluation in Section 3.2.5.; Appendix G.
The methodology should be appropriate for the assessment, described in detail and field tested when appropriate.	The 2010 Mathematics Standards Verification Workshop, Section 3; Derived Cut Scores, in 3.2.4.
Any non-standard methodology must be clearly documented.	The 2010 Mathematics Standards Verification Workshop, Section 3; Mathematics Achievement Standards

	Verification Process, section 3.2.
The precise nature of participants' judgments should be documented, including whether those judgments are of persons, item or test performance, or of other criterion performances predicted by test scores.	Table 22, in section 3.2.5; Placing the Bookmarks in Section 3.2.4; Appendix Q; Appendix T; Target Student Descriptions, in section 3.2.3; Appendices H & I.
The rationale and procedures for establishing cut scores must be documented.	Training, section 3.2.3; Placing the Bookmarks in Section 3.2.4; Table 22, in section 3.2.5; Mathematics Achievement Standards Verification Process, section 3.2; Introduction, section 1.
The methods should be designed so that participants can reasonably contribute their knowledge and experience to produce reasonable, defensible standards.	Training, section 3.2.3; Placing the Bookmarks in Section 3.2.4; Table 22, in section 3.2.5; Mathematics Achievement Standards Verification Process, section 3.2; Introduction, section 1; Appendices L-P.
Participants should be suitably trained on the methodology; training should include a thorough description of the method and practice exercises, practice administration of the assessment, and practice judging task difficulty with feedback on accuracy.	Training, section 3.2.3; Appendix C; Appendix G; Bookmark placement, in section 3.2.3.
Descriptions of performance categories must be clear to the extent that participants are able to use them effectively.	Mathematics Achievement Level Descriptors, section 2.2.3; Achievement level descriptors in section 3.2.3; Process Monitoring and Evaluation, Section 3.2.5; Appendices A, R, G, T.
The process should be conducted efficiently.	Training, section 3.2.3; Placing the Bookmarks, section 3.2.4; Process Monitoring and Evaluation, Section 3.2.5; Appendices G, Q.
Item booklets, rating forms and other provided documents should be easy to use.	Materials review, in section 3.2.3; Process Monitoring and Evaluation, Section 3.2.5; Appendices E, T.
Facilitators should be qualified and capable of leading appropriate discussion among the participants without biasing the process.	Mathematics Consultant Training, in section 3.2.3; Grade Level Group Composition, in section 3.2.2.
Feedback to participants must be clear, understandable, and useful.	Process Monitoring and Evaluation, Section 3.2.5; Appendices G, L-P, T.
Participants should be instructed on the appropriate use of provided data (including performance data, impact data, criterion reference data, etc).	Training, section 3.2.3; Placing the Bookmarks in Section 3.2.4; Table 22, in section 3.2.5; Mathematics Achievement Standards Verification Process, section 3.2; Introduction, section 1; Process Monitoring and Evaluation, Section 3.2.5; Appendices G, L-P, T.
When possible, performance levels should be	Placing the Bookmarks, section 3.2.4; The

established using empirical criterion reference data.	2010 Mathematics Standards Verification Workshop, Section 3; Mathematics Achievement Standards Verification Process, section 3.2.
Process evaluations should be conducted and documented.	Process Monitoring and Evaluation, Section 3.2.5; Appendices G, T.
The entire process must be documented, including participant selection and qualifications, training, feedback to panelists regarding their recommendations, replicability, validity, and variability over participant recommendations.	Panel Participants, section 3.2.2; Grade Level Group Composition, in section 3.2.2; Training, section 3.2.3; Placing the Bookmarks in Section 3.2.4.

The workshop began with orientation, training, and a practice session setting bookmarks. At the conclusion of the first day, participants were asked to complete a training evaluation. The workshop also included three rounds of bookmark placement for grades 4, 6, 8, and high school, which entailed a review of impact data based on assessment results from the 2009-10 academic year and bookmark placement across grade level groups and table teams. The workshop concluded with bookmark placement for grades 3, 5, and 7 and a presentation of the final recommendations and corresponding impact data across all grades. The processes used throughout the workshop are documented in detail below. Additionally, materials used in the workshop are provided in the appendices as noted.

3.1. Goals of the Standards Verification Workshop

The goals of the mathematics achievement standard-setting procedure were as follows:

- Establish what students in each grade (3-8 and high school) should be able to demonstrate on the OAKS in mathematics at each Achievement Level (*Does Not Yet Meet*, *Nearly Meets, Meets*, and *Exceeds*)
- Revise the achievement standards to better prepare students for a competitive international marketplace where students will be competing for jobs with students from states or countries with high expectations
- Ensure that students in earlier grades are held to high standards so they are prepared for even higher standards in later years, never having to "catch up" in later grades
- Consider impact data describing the implications of proposed cut scores in making judgments about item difficulty and the placement of the bookmarks, including national and international contexts
- Provide recommendations to the Oregon State Board of Education on the appropriate cut scores for each Achievement Level

3.2. Mathematics Achievement Standards Verification Process Summary

From August 9 to August 11, 2010, the Oregon Department of Education (ODE) convened a group of educators and stakeholders to participate in the Standards Verification Workshop to recommend achievement standards in mathematics in grades 3-8 and high school on the Oregon Assessment of Knowledge and Skills (OAKS).

Forty-two knowledgeable participants, including educators, higher education representatives, parents, and community members, were recruited from across Oregon to participate in grade-band groups at grades 3-4, 5-6, 7-8, and high school. Using a modified bookmarking procedure (Cizek & Bunch, 2007, Kane, 1994, Mitzel, Lewis, Patz, & Green, 2001) workshop participants completed a sample online mathematics test via OAKS, received training from ODE staff, and completed four rounds of standards setting over three days to determine the *Nearly Meets*, *Meets,* and *Exceeds* cut scores.

Workshop participants participated in one of four grade-band groups (grades 3-4, 5-6, 7- 8, and high school), with two smaller table teams (A and B) in each group. ODE assigned participants to table teams that were balanced in terms of relevant demographic characteristics (e.g., gender, geographic location). Participants used booklets that contained approximately 70 secure test items arranged from least to most difficult to verify the knowledge and skills that students should demonstrate in each assessed grade level. The current cut scores as well as achievement standards for other states and national and international assessments, such as the National Assessment of Educational Progress (NAEP) and Programme for International Student Assessment (PISA), were indicated in the booklets. In addition, each booklet included the ODE projected cut score for that grade level.

Achievement standards were set for one grade at a time. First, participants bookmarked achievement levels for the on-grades (4, 6, 8, and high school). Next, ODE interpolated and extrapolated the standards for grades 3, 5, and 7 from the panelist's on-grade recommendations, and then workshop participants reviewed and revised ODE's derived standards. All derived achievement standards were confirmed through panels' review of the Ordered Item Booklets.

In order to set the on-grade achievement levels, panelists participated in three review rounds in which they individually recommended three cut scores (*Nearly Meets, Meets,* and *Exceeds*) that defined four Achievement Levels: *Does Not Yet Meet, Nearly Meets, Meets,* and *Exceeds.* At the end of Round Three, each grade-band group submitted group consensus cut scores for their target grade. ODE psychometric staff then derived the cut scores for the adjacent off-grades (3, 5, and 7) by analyzing longitudinal student progression from grade 3 to the required high school "Meets" score. This policy model has been previously used successfully by ODE.

The cut scores and associated impact data determined for the adjacent grades by interpolation were presented to the participants during the cross-grade articulation, or "smoothing," discussion on Day 3. The purpose of this smoothing discussion was to establish a system of cut scores that was well articulated and, at the same time, considerate of the participants' original recommendations. All participants reviewed the cross-grade articulation based on the recommended and derived scores. They also considered impact data, an analysis which forecasts the potential percentages of students meeting, not meeting, and exceeding standards at each grade based on prior years' test results. Table 2 above shows the derived cut scores and impact data for Grades 3, 5, and 7 along with the target grade data. As participants reviewed the derived scores and impact data, each grade-band panel and the group as a whole gave careful consideration to the final recommended scores. The

Director of Assessment was present during these discussions to answer policy-related questions, as was the Manager of Psychometrics and Validity to answer technical questions.

Following the workshop, participants completed evaluations that included questions eliciting information about the participants' background and demographics.

3.2.1. Workshop Agenda

During the first day of the training, ODE described to participants the use of assessment scores and the impact of the test scores, cut scores, and the preliminary cut scores determined throughout the verification process. Throughout the training, ODE focused on the goals of the standard setting workshop (see section 3.1 above), emphasizing that one of the goals was to allow Oregon students to be as prepared as students in high performing states and countries. ODE described a linking study conducted to allow for comparisons of Oregon's cut scores to those of other countries (via PISA), the nation (via NAEP), and other states (via other state's NAEP linking studies). The data from the linking study was used demonstrate that Oregon's achievement standards were low compared to other states and countries.

Before reviewing the projected achievement standards, panelists were reminded that high standards are necessary to adequately prepare Oregon students. ODE asked the panelists to provide a rationale for why Oregon should change its cut scores. Panelists responded that if students in earlier grades are not held to high standard, they will lag behind in all subsequent grades, particularly when Oregon adopts the Common Core Standards in four years. Further, Oregon students need to be prepared for an international and national marketplace. Because other states and countries have higher standards, Oregon students are not globally competitive. Community and business leaders provided feedback that high school diplomas should represent that applicants are prepared for employment and engagement in systematic learning.

While ODE did not minimize the impact of raising the achievement standards, it did emphasize that this Standards Verification Workshop was an opportunity to apply expert knowledge to raise standards and expectations in a clear and transparent way. ODE explained that Standards Verification was not an arbitrary discussion, rather it was a systematic process based on expert evaluation of content after in-depth discussion.

The workshop agenda is provided in Appendix B and the training presentations are provided in Appendix C.

3.2.2. Panel Participants

ODE Staff and Mathematics Consultants

Six mathematics consultants were recruited to assist ODE with leading and providing content expertise in the Standards Verification Workshop. These mathematics consultants were external experts who had participated in pre-verification training and assisted with drafting the Achievement Level Descriptors (ALDs).

Standards Verification Workshop Participants

Forty-two Oregonians participated in the Standards Verification Workshop. The panel

was carefully selected to include K-12 educators (79%), community college educators (7%), university educators (5%), business members (5%), and parents (2%). Panels represented the racial makeup of Oregon, which is 90% White (U.S. Census Bureau, 2010). Overall, the panel selected was large and representative of the appropriate constituencies to be judged as suitable for setting achievement standards on the educational assessment (Hambleton, 2001).

The panel composition is described in Appendix D.

Recruitment and Compensation

To recruit workshop participants, the Department solicited involvement from all levels of the education system and from the community. Nominations were solicited from teacher organizations and educator networks including the following:

- Current and past mathematics content panels, mathematics item writers, Oregon Mathematics Education Council (OMEC), Teachers of Teachers of Mathematics (TOTOM), Oregon Council of Teachers of Mathematics (OCTM), The Oregon Mathematics Specialists (teachers on special assignment and curriculum coordinators), and the ODE mathematics instructional contact list.

- Superintendents, Special Education, CTE, and TAG reps at ODE; ELPA and Title III Directors, Oregon School Boards Association

- The Assessment and Accountability Update issue of June 10, 2010

Non-educators in the business and parent communities were recruited via email to the state parent organization and OMEC.

Over 130 individuals expressed interest in participating. From these, the Department selected 42 to represent the needs and demographics of Oregon students, including geographic region, district size, gender, race/ethnicity, educational experience, and role in education or the community.

Participants were provided meals during the workshop, and participants who live more than 70 miles from ODE received reimbursement for travel expenses. Participants who were not employed by their district during the workshop were appointed by ODE as temporary employees and were paid an hourly rate to compensate for their time.

Grade Level Group Composition

The 42 workshop participants were divided into four grade level groups that included a mix of participant characteristics. Each grade level group was divided into two table teams for Rounds One and Two, thereby creating replicate panels to monitor and ensure the consistency of the recommended achievement standards. Each group was assigned two table team leaders, a mathematics consultant, and an ODE representative who facilitated the discussion but had no input in bookmark placement.

Appendix D and the following tables describe panel composition for each grade level group. Note that this information was self-reported on process evaluation forms and

demographic questions were optional.

Table 8. Participant Educational Background by Grade level Group

Grades	N	HSD or GED	Bachelor's	Master's	Doctorate
All	42	0%	17%	74%	10%
3-4	11	0%	27%	73%	0%
5-6	11	0%	0%	82%	18%
7-8	10	0%	10%	80%	10%
HS	10	0%	30%	60%	10%

Table 9 shows the occupation of participants in each grade level group.

Table 9. Participant Occupation by Grade level Group

Grades	N	K-12 educator	Community college educator	University educator	Parent	Community member	Business member	Other
All	42	79%	7%	5%	2%	0%	5%	2%
3-4	11	82%	0%	0%	9%	0%	0%	9%
5-6	11	91%	0%	9%	0%	0%	0%	0%
7-8	10	70%	10%	0%	0%	0%	20%	0%
HS	10	70%	20%	10%	0%	0%	0%	0%

Note. Participants may have self-reported representation in more than one category (*i.e.*, as a business member and community member) or as belonging to another category than that which they were selected to represent (*i.e.*, as a parent instead of community or business member).

Table 10 shows the years of work experience for each grade level group.

Table 10. Years of Work Experience by Grade level Group

Grades	N	1-5	6-10	11-15	16-20	21+
All	42	5%	19%	7%	12%	57%
3-4	11	0%	18%	0%	18%	64%
5-6	11	18%	18%	9%	0%	55%
7-8	10	0%	10%	0%	20%	70%
HS	10	0%	30%	20%	10%	40%

Table 11 shows participants' experience teaching English language learners (ELL), special education (SPED), vocational education (Voc. Ed.), alternative education (Alt. Ed.), and adult education (Ad. Ed).

Table 11. Participant Teaching Experience with Diverse Populations by Grade level Group

Grades	N	SPED	N	ELL	N	Voc. Ed	N	Alt. Ed.	N	Ad. Ed
All	42	17%	42	19%	42	7%	41	12%	42	45%
3-4	11	18%	11	27%	11	0%	11	18%	11	45%
5-6	11	18%	11	18%	11	18%	11	9%	11	45%
7-8	10	10%	10	30%	10	10%	9	22%	10	50%
HS	10	20%	10	0%	10	0%	10	0%	10	40%

Participant Roles and Responsibilities

Workshop participants included the following:

- ODE staff
- Mathematics consultants
- Grade level Group Leads (grades 3-4, 5-6, 7-8, and high school)
- Table Team Leads (A/B)

ODE staff planned and ran the workshop. During the workshop, their responsibilities included training, keeping secure materials secure, monitoring questions for additional clarification, keeping groups on task and on time, and facilitating discussions. ODE staff was also responsible for collecting data sheets from each participant, team, and table.

Mathematics consultants were available throughout the process to clarify content-related questions and to facilitate discussions. They were not expected to have a voice in standards verification decisions but could share their mathematics expertise with panelists and assist table leaders with keeping each table on task.

Table Team Leaders anticipated the questions of panelists, discussed and agreed on explanations, and also suggested additions to the instructions provided to all participants on the first day of training.

Table team leaders led the discussion at each table. Each table team also selected a recorder to record and document the group's decisions in Rounds Two and Three and a table reporter to speak for the group.

Two external evaluators from the Educational Policy Improvement Center were non-participatory observers for the entire process, as was a member of the press.

Key Definitions and Table Norms

Prior to beginning their work, workshop participants engaged in a team building activity to ensure shared understanding of important terms used in the process. Each table team also brainstormed norms and identified rules to follow to facilitate collaboration and efficiency.

Norms for each table team were posted on the wall near each table and remained visible throughout the workshop. As needed, mathematics consultants and ODE staff reminded table teams of the norms agreed upon during day one. During process evaluation interviews, participants reported that the team norms were helpful and followed throughout the process.

The norming and rules presentation is included in Appendix E and grade level group norms are provided in Appendix F.

Maintaining Security of Secure Test Materials

All workshop participants signed a confidentiality agreement during registration and were instructed that the use of laptops, PDAs, and cell phones was prohibited while secure test materials were in the room and that violators would be immediately excused from the process. One ODE laptop was provided to each table for participants to use for note taking. Participants were frequently reminded to not disclose or discuss secure test items. Posters reminded participants to maintain item security during the process and that they were not to disclose or discuss secure test items outside of the standards verification meeting. Secure materials were kept in sight of ODE staff and were moved to a secure vault near the meeting room during breaks.

3.2.3. Training

Training was provided by ODE staff, including Oregon's Director of Assessment and Accountability, Manager of Test Design and Implementation, and Manager of Psychometrics and Validity.

ODE staff trained the panelists on using the bookmark method, Oregon's content standards, assessment, and materials necessary for recommending performance standards. Panelists internalized the concept of target students, who are just barely able to complete the work at the Meets Achievement Level (and *Does Not Yet Meet, Nearly Meets*, and *Exceeds* levels) and came to understand how their understanding of these students would contribute to the bookmark placement task.

Prior to the workshop, ODE provided training to the mathematics consultants. At the end of the workshop each day, the ODE staff met with the grade level group leaders and mathematics consultants to review 1) the perceived effectiveness of the days training, 2) any possible areas of confusion that may benefit from clarification the next day, and 3) their role as small-group leaders and facilitators.

All training activities are discussed in depth below. Training presentations are included in Appendix C.

Workshop Participant Training Overview

Training consisted of a review and discussion of the Oregon Achievement and Content Standards, sample test items, the purpose of the OAKS, the standards setting process, and the ALDs for each performance standard.

Prior to the workshop, participants were sent a packet of materials including links to the following:

- Grade level assignment for the workshop
- The Achievement Level Descriptors
- The mathematics content standards
- An article summarizing best-practices in performance level descriptor development (Perie, 2008).

ODE also provided participants with access to an online OAKS mathematics test to familiarize them with the assessment.

The workshop began with orientation that included a review of the purpose for reviewing the cut scores, current educational context and Oregon's standing within that context, and the workshop agenda. Participants were trained during the afternoon of the first day of the workshop and the morning of the second.
The training covered the following topics:

- The purpose and goals of the Standards Verification Workshop
- A general overview of standard setting and training on the bookmark procedure
- Orientation to Oregon's content standards, test items, and Achievement Level descriptors
- Key concepts and materials, including the Ordered Item Booklet (OIB), Ordered Item Map (OIM), and the Achievement Level Descriptors (ALDs)
- The role of table leaders, facilitating discussion at their tables and helping participants complete tasks in a timely manner
- The agenda for each day

At the end of the training, participants engaged in a brief, mock standard setting using non-secure, released Grade 4 Mathematics items from the National Assessment of Educational Progress (NAEP) to ensure task understanding. During this mock standard setting, participants reviewed and used non-secure sample materials including a sample Ordered Item Booklet (OIB), which can be viewed in Appendix H; Ordered Item Map (OIM), which can be viewed in Appendix I; and the preliminary Achievement Level Descriptors (ALDs), which can be viewed in Appendix A.

Participants evaluated the training; results are described below in section 3.2.5 and in detail in Appendix G.

General Overview of Mathematics Assessment

During the first day of the workshop, participants were provided an overview of OAKS and a description of how assessment scores are used and how changes to cut scores determined throughout the verification process may impact Oregon students and educators. ODE described the external data used in the creation of the projected cut scores and explained how this data allowed for participants to compare Oregon's standards to those of other countries, the nation, and other states.

Workshop leaders described the task and the reasons for reviewing the achievement standards. They reviewed Oregon's achievement standards in relation to other states' and

countries' standards, and discussed the importance and implications of changes to the cut scores, including the impact higher cut scores would have on students, in terms of holding them to higher expectations for learning more challenging content and OAKS pass rates. Throughout the overview and orientation, ODE staff defined and discussed key terms and concepts. At the conclusion of the overview session, workshop participants completed a task to ensure they had internalized shared understanding of these key concepts.

General Overview of Mathematics Content and Achievement Standards

During the training, workshop participants reviewed materials including non-secure sample Ordered Item Booklets (OIBs) and Ordered Item Maps (OIMs), Achievement Level Descriptors, content standards, and non-secure, released test items. Participants created target student descriptions and were trained on bookmark placement.

Materials Review

The following materials were created or used during the workshops. Workshop participants reviewed and received training on each.

Ordered Item Booklets and Ordered Item Maps. The Ordered Item Booklets (OIBs) contained one assessment item per page, ranked in order of increasing difficulty on Oregon's RIT scale. Some scale scores (RITs) were represented by more than one item, particularly around the cut scores and external references. Item difficulty was based on operational 2009-10 data. Workshop participants were not provided the RIT values for items, as the focus was on content and the ordered difficulty.

There was one OIB per grade. Each item was presented with an item ID, the item prompt, and response options. Within each booklet, the current and projected cut points for each Achievement Level were noted on items. ODE also included external reference data providing context for how Oregon's current and projected achievement levels compared to those of other states (through NAEP equivalent state cut scores), the nation (NAEP Basic and Proficient cut scores), and other countries (from PISA).

A newly piloted, innovative polytomous item type, machine scored graphic-response items, were also included in the booklets, at each score point. Item difficulty was based on the 2009-10 field test data. A separate booklet contained sample student responses for each score point so workshop participants could see the responses required to earn each point value.

The Ordered Item Maps contained the page number of each item in the OIB, the external reference data, the Oregon item ID, the ITS item ID, the answer key, the content standard the item represents, and a column for participant notes.

Appendices F and G include non-secure sample Ordered Item Booklets and Ordered Item Maps.

Achievement Level Descriptors. Prior to the standard setting workshop, ODE convened a panel of experts to develop Achievement Level Descriptors (ALDs) for each of the following achievement levels: *Does Not Yet Meet, Nearly Meets, Meets*, and *Exceeds*.

The ALDs were drafted such that each of the four achievement levels differentiated student performance in terms of increasing cognitive demand and task complexity. During the training, ODE reviewed the ALDs with participants and provided sample OIBs containing non-secure released items for participants to use in the training.

After Round Four bookmarks had been placed, participants provided revisions to the original ALDs based on the newly recommended cut scores so they were consistent with the recommended cut scores and described the content necessary for each level as determined by the workshop participants.

Appendix A contains the preliminary Achievement Level Descriptors for each grade level provided to workshop participants. Appendix R contains the revised ALDs recommended at the end of Round Four.

Target Student Descriptions

Once participants were in their pre-assigned, breakout rooms, the group leader within each grade and content area facilitated a target student discussion to help participants articulate the achievement levels. The target student descriptions depict the minimum knowledge and skills that a student must demonstrate on the OAKS in order to reach each Achievement Level.

Participants visualized target students for each Achievement Level using the appropriate content standards, the ALDs, and the workshop participants' expert judgment. Defining target students began individually and then ideas were shared with tables and with grade level groups. Once target students were defined for the *Meets* Achievement Level, participants created them for the *Does Not Yet Meet, Nearly Meets*, and *Exceeds* Achievement Levels. ODE staff facilitated the process, and mathematics consultants provided content expertise as participants developed the target student descriptions.

At the end of the first day, there was evidence that panelists were uncomfortable with the process of visualizing target students and placing their bookmarks. As a result, ODE started the second day with a second round of training in which mathematics consultants role-modeled visualizing a target student and placing bookmarks using the sample OIB. Workshop participants then worked again with their table teams to refine target student descriptors for their grade levels. ODE staff facilitated the process, and mathematics consultants provided content expertise as participants developed the target student descriptions.

Participants were encouraged to take notes during the target student discussion and were asked to refer to the target student descriptors throughout the standard setting. Once finalized, characteristics of target students at each achievement level were recorded and posted near each table. These target student definitions served as a basis for establishing a common understanding of the type of student that should be considered when setting each cut score.

Appendix J contains the presentation and instructions for creating target student descriptions. Appendix K contains each grade level group's target student descriptions.

Bookmark Placement

Each panelist practiced placing bookmarks using their target student description and sample OIB prior to placing Round One bookmarks. Following the practice round, the group discussed the process and ODE staff and mathematics consultants answered questions.

Participants were instructed to use the following tools when placing their bookmarks: the Oregon content standards, their group's target student descriptions, the Achievement Level Descriptors, the content as represented by the items in the Ordered Item Booklets (OIBs), current cut scores, projected cut scores, and external reference data for each Achievement Standard.

Workshop participants were instructed to place their bookmarks considering the likelihood that a just barely proficient student has a 67% likelihood of successfully completing the item. The item in front of the bookmark was the last item in the OIB where the target student had a 67% probability of answering correctly, and the item behind the bookmark was the first item in the OIB where the target student had less than a 67% probability of answering correctly. Workshop participants placed bookmarks between the two items and wrote the first item in the higher category on the bookmark. Bookmarks placed between the last item in one level and in front of first item in the higher level, such that their placement identified the point at which students minimally should know and be able to do. After the cut score, students then would fall into the category defined by that cut score. Participants were instructed to begin by placing the *Meets* bookmark, then the *Nearly Meets*, then the *Exceeds* bookmarks.

Mathematics Consultant and Facilitator Training

Prior to the Standards Verification Workshop, ODE staff leading the workshop provided a half-day training for the mathematics consultants. Senior ODE staff led the training and defined roles and responsibilities. They provided a detailed overview of the workshop process; reviewed materials that would be used by workshop participants, including ordered item booklets and ordered item maps; presented the NAEP and PISA linking methodology, data, analyses, and resulting projected achievement standards and impact data for those standards; and summarized the workshop goals. The mathematics consultants critically reviewed materials to identify and note any errors.

Mathematics consultants completed training evaluation forms at the conclusion of the pre-workshop training session. Results were overwhelmingly positive, 100% of responses indicated mathematics consultants agreed or strongly agreed that they were well trained and prepared to complete their roles in the workshop.

Item level responses to the mathematics consultant training evaluation are provided in Appendix G.

3.2.4. Placing the Bookmarks

The panel followed the bookmarking standard-setting method (Kane, 1994, Mitzel, Lewis, Patz, & Green, 2001) with the addition of external data and projected cut scores. ODE provided this data to give participants the most information possible to use in conjunction with their professional judgment in bookmark placement (recommended practice in Hambleton, 2001, Kane, 1994).

Workshop participants placed the bookmarks at the location in the OIB where the target student defined for that level had a two-thirds chance of correctly responding to the item at that location.

In Round One, participants worked independently to place bookmarks for the *Nearly Meets*, *Meets*, and *Exceeds* Achievement Levels for the even numbered grades (4, 6, 8, and high school). In Round Two, participants reviewed the data from Round One and discussed their bookmark placement in their table teams. In Round Three, workshop participants worked in grade level groups to reach a group consensus around bookmark placement. Once all grade-band panels completed Round Three, ODE psychometric staff derived the cut scores for the adjacent grades (Grades 3, 5, and 7) by analyzing longitudinal student progression from grade 3 to the required high school "Meets" score. To ensure internal validity, the methods were consistent through all four rounds so that ratings indicate increased internal consistency across rounds and panelists (NAGB, 2010).

Round One

Prior to Round One, participants reviewed the instructions for the bookmarking process, the ALDs, and the OIBs to ensure a shared and thorough understanding of the task. ODE staff and the table leads introduced each task, monitored the group during completion of each task, and were available for content related questions.

During Round One, participants worked independently for approximately one hour to determine bookmarks for the even numbered grades (4, 6, 8, and high school). Upon completion of the task, ODE analysts summarized the Round One data as the percent falling into each performance level category for the median OIB page numbers.

Results of Round One are summarized in Tables 12 and 13 below and provided in Appendix L.

Table 12. Round One Median Bookmark Placement by Grade level Group

	Grade 4	Grade 6	Grade 8	HS
Nearly Meets	21	31	26	19
Meets	33	38	34	28
Exceeds	49	48	46	44

Table 13. Round One Impact Data by Grade level Group

	Grade 4	Grade 6	Grade 8	HS
Does Not Yet Meet	18.1%	30.0%	27.9%	29.9%
Nearly Meets	35.6%	23.3%	19%	20.9%
Meets	27%	19%	24.7%	37.8%
Exceeds	19.3%	17.7%	28.4%	11.4%

Round Two

During Round Two, workshop participants reviewed the data from Round One and discussed their bookmark placement in their table teams. Workshop participants took turns explaining their rationale for the low and high individual bookmarks and began to work towards consensus. Informal interviews at the end of the second day indicated that the groups worked efficiently and followed the established protocols and norms.

Overall median recommendations did not change much from Round One, but the variability around medians decreased. Table medians were 1-3 pages apart at the end of Round Two.

Results of Round Two are summarized in Tables 14 and 15 below and provided in Appendix M.

Table 14. Round Two Median Bookmark Placement by Grade level Group

	Grade 4	Grade 6	Grade 8	HS
Nearly Meets	21	30	29	19
Meets	33	38	36	29
Exceeds	49	50	50	48

Table 15. Round Two Impact Data by Grade level Group

	Grade 4	Grade 6	Grade 8	HS
Does Not Yet Meet	18.1%	30.0%	37.8%	29.9%
Nearly Meets	35.56%	23.34%	13.44%	27.25%
Meets	26.99%	33.23%	28.33%	36.91%
Exceeds	19.31%	13.46%	20.42%	5.98%

Round Three

The workshop participants worked in grade level groups for Round Three to reach a group consensus around bookmark placement. Participants reported increased confidence in their bookmarks after Round Three. ODE analysts presented the impact data from the Round Two bookmarks, which represented a marked change in the percentages of students who would obtain Meets or Exceeds scores on the OAKS. This impact data provided the participants with more information to use to judge the reasonableness of their

recommendations and to make modifications if they felt it was appropriate to do so (Hambleton, 2001).

Results of Round Three are summarized in Tables 16 and 17 below and provided in Appendix N.

Table 16. Round Three Median Bookmark Placement by Grade level Group

	Grade 3	**Grade 4**	Grade 5	**Grade 6**	Grade 7	**Grade 8**	**HS**
Nearly Meets	21	**23**	31	**30**	23	**29**	**19**
Meets	34	**33**	39	**38**	27	**35**	**28**
Exceeds	45	**49**	50	**48**	43	**52**	**53**

Table 17. Participant-recommended Mathematics Cut Scores and Associated Impact Data for Target Grades

	Cut Scores			**Impact Data***			
Grade	**Nearly Meets**	**Meets**	**Exceeds**	**Does Not Yet Meet**	**Nearly Meets**	**Meets**	**Exceeds**
4	212	**219**	227	20.4%	33.3%	**27%**	19.3%
6	222	**227**	237	30%	23.3%	**29%**	18%
8	230	**234**	245	27%	19%	**37%**	16.4%
HS	232	**237**	251	30%	21%	**43%**	6%

* Impact data indicate percentage of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment.

Round Four

Derived Cut Scores

Once all grade-band panels completed Round Three, ODE psychometric staff derived the cut scores for the adjacent grades (Grades 3, 5, and 7) analyzing longitudinal student progression from grade 3 to the required high school "Meets" score. This model has been previously used successfully by ODE. Table 18 below shows the derived cut scores and impact data for Grades 3, 5, and 7 along with the target grade data.

Table 18. Recommended and Derived Cut Scores and Impact Data for All Grades Showing Cross-Grade Articulation.

	Cut Scores			Impact Data**				
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds	Meets & Above
3*	205	212	219	25%	28%	29%	18%	47%
4	**212**	**219**	**227**	**20%**	**33%**	**27%**	**19%**	**46%**
5*	219	225	234	21%	30%	34%	16%	50%
6	**222**	**227**	**237**	**30%**	**23%**	**29%**	**18%**	**47%**
7*	228	232	242	29%	19%	34%	18%	52%
8	**230**	**234**	**245**	**27%**	**19%**	**37%**	**16%**	**53%**
HS	**232**	**237**	**251**	**30%**	**21%**	**43%**	**6%**	**49%**

Cross-grade Articulation (Smoothing)

The cut scores and associated impact data determined for the adjacent grades by interpolation were presented to the participants during the cross-grade articulation, or "smoothing," discussion on Day 3. The purpose of this smoothing discussion was to establish a system of cut scores that was well articulated and, at the same time, reflective of the participants' original recommendations. As participants reviewed the derived scores and impact data, each grade-band panel and the group as a whole gave careful consideration to the final recommended scores.

The grade level groups were allowed to discuss and revise their suggested cut scores based on the following factors:

- The content required by the extrapolated cut scores in the extrapolated grades
- The impact data
- The cut scores across grades as a whole

The grade level groups maintained their judgment-based recommendations to raise cut scores with some minor revisions. Revisions made to the original Round Four data addressed the following:

- The high school group lowered their recommended *Exceeds* cut score
- The 8th grade group created a wider difference between the *Nearly Meets* and *Meets* cut scores
- The 5th grade group raised the *Exceeds* cut score

The grade 3-4 grade level group was allowed to keep working until they were confident with their review of the data and the final placement of their cut scores. ODE provided a separate debrief session to the participants of this group.

ODE's smoothing presentation is presented in Appendix O. Results of Round Four are summarized in Table 19 below and provided in Appendix P.

Table 19. Round Four Bookmark Placement by Grade Level

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Nearly Meets	23	23	28	30	22	26	19
Meets	33	33	39	38	26	35	28
Exceeds	45	49	51	49	42	51	48

The recommedations of the panel were to increase the standards for all students, with the largest overall increases in the Nearly Meets and grade 5 cut scores.

Table 20. Recommended Cut Scores and Impact Data for All Grades Showing Cross-Grade Articulation

	Cut Scores			Impact Data*				
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds	Meets & Above
3*	205	212	219	25%	28%	29%	18%	47%
4	**212**	**219**	**227**	**20%**	**33%**	**27%**	**19%**	**46%**
5*	219	225	234	21%	30%	34%	16%	50%
6	**222**	**227**	**237**	**30%**	**23%**	**29%**	**18%**	**47%**
7*	228	232	242	29%	19%	34%	18%	52%
8	**230**	**234**	**245**	**27%**	**19%**	**37%**	**16%**	**53%**
HS	**232**	**237**	**251**	**30%**	**21%**	**43%**	**6%**	**49%**

*Impact data indicate % of Oregon students who would fall within certain achievement levels based on 2009-2010 student assessment.

Variability

As panelists discuss their reasons for placing bookmarks and impact data, variability across tables and individuals often decreases over the rounds of decision making. Taking the standard deviations across bookmark placements for individuals within grade level provides a measure of variability across participants at each round. Variability does decrease with each round, to zero in the 3-4 and 7-8 grade level groups.

Individual bookmarks for each panelist are presented in Appendix Q and are summarized in Table 21 below.

Table 21. Standard deviations and ranges for individual Meets bookmark placement in each round.

	Round 1		Round 2		Round 3	
	St. Dev.	Page Range	St. Dev.	Page Range	St. Dev.	Page Range
Grade 3-4	4.8	26-44	1.0	30-33	0	33-33
Grade 5-6	5.1	30-47	1.9	36-42	.3	38-39
Grade 7-8	3.5	32-42	2.0	33-38	0	35-35
HS	4.5	24-40	3.0	28-38	.7	28-30

Revision of the Achievement Level Descriptors

After the Round Three cut scores were determined, workshop participants refined the ALDs. During this revision, workshop participants were encouraged to review the ALDs to be consistent with their recommended cut scores and the content of the OIB. Revised ALDs are presented in Appendix R.

Workshop Conclusion

The workshop concluded with recommendations from ODE regarding how participants can convey the results of the workshop to their constituents. They stressed the importance of maintaining confidentiality until the standards were released to the public and encouraged participants to share with others the importance of raising standards in order to produce globally competitive students.

Debriefing

Because the recommendations are not final until they have been vetted by the ODE Management Team and released for public comment, panelists were asked not to disclose the specific cut scores recommended by the panel. Upon completion of the workshop, panelists were provided with talking points, including specification of process components that were a) confidential and could not be discussed at any time (secure test items, specific cut scores, impact data), b) those that could be immediately shared with others (the process followed, the types of materials used, the external reference data, and general statements that the panel recommended raising current standards) and c) those that could be shared with others as soon as results of the Standards Verification process were released for public comment (specific recommendations for cut scores).

3.2.5. Process Monitoring and Evaluation

In order to ensure procedural and internal validity, participants and leaders were provided with opportunities to evaluate the process using process check-ins, formal and informal interviews, and training and workshop evaluations (recommended by Hambleton, 2001, National Assessment Governing Board, 2010).

All of the above were utilized throughout the workshop, and results are summarized in the sections below. Additionally, comment cards were left in the back of the room for participants to provided feedback about the workshop process or materials or secure test

items, and some participants used these to note issues or questions that may be important but were not directly relevant to the standard verification task.

Overall, panelists had confidence in the workshop training, methods, and outcomes and felt capable of performing the bookmarking task.

Training Evaluation Forms

At the completion of training, prior to beginning Round One, participants completed a training evaluation comprised of nine Likert type items with a 5-point response scale from "Strongly Disagree" to "Strongly Agree" and one open-ended item for additional comments. A copy of the training evaluation form is provided in Appendix G.

Overall, feedback on the training was positive, for example:

- 88.1% of participants agreed or strongly agreed with the statement, "The training materials were helpful."
- 92.9% of participants agreed or strongly agreed with the statement, "I am confident I understand my role in the standards verification process."
- 71.4% of participants agreed or strongly agreed with the statement, "Overall, I feel well trained and prepared to complete the standards verification task."

While there were some participants who did not feel prepared for the task, later evaluations and interviews indicate that participants felt much more confident and prepared once they engaged in the tasks during Round One.

Response data for each of the training evaluation questions are provided in Appendix G.

Workshop Participant Interviews

On days 2 and 3, panelists were selected for informal and formal interviews with the evaluation team. Panelists who could represent the perspectives of a range of stakeholder groups, or who may have been unfamiliar with the task were selected for formal interviews. Informal interviews were conducted with participants selected at random from each grade level group. The interviews followed a standardized process and protocol. They were conducted in semi-private or private settings.

The interview protocols for both the formal and informal interviews are provided in Appendix S.

Formal Interviews

Five participants were selected for short interviews throughout the process. Selection criteria included participants who may have been unfamiliar or more challenged by the task (parents, community and business representatives), those who could represent the perspectives of the various stakeholder groups in the workshop (higher education, educators of special populations). Interviews were conducted individually at the conclusion of the workshop.

Responses were coded for broad themes, which are summarized as follows:

- Interviewed participants reported that the training was adequate, particularly after the additional training on day two.

- Interviewees reported that they felt quite comfortable with the process; they agreed that raising standards was necessary, even though it may be difficult for educators.
- Interviewees were confident with the outcomes of the workshop. They reported that they were satisfied with their final recommended cut scores.
- Overall, the interviewees reported that the groups worked well together and respected their established norms.
- Interviewees reported wanting more time to accomplish the workshop tasks.

Informal Interviews

Informal interviews were conducted with six workshop participants at the end of the second day to elicit feedback from participants about their progress in reaching consensus during Round Two. Interviews were conducted individually; participants were randomly selected and approached during break times.

The interviews were coded for broad themes, which are summarized as follows:

- Participants reported that the groups worked well together. Interviewed participants made statements such as "the group was open with their opinions," "the group was productive together," and "everyone listened."
- Participants from the high school and 8th grade groups reported that their groups were close to reaching consensus.
- Participants from the 4th grade group reported that they were not yet working to reach group consensus but were still working individually to set their bookmarks for Round One.
- Participants reported that the additional training in the morning of the second day improved their understanding of the bookmarking process.

Workshop Evaluation Forms

At the completion of the standards verification, participants completed an evaluation about the workshop process and outcomes. The evaluation form is provided in Appendix T, and data are provided in Appendices C and T and results are summarized below.

Generally, feedback was positive and included the following:

- 81.4% of participants agreed or strongly agreed with the statement, "The Bookmark Procedure was well described."
- 97.6% of participants agreed or strongly agreed with the statement, "Overall, I am satisfied with my group's final bookmarks."
- 83.7% of participants agreed or strongly agreed with the statement, "I am confident that the Bookmark Procedure used produced valid cut scores."
- 100% of participants agreed or strongly agreed with the statement, "I feel this procedure was fair."

Participants were asked to write comments on their evaluation forms. These comments were coded for broad themes, which included the following:

- K-12 educators commented that the workshop gave them a better understanding of standards and assessments; however, participants from the business community reported wanting more explanation of the standards and accountability system.
- Participants from the grade 3-4 grade level group reported that the extra time given for Round Three was very beneficial and helped them feel comfortable with their recommended cut scores.
- Participants appreciated the additional training on the second day.
- Participants reported wanting more time for the tasks.

Participants were asked to evaluate which factors were the most important for their bookmark placement. They rated the panel discussions as most important, followed by the OIBs and the ALDs. Participants indicated that the sample student responses were least important, which is not unreasonable given that sample responses were provided for the new technology-enhanced constructed response item types and there were only a few (approximately 2) in each grade.

These data are provided in Appendix T and results are summarized in Table 22 below.

Table 22. Importance of factors used to place bookmarks.

Factor	N	Importance 1="Not Important", 2="Somewhat Important", 3="Important", 4="Very Important"				N/A	Important + Very Important
		1	2	3	4		
The Achievement Level Descriptions (ALDs) of *Does Not Yet Meet*, *Nearly Meets, Meets, Exceeds.*	42	0%	5%	30%	65%	0%	95%
Your perceptions of the difficulty of the items in the Ordered Item Booklet.	42	0%	2%	33%	65%	0%	98%
Your perceptions of the quality of the sample student responses.	42	10%	24%	24%	29%	14%	52%
Your own classroom experience.	42	2%	5%	24%	62%	7%	86%
Visualizing a Target Student.	42	2%	7%	30%	58%	2%	88%
The impact data.	25	8%	16%	32%	44%	0%	76%
The PISA, NAEP & OUS calibration data.	42	2%	19%	40%	38%	0%	79%
The presentation of impact data.	42	5%	12%	40%	42%	2%	81%
Your initial classification of student performance in	42	2%	14%	49%	33%	2%	81%

Round One.							
Panel discussions.	42	0%	2%	9%	86%	2%	95%
The initial classifications of other panelists	42	0%	19%	38%	40%	2%	79%

Process Check-ins

At the end of each day, ODE staff met with the mathematics consultants to ensure shared understanding of process and key concepts and to review timeline revisions or new tasks for the following day. These meetings provided an opportunity to maintain consistent communication and expectations across tables (such as keeping panelists focused and on-task). ODE staff implemented the suggestions and adjusted the timeline each night for the next day's activities.

3.2.6. Formal Adoption of Challenging Academic Content Standards

The State Board of Education will consider adoption of the cut scores on October 28, 2010.

REFERENCES

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2008). Standards for Educational and Psychological Testing. Washington, D.C.: American Educational Research Association.

Cizek, G. J., & Bunch, M.B. (2007). Standard setting: A practitioner's guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: SAGE.

Bandeira de Mello, V.P., Blankenship, C., McLaughlin, D.H. (2009) *Mapping State Proficiency Standards onto NAEP Scales: 2005-2007*. Washington, D.C.: National Center for Education Statistics. (NCES 2010-456)

Haertel, E. H., & Lorié, W. A. (2004). Validating standards-based test score interpretations. *Measurement: Interdisciplinary Research and Perspectives*, *2*, 61-103.

Hambleton, R.K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G.J. Cizek (Ed.), *Setting Performance standards: Concepts, methods, and perspectives* (pp. 89-116). Mahwah, N.J.: Lawrence Earlbaum Associates.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425-461.

Mitzel, Lewis, Patz, & Green (2001). The bookmark procedure: Psychological perspectives. In G.J. Cizek (Ed.), *Setting Performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, N.J.: Lawrence Earlbaum Associates.

National Assessment Governing Board (2010). *Work Statement for Judgmental Standard Setting Workshops for the 2009 Grade 12 Reading and Mathematics National Assessment of Educational Progress to Reference Academic Preparedness for College Course Placement* (Higher Education Solicitation number ED-R-10-0005). Retrieved from http://www.nagb.org/what-we-do/register-notice.htm

Perie, M. (2008). A guide to understanding and developing performance-level descriptors.

Phillips, G.W., (2009). The second derivative: International benchmarks in mathematics for US states and school districts. Washington, D.C.: American Institutes for Research.

PISA (2007). Science Competencies for Tomorrow's World Executive Summary. Retrieved from http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

PISA (2009). Proficiency scale construction. In *PISA 2006 Technical Report* (15). Paris, France: OECD Publishing.

U.S. Department of Education (2007). Standards and assessments: Peer review information and examples for meeting requirements of the No Child Left Behind Act of 2001. Washington, D.C.: U.S. Department of Education Office of Elementary and Secondary Education.

Zieky, M., & Perie, M. (2006). A primer on setting cut scores on tests of educational achievement. Princeton, NJ: Educational Testing Service.