Oregon Science Standards Verification Technical Report

Grades 5, 8, and High School

October 2011

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

EXECUTIVE SUMMARY

The Oregon Department of Education (ODE) conducted an achievement standards verification process for the Oregon Assessment of Knowledge and Skills for Science on July 19–21, 2011, using the empirically-based bookmark procedure to recommend Achievement Standards for Grades 5, 8, and high school. This document summarizes the resulting recommendations. The recommended Achievement Standards will be reviewed through a public process August through September and will be considered by The State Board of Education on October 20, 2011.

Why is ODE setting new Achievement Standards for Science?

- Science teachers are basing their instruction on the Content Standards adopted in 2009, and the assessment needs to follow.
- Given the scope of the 2009 change in Content Standards, the U.S. Department of Education requires that ODE assess these new Content Standards in concert with the achievement standard-setting review process.

The Extended Assessment alternate Achievement Standards (including Achievement Level Descriptors and Achievement Standards) were reviewed and proposed in August 2011.

How were recommended Achievement Standards identified?

Participants were recruited from across Oregon to participate in benchmark 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 key characteristics (e.g., gender, geographic location). Participants used booklets that contained 70 secure test items arranged from least to most difficult, as well as the new Science Content Standards and Achievement Level Descriptors—to verify the knowledge and skills that students should demonstrate at each assessed grade level. For participants' reference in making their judgments, the current Oregon Achievement Standards, the performance of other states and countries in the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA), and NAEP achievement standards were marked in the booklets.

Panelists participated in three review rounds in which they individually recommended three Achievement Standards (*Nearly Meets, Meets,* and *Exceeds*) that defined four performance levels: *Does Not Yet Meet, Nearly Meets, Meets,* and *Exceeds* for grades 5, 8, and high school. They also considered impact data, an analysis that forecasts the potential percentages of students meeting, not meeting, and exceeding Achievement Standards at each grade, based on prior year's test results.

What Achievement Standards did the group recommend for 2011–12 and beyond?

Table 1 summarizes the Achievement Standards and associated impact data for the three 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 the 2010–11 test administration.

Table 1. Participant-Recommended Science Achievement Standards and Associated Impact Data for Grades 5, 8, and High School*

	Achievement Standards			Impact Data**			
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds
5	216	**226**	239	8%	22%	**51%**	19%
8	229	**235**	247	15%	18%	**49%**	19%
HS	235	**240**	252	21%	8%	**56%**	15%

* 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 2010–2011 science assessment.

Cross-Grade Articulation (Smoothing)

The Achievement Standards and associated impact data determined for grades 5, 8, and high school 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 Achievement Standards 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.

Summary

Figure 1 displays the recommended Science Achievement Standards at the *Nearly Meets*, *Meets,* and *Exceeds* levels at grades 5, 8, and high school.

Figure 1. Cross-Grade Progression of Recommended Science Achievement Standards

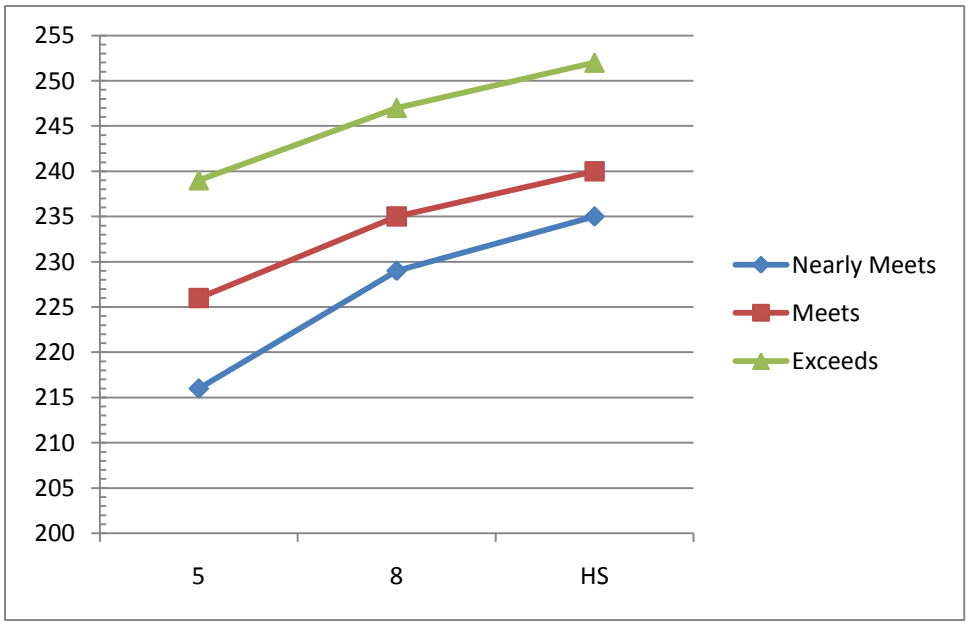

Table 2 shows the final recommended Achievement Standards, current Achievement Standards, and the changes to the Achievement Standards.

Table 2. Final Recommended Achievement Standards, Current Achievement Standards, and Changes to Achievement Standards

	Recommended Achievement Standards Science 2011–2012			**Current Achievement Standards Science 2010–2011**			**Change to Achievement Standard** (+/- Resulting from Recommended Minus Current Achievement Standard)		
Achievement Level	**5**	**8**	**HS**	**5**	**8**	**HS**	**5**	**8**	**HS**
Nearly Meets	216	229	235	216	229	235	0	0	0
Meets	**226**	**235**	**240**	**225**	**234**	**240**	**+1**	**+1**	**0**
Exceeds	239	247	252	238	246	249	+1	+1	+3

How Can We Provide Feedback?

ODE welcomed and encouraged feedback from Oregon stakeholders. A survey gathering structured input on these recommendations was available through a link in the Assessment and Accountability Update the week of August 29, 2011.

Conclusion

This Executive Summary provides an overview of the process used to develop recommended Achievement Standards for the Oregon Assessment of Knowledge and Skills (OAKS) Science and the resulting recommendations from the participants. More detailed information will be posted on the ODE website at http://www.ode.state.or.us/search/page/?id=3319.

1. INTRODUCTION

In July 2011, staff from the Oregon Department of Education (ODE) conducted the Oregon Assessment of Knowledge and Skills (OAKS) Science standards verification 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 Science Achievement Standards fall in relation to other states and countries for panelists to consider while reviewing and setting the new Achievement Standards (cut scores). Participants were provided reference cut points obtained by embedding items from the PISA in the OAKS, linking the NAEP and OAKS scales, and through analysis of postsecondary student performance data obtained through ODE's partnership with the Oregon University System (OUS).

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 Science at each grade level.

ODE sought stakeholder review of the current Achievement Standards because the State Board of Education adopted new Content Standards in 2009 and because Science achievement reporting was moved from grade 10 to grade 11.

To set the Science Achievement Standards, ODE recruited a diverse set of panelists from across the state. Panelists brought expertise in science and represented the range of stakeholder characteristics. 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 5–8 and high school.

The final recommendations from the panel are described in Table 3, which summarizes the standards recommended by the panel and the associated impact data. The change to the Achievement Standards is described in Table 4.

Table 3. Recommended Achievement Standards and Impact Data for All Grades Showing Cross-Grade Articulation

	Achievement Standards			Impact Data*				
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds	Meets & Above
5	216	**226**	239	8%	22%	**51%**	19%	70%
8	229	**235**	247	15%	18%	**49%**	19%	68%
HS	235	**240**	252	21%	8%	**56%**	15%	71%

*Impact data indicate % of Oregon students who would fall within certain Achievement Levels based on 2010–2011 student assessment

Table 4. Change to Achievement Standards (+/- Resulting from Recommended Minus Current Achievement Standards)

Achievement Level	Grade 5	Grade 8	High School
Nearly Meets	0	0	0
Meets	**+1**	**+1**	**0**
Exceeds	+1	+1	+3

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 and computer-scored 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 Science Standards

Oregon's standards system consists of Oregon's Content Standards and Academic Achievement Standards. Content standards define the knowledge and skills that 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. For science, grade 3–5 content is tested at grade 5, grade 6–8 content is tested at grade 8, and high school content is tested at grade 11.

2.2.1. Science 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 and describe what Oregon students are expected to know at each grade level.

The Science Content Standards were most recently revised in 2009 and were subsequently adopted by the State Board of Education. The next anticipated revision of the Science Content Standards will occur in 2015–16.

Oregon's Academic Content Standards are available on the ODE website 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. Science 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, and in 1999 for science. Prior to 2010–11, 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 Science are provided in Table 5.

Table 5. Current Achievement Standards for Science 2010–2011

Achievement Level	Grade 5	Grade 8	High School
Nearly Meets	216	229	235
Meets	**225**	**234**	**240**
Exceeds	238	246	249

Starting with the 9th 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, and applying mathematics.

2.2.3. Science Achievement Level Descriptors (ALDs)

Oregon's science assessments use four levels of achievement—*Exceeds*, *Meets*, *Nearly Meets*, and *Does Not Yet Meet*. The grade- and content-specific Achievement Level descriptors describe the knowledge and skills required by students at each level of performance. The preliminary Achievement Level Descriptors (ALDs) are available in Appendix A and on the ODE website at http://www.ode.state.or.us/search/results/?id=223. The ALDs recommended by the 2011 Standard Verification Panel are available in Appendix R.

Prior to the July 2011 workshop, ODE worked with stakeholders to draft preliminary ALDs that include general Policy Definitions. The Policy Definitions provide an overarching definition (across grade and content area) for each Achievement Level and describe how rigorous and challenging the Achievement Standards 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 science

content panel. 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 2010, ODE received feedback from panelists from around the state. One survey was for members of the English Language Arts content panel who were asked about the Reading/Literature ALDs and the Policy Definitions. The second survey was for members of the other content panels (Mathematics, Science, Social Sciences, and ELPA) and just concerned the Policy Definitions. Feedback from both surveys was supportive of the direction of these drafts.

Based on feedback from the field, ODE staff made adjustments and improvements to the Policy Definitions and Science ALDs. 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 2011 SCIENCE STANDARDS VERIFICATION WORKSHOP

The Science Achievement Standards were established in a workshop held in July 2011 using a modified bookmarking standard setting procedure (Cizek & Bunch, 2007; Kane, 1994; Mitzel, Lewis, Patz, & Green, 2001). The same procedure was previously implemented for mathematics verification in August 2010 and reading in January 2011. Twenty-four Oregonians recommended Achievement Standards for grades 5, 8, and high school in science. ODE science 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 locations in this report for specific evidence that the process met the expectations described for appropriate, high-quality achievement standards are summarized in Table 6.

Table 6. Evidence of Validity Documented During Evaluation

Standard	Evidence
Panels should be large enough and representative enough of the appropriate constituencies.	Grade Level Group Composition, 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, 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, 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, Section 3.2.4.
To ensure procedural validity, the procedures must be reasonable, carried out as intended, and understood by panelists.	The 2011 Science Standards Verification Workshop, Section 3; Training, Section 3.2.3; Placing the Bookmarks, Section 3.2.4; Training Evaluation, Section 3.2.5; Appendix G.
The methodology should be appropriate for the assessment, described in detail, and field-tested when appropriate.	The 2011 Science Standards Verification Workshop, Section 3; Derived Achievement Standards, Section 3.2.4.

Table 6. Evidence of Validity Documented During Evaluation (cont.)

Standard	Evidence
Any non-standard methodology must be clearly documented.	The 2011 Science Standards Verification Workshop, Section 3; Science 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, Section 3.2.5; Placing the Bookmarks, Section 3.2.4; Appendix Q; Appendix S; Target Student Descriptions, Section 3.2.3; Appendices G & I.
The rationale and procedures for establishing achievement standards must be documented.	Training, Section 3.2.3; Placing the Bookmarks, Section 3.2.4; Table 22, Section 3.2.5; Science 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, Section 3.2.4; Table 22, Section 3.2.5; Science 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, Section 3.2.3.
Descriptions of performance categories must be clear to the extent that participants are able to use them effectively.	Science Achievement Level Descriptors, Section 2.2.3; Achievement Level Descriptors, 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, Section 3.2.3; Process Monitoring and Evaluation, Section 3.2.5; Appendices E, F, T.
Facilitators should be qualified and capable of leading appropriate discussion among the participants without biasing the process.	Science Consultant Training, Section 3.2.3; Grade Level Group Composition, 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.

Table 6. Evidence of Validity Documented During Evaluation (cont.)

Standard	Evidence
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, Section 3.2.5; Science 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 established using empirical criterion reference data.	Placing the Bookmarks, Section 3.2.4; The 2011 Science Standards Verification Workshop, Section 3; Science 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, Section 3.2.2; Training, Section 3.2.3; Placing the Bookmarks, 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 5, 8, and high school, which entailed a review of impact data based on assessment results from the 2010–11 academic year and bookmark placement across grade level groups and table teams. The workshop concluded with 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 Science Achievement Standard-setting procedure were as follows:

- Establish what students in each grade (5, 8, and high school) should be able to demonstrate on the OAKS in Science 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 Achievement

Standards 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 Achievement Standards for each Achievement Level

3.2. Science Achievement Standards Verification Process Summary

From July 19 to July 21, 2011, 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 Science in grades 5, 8, and high school on the Oregon Assessment of Knowledge and Skills (OAKS).

Twenty-four knowledgeable participants, including educators, higher education representatives, parents, and community members, were recruited from across Oregon to participate in grade-level groups at grades 5, 8, and high school. Using a modified bookmarking procedure (Cizek & Bunch, 2007; Kane, 1994; Mitzel, Lewis, Patz, & Green, 2001), workshop participants received training from ODE staff and completed four rounds of standards verification over three days to determine the *Nearly Meets*, *Meets,* and *Exceeds* Achievement Standards.

Workshop participants were assigned to one of three grade-level groups at grades 5, 8, and high school, with two smaller table teams (A and B) in each group. ODE assigned participants to table teams to ensure that they were balanced in terms of relevant demographic characteristics (e.g., gender, geographic location). Participants used booklets that contained 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 Achievement Standards and the achievement standards for other states and national and international assessments, such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA), were indicated in the booklets.

In order to set the Achievement Levels, panelists participated in three review rounds in which they individually recommended three Achievement Standards (*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 Achievement Standards. This policy model has been previously used successfully by ODE.

The Achievement Standards and associated impact data were presented to the participants during Round Four, the cross-grade articulation, or "smoothing," discussion on Day 3. The purpose of this smoothing discussion was to establish a system of Achievement Standards 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 scores. They also considered impact data, an analysis that forecasts the potential percentages of students meeting, not meeting, and exceeding standards at each grade based on a prior year's test results. Table 2 above shows the final Achievement Standards and impact data for grades 5, 8, and high school. 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. Senior ODE staff were available to answer

policy-related questions, and the Manager of Psychometrics and Validity summarized the results to panelists and answered technical questions.

Following the workshop, participants completed evaluations about the process and outcomes.

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, Achievement Standards, and the preliminary Achievement Standards 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 Achievement Standards to those of other countries (via PISA), the nation (via NAEP), and other states (via other state's NAEP linking studies).

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. Before reviewing the Achievement Standards, panelists were reminded that high standards are necessary to adequately prepare Oregon students.

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 Science Consultants

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

Standards Verification Workshop Participants

Twenty-four Oregonians participated in the Standards Verification Workshop. The panel was carefully selected to represent Oregon stakeholders, to include K–12 educators (95%), and business members (5%) from both rural and urban communities. Parents and representatives from higher education also attended; these participants indicated a secondary role on the participant characteristic section of the workshop evaluations, and thus are not included in the percentages above. 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, ODE solicited involvement from all levels of the education system and from the community. Nominations were solicited from teacher organizations and educator networks. Non-educators in the business and parent communities were recruited via email to the state parent organization.

From the individuals who expressed interest in participating, ODE selected 24 to represent the needs and demographics of Oregon students, taking into account 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 24 workshop participants were divided into three 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 science 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. As a result, some background data reported by participants on the evaluations differs from the background data used by ODE for recruitment.

Table 7 shows the educational background of participants in each grade level group.

Table 7. Participant Educational Background by Grade Level Group

Grades	N	HSD or GED	Bachelor's	Master's	Doctorate
All	20	0%	0%	100%	0%
5	7	0%	0%	100%	0%
8	6	0%	0%	100%	0%
HS	7	0%	0%	100%	0%

Note. Data reported in table represents self-report data collected on workshop evaluations. Four participants did not indicate educational attainment.

Table 8 shows the occupation of participants in each grade level group.

Table 8. Participant Occupation by Grade Level Group

Grades	N	K–12 educator	Community college educator	University educator	Parent	Community member	Business member	Other
All	20	95%	0%	0%	0%	0%	5%	0%
5	7	85.7%	0%	0%	0%	0%	14.3%	0%
8	6	100%	0%	0%	0%	0%	0%	0%
HS	7	100%	0%	0%	0%	0%	0%	0%

Note. Although participants were selected to represent a range of occupational levels, some self-reported otherwise. For example, although all participants identified themselves as either K-12 educators or members of the business community, 29% of grade 5 participants and 17% of grade 8 participants were selected by ODE to represent parents of Oregon students, and 17% of participants in grade 8 were selected to represent postsecondary educators. Data reported in table represents self-report data collected on workshop evaluations. Four participants did not indicate occupational affiliation.

ODE was diligent in recruiting diverse participants who represent broad constituencies, educational environments, and expertise, including rural/urban and content strand knowledge and expertise (Physical Science, Life Science, and Earth and Space Science). However, as noted above, many participants self-reported on their evaluations that they belong to another category than that which they were selected to represent (e.g., they reported being a teacher but were selected because they are a parent).

Table 9 shows the years of work experience for each grade level group.

Table 9. Years of Work Experience by Grade Level Group

Grades	N	1–5	6–10	11–15	16–20	21+
All	21	14.3%	23.8%	14.3%	9.5%	38.1%
5	7	14.3%	42.9%	14.3%	0%	28.6%
8	7	14.3%	0%	14.3%	14.3%	57.1%
HS	7	14.3%	28.6%	14.3%	14.3%	28.6%

Note. Data reported in table represents self-report data collected on workshop evaluations. Three participants did not indicate work experience.

Table 10 shows participants' experience teaching special education (SPED), English language learners (ELL), vocational education (Voc. Ed.), alternative education (Alt. Ed.), and adult education.

Table 10. Participant Teaching Experience with Diverse Populations by Grade Level Group

Grades	N	SPED	ELL	Voc. Ed.	Alt. Ed.	Adult Ed.
All	21	23.8%	19%	4.8%	9.5%	33.3%
5	7	42.9%	28.6%	0%	0%	28.6%
8	7	14.3%	0%	0%	0%	14.3%
HS	7	14.3%	28.6%	28.6%	28.6%	57.1%

Note. Data reported in table represents self-report data collected on workshop evaluations. Three participants did not indicate educational teaching experience with diverse populations.

Participant Roles and Responsibilities

Workshop participants included the following:

- ODE staff
- Science consultants
- Grade level group leads (grades 5, 8, and high school)
- Table team leads (A/B)
- External process evaluators

ODE staff planned and ran the workshop. During the workshop, their responsibilities included training, keeping secure materials confidential, 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.

Science 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 science 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 facilitated 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.

Three external evaluators from the Educational Policy Improvement Center were non-participatory observers for the entire process.

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. ODE provided norms for all groups to follow based on the norms generated in the two previous verification workshops for mathematics and reading. 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, science consultants and ODE staff reminded table teams of the norms agreed upon during the first day. During process evaluation interviews, participants reported that the team norms were helpful and followed throughout the process.

The grade level group norms are provided in Appendices E and 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. 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 Manager of Test Design and Implementation and the Manager of Psychometrics and Validity.

ODE staff trained panelists on the bookmark method, Oregon's Content Standards, the OAKS, 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 corresponding targets at the *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 science consultants. At the end of the workshop each day, the ODE staff met with the grade level group leaders and science consultants to review 1) the perceived effectiveness of the day's training, 2) identification of 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 Science Content Standards
- An article summarizing best practices in performance level descriptor development (Perie, 2008).

The workshop began with a day-long orientation and training that included a review of the purpose for reviewing the Achievement Standards, current educational context and Oregon's standing within that context, and the workshop agenda.

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), Computer Scored Constructed Response Item Booklets, 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, practice standard setting using released science items from the OAKS to ensure task understanding. During this practice session, participants reviewed and used sample materials including sample Ordered Item Booklets (OIB), which can be viewed in Appendix H; Ordered Item Maps (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 Science 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 Achievement Standards determined throughout the verification process may affect Oregon students and educators. ODE described the external data used in the standard-setting process, and explained how these 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 standards from other states and countries, and discussed the importance and implications of changes to the Achievement Standards, including the impact that higher Achievement Standards would have on students, in terms of holding them to higher expectations for learning more challenging content and meeting 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 that they had internalized shared understanding of these key concepts.

General Overview of Science Content and Achievement Standards

During the training, workshop participants reviewed materials including sample Ordered Item Booklets (OIBs), Computer Scored Constructed Response Booklets, Ordered Item Maps (OIMs), Achievement Level Descriptors, and the Science Content Standards. 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, Computer Scored Constructed Response 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 Achievement Standards and external references. Item difficulty was based on operational 2010–11 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 included an item ID, the item prompt, and response options. Within each OIB, the current cut points for each Achievement Level were noted on items. ODE also included external reference data, providing context for how Oregon's current Achievement Levels compared to other states (through NAEP equivalent state achievement standards), the nation (NAEP Basic and Proficient achievement standards), to other countries (from PISA), and to student performance in first-year science courses in Oregon's University System (OUS).

The Ordered Item Maps contained the page number of each item in the OIB, the external reference data (NAEP, PISA, and OUS), the location of current Oregon Achievement Standards, the Oregon item ID, the answer key, the Content Standard the item represents, and a column for participant notes.

Appendices H and I include 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 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 Achievement Standards so they were consistent with the recommended Achievement Standards 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

During the first day of training, ODE led the participants in an exercise to develop target student descriptions. The target student descriptions depict the minimum knowledge and skills that a student must demonstrate on the OAKS in order to "just barely" reach each Achievement Level. After modeling the process, ODE asked participants to read through the on-grade ALDs silently and begin to think about target students at each of the Achievement Levels.

After ODE trained participants, the group leader within each grade and content area facilitated a discussion to help participants articulate what a target student could demonstrate at each of the Achievement Levels. 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 science 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 Achievement Standard.

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 science 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 Achievement Standards, 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 were placed after the last item in one Level and in front of the first item in the higher Level, such that their placement identified the items students in each Achievement Level should be able to answer correctly 67% of the time. Participants were instructed to begin by placing the *Meets* bookmark, then the *Nearly Meets*, then the *Exceeds* bookmarks.

Science Consultant and Facilitator Training

Prior to the Standards Verification Workshop, ODE staff leading the workshop provided two days of training for the science 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 impact data for those standards; and summarized the workshop goals. The science consultants critically reviewed materials to identify and note any errors.

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 references. ODE provided these 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 (67%) 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. 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 analyzed the longitudinal student progression from grade 5 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 grades 5, 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 below and provided in Appendix L.

Table 11. Round One Median Bookmark Placement by Grade Level Group

	Grade 5	Grade 8	High School
Nearly Meets	20	16	22
Meets	**36**	**26**	**32**
Exceeds	54	48	54

Table 12. Round One Impact Data by Grade Level Group

	Grade 5	Grade 8	High School
Does Not Yet Meet	6.8%	14.9%	18.9%
Nearly Meets	21.6%	13.1%	10.2%
Meets	**48.3%**	**49.9%**	**53.4%**
Exceeds	22.3%	22.1%	17.5%
Meets and Above	70.6%	72%	70.9%

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 toward consensus.

Overall median recommendations changed from Round One, and the variability around medians decreased. Table medians were 1–7 pages apart at the end of Round Two.

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

Table 13. Round Two Median Bookmark Placement by Grade Level Group

	Grade 5	Grade 8	High School
Nearly Meets	17	16	20
Meets	**36**	**30**	**33**
Exceeds	53	50	58

Table 14. Round Two Impact Data by Grade Level Group

	Grade 5	Grade 8	High School
Does Not Yet Meet	5.1%	14.9%	15.2%
Nearly Meets	24.3%	27%	20.5%
Meets	**48.3%**	**39.2%**	**52.1%**
Exceeds	22.3%	18.9%	12.2%
Meets and Above	70.6%	58.1%	64.3%

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 below and provided in Appendix N.

Table 15. Round Three Median Bookmark Placement and Associated Impact Data

	Achievement Standards			Impact Data*			
Grade	**Nearly Meets**	**Meets**	**Exceeds**	**Does Not Yet Meet**	**Nearly Meets**	**Meets**	**Exceeds**
5	21	**36**	56	7.8%	21.6%	**51.3%**	19.3%
8	17	**27**	50	14.9%	17.8%	**48.5%**	18.9%
HS	23	**32**	57	18.9%	10.2%	**56.3%**	14.6%

* Impact data indicate percentage of Oregon students who would fall within certain Achievement Levels based on 2010–2011 student assessment.

Round Four

Derived Achievement Standards

Once all grade-band panels completed Round Three, ODE group leaders facilitated the sharing out of Achievement Standards and impact data from Round Three, including the recommendations regarding cross-grade articulation.

Cross-grade Articulation (Smoothing)

The Achievement Standards and associated impact data 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 Achievement Standards 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 Achievement Standards based on the following factors:

- The impact data
- The Achievement Standards across grades as a whole

The grade level groups maintained their judgment-based recommendations to raise Achievement Standards with no revisions. The final recommended Achievement Standards and impact data for grade 5, 8, and high school are below and presented in Appendix P.

Table 16. Final Recommended Achievement Standards and Impact Data for All Grades.

	Achievement Standards			Impact Data*			
Grade	Nearly Meets	Meets	Exceeds	Does Not Yet Meet	Nearly Meets	Meets	Exceeds
5	216	**226**	239	8%	22%	**51%**	19%
8	229	**235**	247	15%	18%	**49%**	19%
HS	235	**240**	252	21%	8%	**56%**	15%

* Impact data indicate percentage of Oregon students who would fall within certain Achievement Levels based on 2010–2011 student assessment.

The recommedations of the panel were to increase the standards, with the largest overall increases in the *Exceeds* and high school Achievement Standards.

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 grade 5 and high school groups.

Individual bookmarks for each panelist are presented in Appendix Q and are summarized in the table below.

Table 17. Standard Deviations and Ranges for Individual *Meets* Bookmark Placement in Each Round.

	Round 1		Round 2		Round 3	
Grade	St. Dev.	Page Range	St. Dev.	Page Range	St. Dev.	Page Range
5	9.4	11–40	0	36–36	0	36–36
8	7.5	25–46	3.2	25–33	2.4	27–33
HS	9.2	18–51	1.9	30–36	0	32–32

Revision of the Achievement Level Descriptors

After the Round Three Achievement Standards were determined, workshop participants refined the ALDs. During this revision, workshop participants were encouraged to review the ALDs to be consistent with their recommended Achievement Standards 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. ODE 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 approved by the Board and are not public until they have been released by ODE for public comment, panelists were asked not to immediately disclose the specific recommended Achievement Standards. 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 Achievement Standards, 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 Achievement Standards).

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, 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 provide 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 composed 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:

- 100% of participants agreed or strongly agreed with the statement, "The training materials were helpful."
- 100% of participants agreed or strongly agreed with the statement, "I am confident I understand my role in the standards verification process."
- 100% of participants agreed or strongly agreed with the statement, "The training clearly identified the goals for the standards verification procedure."
- 95.7% of participants agreed or strongly agreed with the statement, "Overall, I feel well trained and prepared to complete the standards verification task."

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 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. The interviews followed a standardized process and protocol. They were conducted in semi-private or private settings.

The interview protocols are provided in Appendix S.

Interviews

Nine 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), and 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 prepared them for the task and that they felt comfortable with the process.
- Interviewees were confident with the outcomes of the workshop. They reported that they were very satisfied with their final recommended Achievement Standards.
- Interviewees found the guidelines and process of manipulating target student descriptors somewhat challenging.
- Interviewees reported that the groups worked well together and respected their established norms.
- Overall, interviewees found external references and impact data helpful in the later parts of the workshop.

Workshop Evaluation Forms

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

Feedback was very positive and included the following:

- 100% of participants agreed or strongly agreed with the statement, "The bookmark procedure was well described."
- 100% of participants agreed or strongly agreed with the statement, "I understood how to place my bookmarks."
- 100% of participants agreed or strongly agreed with the statement, "Overall, I am satisfied with my group's final bookmarks."
- 100% of participants agreed or strongly agreed with the statement, "I am confident that the bookmark procedure used produced valid Achievement Standards."

Participants were asked questions about the relative importance they placed on the factors used in their bookmark placement, including the materials provided, the external referents,

and the impact data. These data are provided in Appendix T and results are summarized below.

Overall, participants placed the most importance on the following:

- Participants placed equal importance on their perceptions of the difficulty of the items in the Ordered Item Booklet and their own classroom experience, with 100% responding that they were important or very important to their bookmark placement.
- Participants rated the panel discussions and the ALDs equally, with 95.2% responding that they were important or very important to their bookmark placement.
- Participants placed the least importance on the PISA, NAEP, and OUS calibration data, with 50% responding that the data were important or very important to their bookmark placement.

Table 18. Importance of Factors Used to Place Bookmarks.

Factor	N	Not Important	Somewhat Important	Important	Very Important	N/A	Important + Very Important
The Achievement Level Descriptors (ALDs) of *Does Not Yet Meet, Nearly Meets, Meets, Exceeds.*	21	0%	4.8%	9.5%	85.7%	0%	95.2%
Your perceptions of the difficulty of the items in the Ordered Item Booklet.	21	0%	0%	42.9%	57.1%	0%	100%
Your perceptions of the quality of the sample student responses.	20	15.0%	10%	45.0%	20%	10%	65%
Your own classroom experience.	21	0%	0%	38.1%	52.4%	0%	100%
Visualizing a target student.	21	0%	9.5%	52.4%	38.1%	0%	90.5%
The impact data.	21	0%	14.3%	57.1%	28.6%	0%	85.7%
The PISA, NAEP, & OUS calibration data.	20	15%	25%	40%	10%	10%	50%
Your initial classification of student performance in Round One.	21	0%	14.3%	61.9%	23.8%	0%	85.7%
Panel discussions.	21	0%	4.8%	19.0%	76.2%	0%	95.2%
The initial classifications of other panelists.	21	0%	23.8%	28.6%	38.1%	4.8%	66.7%

Note. Some participants did not respond to these items.

Process Check-Ins

At the end of each day, ODE staff met with the science 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 Achievement Standards on October 20, 2011.

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. *Educational Measurement: Issues and Practice, 27*(4), 15–19.

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.