Oregon English Language Proficiency Assessment Performance Standards Verification Technical Report

Grades Kindergarten–High School

November 2012 and February 2013

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

Liz Gilkey, JD, MA
Mary Seburn, PhD
Cristen McLean, MA
David T. Conley, PhD

epic
Educational Policy
Improvement Center

List of Tables and Figures

List of Appendices

1. Introduction

In November of 2012 and February of 2013, the Oregon Department of Education (ODE) held two workshops to conduct the English Language Proficiency Assessment (ELPA) performance standards verification using the bookmark procedure (Cizek & Bunch, 2007; Kane, 1994; Mitzel, Lewis, Patz, & Green, 2001).

To set the bookmarks, ODE recruited a diverse set of participants from across the state. Participants brought expertise in English Language instruction and represented the range of stakeholder characteristics. They were split into grade-level groups and table teams within those groups. They then participated in six rounds of bookmarking and set four achievement standards defining five Proficiency Levels (*Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced)* for grades kindergarten (K) through high school.

The final recommendations from the panel are described in Table 1, which summarizes the standards recommended by the panel and the associated impact data. The changes to the current cut scores based on participants' recommendations are described in Table 2.

Table 1. Recommended Cut Scores and Impact Data for All Grades Showing Cross-Grade Articulation

Grade	Cut Scores				Impact Data			
	EI	I	EA	A	EI	I	EA	A
K	481	491	497	505	36%	21%	17%	6%
1	491	503	512	522	27%	32%	21%	9%
2	492	504	514	521	28%	37%	18%	7%
3	500	511	521	526	22%	41%	17%	10%
4	494	504	514	522	9%	29%	35%	23%
5	496	508	515	524	10%	17%	41%	28%
6	493	504	516	522	7%	29%	31%	30%
7	495	508	518	524	9%	25%	29%	32%
8	497	509	520	527	8%	25%	34%	29%
HS	494	500	513	523	4%	20%	39%	32%

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 2. Change to Cut Scores (+/- Resulting from Recommended Minus Current Cut Scores)

	Grade Level									
	K	1	2	3	4	5	6	7	8	HS
EI	-1	-1	-3	-1	-3	-1	-4	-2	-2	0
I	-1	-4	-4	-3	-4	0	-2	1	1	-1
EA	-1	-2	0	0	0	-1	1	1	2	-2
A	-2	-1	-2	-3	1	1	0	0	1	-5

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

2. Overview
2.1. Oregon's English Language Proficiency Assessment (ELPA)
Oregon's English Language Proficiency Assessment (ELPA) provides placement information to educators about student mastery of the English language described by the English language proficiency standards. The ELPA is an online computer-adaptive test designed to measure the content described in the standards. ELPA assesses knowledge and skills using multiple-choice and polytomous spoken and written responses. 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 English Language Proficiency Standards System
Oregon's English Language Proficiency standards system consists of Oregon's English Language Proficiency Standards and English Language Proficiency Levels (i.e., performance standards or "cut scores"). The English language proficiency levels define five levels of proficiency (*Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced*) that students in each grade can demonstrate on the ELPA.

During the 2012-13 standards verification, the high school grade band was modified from four independent high school cut scores (at 9th, 10th, 11th, and 12th grades) for each proficiency level, to a single high school-level cut score for each proficiency level. Grade 11, the year of accountability for high school, was used as the basis for the new ELPA cut scores for high school. This action was based on teacher and stakeholder input that there needed to be a single cut score for each proficiency level for high school.

2.2.1. Oregon's English Language Proficiency Standards
Oregon's English Language Proficiency (ELP) standards define progressive levels of competence in the use of English. The ELP standards were developed by a content panel that convened in February 2003 and were subsequently adopted by the State Board of Education. The standards were posted June 2004. The next anticipated revision of the ELP standards will occur in 2013–2014.

The ELP standards are written as pathways to the Oregon English Language Arts standards and are designed to ensure that students with limited English proficiency develop proficiency in both the English language and the concepts and skills contained in the English Language Arts standards. Oregon's English Language Proficiency standards are available via the state's Searchable Standards Tool that allows you to locate, view, and export standards by subject at http://www.ode.state.or.us/teachlearn/real/standards/

2.2.2. Oregon's English Language Proficiency Levels
Oregon's English Language Proficiency Levels include *Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced.* These proficiency levels define, in terms of performance on the ELPA, what students must do to demonstrate Oregon's English Language Proficiency standards: they specify the English language skills and abilities the student needs in order to be classified in a given proficiency level based on the ELPA, and they set clear benchmarks of progress that reflect differences for students entering school at various grade levels.

In November 2007, CTB/McGraw-Hill worked in collaboration with ODE to set performance standards for the ELPA. Approximately 15 participants convened to set performance standards, which establish how well students must be able to use English language, at six grades. Performance standards for the remaining grades were statistically interpolated. The Oregon State Board of Education adopted the recommended proficiency levels in March 2008, and they were applied to the ELPA during the 2008–2009 school year.

2.2.3. English Language Proficiency Performance Level Descriptors (PLDs)

Oregon's grade-specific Performance Level Descriptors (PLDs) describe the knowledge and skills required by students within each proficiency level (*Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced*) at each grade. The PLDs include general policy definitions that are more general statements that describe rigor across grade levels. The Oregon State Board of Education adopted the current PLDs in March 2008. Prior to the 2012–13 workshops, ODE worked with stakeholders to draft preliminary revisions to the 2008 PLDs. Panelists worked to revise these PLDs during the 2012–13 workshops.

The preliminary PLDs are available on the Oregon Department of Education website at http://www.ode.state.or.us/search/results/?id=223. The PLDs recommended by the November participants are available in Appendix A, the PLDs recommended by the February participants and are in Appendix AC.

3. The 2012–2013 English Language Proficiency Assessment Performance Standards Verification Workshops

The ELPA performance standards were set in two workshops held in November 2012 and February of 2013 using a modified bookmarking standard setting procedure (Cizek & Bunch, 2007; Kane, 1994; Mitzel, Lewis, Patz; & Green, 2001). A similar procedure was implemented for mathematics verification in August 2010, reading in January 2011, and science in July 2011. Seventy-six participants in November and twenty participants in February recommended preliminary achievement standards for grades kindergarten through high school in English language proficiency. ODE English language proficiency (ELPA) consultants and senior staff developed materials, planned the workshop, conducted the training, and led the participants through the workshop. Section 3.2 describes the November workshop and section 3.3 describes the February workshop.

ODE contracted with the Educational Policy Improvement Center (EPIC) to review materials and 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 3.

Table 3. Evaluation Standards and Evidence

Standard	Evidence
Panels should be large enough and representative enough of the appropriate constituencies.	Grade-Level Group Composition, in sections 3.2.2 and 3.3.2.
Selection and qualification of participants should be documented.	Panel Participants, sections 3.2.2 and 3.3.2.
Two panels or subpanels should be used to check the generalizability of the standards.	Grade-Level Group Composition, in sections 3.2.2 and 3.3.2; Placing the Bookmarks, sections 3.2.4 and 3.3.4.
Background and demographic information about participants should be collected and documented.	Grade-Level Group Composition, in sections 3.2.2 and 3.3.2; Appendices D and U.
To ensure internal validity, the methods must be consistent so that ratings indicate increased internal consistency across rounds and panelists.	Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4; Variability, in sections 3.2.4 and 3.3.4.
To ensure procedural validity, the procedures must be reasonable, carried out as intended, and understood by panelists.	The 2012–2013 ELPA Performance Standards Verification Workshops, section 3; Training, sections 3.2.3 and

	3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4; November Training Evaluation Forms, in section 3.2.5; Appendix F.
The methodology should be appropriate for the assessment, described in detail, and field tested when appropriate.	The 2012–2013 ELPA Performance Standards Verification Workshops, section 3.
Any nonstandard methodology must be clearly documented.	The 2012–2013 ELPA Performance Standards Verification Workshops, section 3; ELPA Performance Standards Verification Process Summary, sections 3.2 and 3.3.
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 15, section 3.2.4; Table 30, section 3.3.4; Placing the Bookmarks, sections 3.2.4 and 3.3.4; Target Student Descriptions, in section 3.2.3.
The rationale and procedures for establishing cut scores must be documented.	The 2012–2013 ELPA Performance Standards Verification Workshops, section 3; Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4.
The methods should be designed so that participants can reasonably contribute their knowledge and experience to produce reasonable, defensible standards.	Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4; Table 15, section 3.2.4; Table 30, section 3.3.4; ELPA Performance Standards Verification Process Summary, sections 3.2 and 3.3.
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, sections 3.2.3 and 3.3.3; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5; Workshop Evaluations, Appendices C, J, and R; 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.	ELPA Performance Level Descriptors, section 2.2.3 and Appendices A and I; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5.
The process should be conducted efficiently.	Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections

	3.2.4 and 3.3.4; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5.
Item booklets, rating forms, and other provided documents should be easy to use.	Materials review, in section 3.2.3; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5; Appendices G and H.
For a test of foreign-language speaking ability or musical performance, the judges can listen to the actual performance, or a portion of it (either live or recorded). (Purpose: Helps judges become familiar with test takers' knowledge and skills; gives them a chance to observe a demonstration or an example of the product of each test taker's knowledge and skills.)	Oregon's English Language Proficiency Assessment (ELPA), section 2.1; Placing the Bookmarks, sections 3.2.4 and 3.3.4.
Facilitators should be qualified and capable of leading appropriate discussion among the participants without biasing the process.	ELPA Consultant and Facilitator Training, in section 3.2.3; Grade-Level Group Composition, in sections 3.2.2 and 3.3.2.
Feedback to participants must be clear, understandable, and useful.	Materials Review, in section 3.2.3; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5.
Participants should be instructed on the appropriate use of provided data (including performance data, impact data, criterion reference data, etc.).	Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4; Table 15, section 3.2.4; Table 30, section 3.3.4; ELPA Performance Standards Verification Process Summary, sections 3.2 and 3.3; Introduction, section 1; Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5.
When possible, performance levels should be established using empirical criterion reference data.	Placing the Bookmarks, sections 3.2.4 and 3.3.4; ELPA Performance Standards Verification Process Summary, sections 3.2 and 3.3.
Process evaluations should be conducted and documented.	Process Monitoring and Evaluation, sections 3.2.5 and 3.3.5; Appendices J, Q, R, AB.
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, sections 3.2.2 and 3.3.2; Grade-Level Group Composition, in sections 3.2.2 and 3.3.2; Training, sections 3.2.3 and 3.3.3; Placing the Bookmarks, sections 3.2.4 and 3.3.4.

The November workshop began with orientation, training, and a practice session on setting bookmarks. At the conclusion of the training, participants were asked to complete a training evaluation. The November workshop included three rounds of bookmark placement for grades kindergarten through high school, which entailed a review of impact data based on assessment results from the 2011-12 academic year and bookmark placement across grade-level groups and table teams. The November workshop concluded with a presentation of the final recommendations and corresponding impact data across all grades. This presentation and the ensuing discussion led ODE and EPIC to decide to hold a second workshop with a subset of panelists who would reconvene in February to recommend a final set of cut scores. The processes used throughout the two workshops are documented in detail below. Additionally, materials used in the workshops are provided in the appendices as noted.

3.1 Goals of the Standards Verification Workshops

The goals of the English language proficiency performance standard-setting procedure were as follows:

- Quantify the rigor of each Proficiency Level: *Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced*
- Finalize the Performance Level Descriptors to recommend to the State Board of Education
- Describe student performance and proficiency in a systematic manner that can be used for student descriptive information and student growth in English language proficiency
- Provide information to students, parents, educators, policymakers, and others about what students can know and do in the English language

3.2. November 2012 English Language Proficiency Assessment Performance Standards Verification Process Summary

From November 6 to November 9, 2012, the Oregon Department of Education (ODE) convened a group of educators and stakeholders to participate in a standards verification workshop to recommend performance standards in grades kindergarten through high school on the Oregon English Language Proficiency Assessment (ELPA).

Seventy-six knowledgeable participants, including educators, higher education representatives, parents, and community members, were recruited from across Oregon to participate in groups at grades kindergarten through 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 *Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced* cut scores.

Workshop participants participated in one grade-level group (grades kindergarten through 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 between 52 and 62 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 score references were indicated in the booklets.

Performance standards were set for one performance level at a time within each grade-level group. First, participants bookmarked the *Advanced* performance level, then the *Early Advanced*, followed by the *Intermediate* and *Early Intermediate* performance levels for each grade. In order to set the performance levels, panelists participated in three review rounds in which they individually recommended cut scores during the first round, reached consensus as a table team in the second round, and reached consensus as a grade-level group by the end of the third round. This bookmarking method 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 4. For the purposes of these discussions, the high school impact data was the average of the grades 9 through 12 impact data from the 2011-12 school year. 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 the prior year's test results. As participants reviewed the derived scores and impact data, each grade-level group and the participants as a whole gave careful consideration to the recommended scores. The Manager of Test Design and Implementation introduced these discussions to answer policy-related questions, and the Manager of Psychometrics and Validity summarized the results to panelists and answered technical questions.

During this discussion, participants were not able to reach consensus on a set of final proficiency standards and the grade-level groups indicated that they wished to have more time to meet with other groups for cross-grade discussions. ODE, in consultation with EPIC, decided that a smaller group of panelists should reconvene at a later date to set final bookmarks.

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

3.2.1. November Workshop Agenda
During the first day of the training, ODE described to participants the use of English language 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), emphasizing the state and federal context of English language proficiency assessments. ODE reminded participants of three crucial questions at the core of the ELPA standards verification process:

- When are students proficient enough in English to participate meaningfully in the general education program?

- When exited, what is the evidence that ELL students are performing as well as their non-ELL peers?
- When are ELL students participating in essentially all aspects of the district's curriculum?

ODE explained that this Standards Verification Workshop was an opportunity to apply expert knowledge to set standards and expectations in a clear and transparent way. ODE emphasized that Standards Verification was not an arbitrary discussion, rather it was a systematic process based on expert evaluation of content after in-depth discussion. ODE reminded participants that, while the policy decisions around exiting ELL students are very important, they are not appropriate topics for the workshop. Policy decisions based on judgments outside the assessment system should not play a role in the bookmarking process.

On the second and third days of the workshop, ODE and researchers from the American Institute of Research described two studies conducted on the ELPA:

- The ELPA to ELPA Longitudinal Look Back study, which examined ELPA scores for a cohort of 40,000 students from 2009–10 through 2011–12
- The ELPA to 2011–12 OAKS Reading Comparison study, which examined the performance on both the 2011–12 ELPA and the 2011–12 OAKS Reading for a cohort of 24,183 ELL students

On the fourth day of the workshop, ODE psychometricians presented the results from the Contrasting Groups study, which compared teachers' judgments about which proficiency level students would achieve on the ELPA with the actual proficiency level students scored on the ELPA.

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 English Language proficiency (ELPA) Consultants
Eight English language proficiency consultants were recruited to assist ODE with leading and providing content expertise in the Standards Verification Workshop. These ELPA consultants were external experts who had participated in pre-verification training and assisted with drafting the Performance Level Descriptors (PLDs).

Standards Verification Workshop Participants
Seventy-six Oregonians participated in the November Standards Verification Workshop. The panel was carefully selected to represent Oregon stakeholders, to include ELL teachers, specialists, and coordinators (81%); school administrators (4%); university educators (10%); and parents, business people, and others (3%)[1]. 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,

---

[1] *Note.* Some participants did not respond to this item on workshop evaluation forms.

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

From the 126 individuals who expressed interest in participating, the Department selected 76 participants 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. A small number of participants chose to volunteer their time.

Recruitment criteria are included in Appendix E

Grade-Level Group Composition

The seventy-six November workshop participants were divided into 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, an ELPA 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 4 shows the educational background of participants in each grade-level group.

Table 4. Participant Educational Background by Grade-Level Group

Grades	N	HSD or GED	Bachelor's	Master's	Doctorate
All	66	2%	12%	79%	8%
K	6	0%	0%	50%	33%
1	8	0%	0%	75%	0%
2-3	15	7%	20%	67%	7%
4-5	13	0%	8%	92%	0%
6, 7, 8	20	0%	10%	80%	10%
HS	7	0%	29%	71%	0%

Note. Some participants did not respond to this question.

Table 5 shows the occupation of participants in each grade-level group.

Table 5. Participant Occupation by Grade-Level Group

Grades	N	ELL Teacher	ELL Specialist	ELL Coordinator	Community College/ University Faculty	Administrator / Supervisor of Title III school	Parent of ELL Student	Business Member	Other
All	67	46%	19%	15%	10%	4%	1%	1%	1%
K	6	80%	0%	0%	0%	0%	20%	0%	0%
1	8	25%	13%	50%	13%	0%	0%	0%	0%
2-3	15	47%	27%	13%	13%	0%	0%	0%	0%
4-5	13	46%	23%	8%	8%	8%	0%	0%	8%
6, 7, 8	19	58%	16%	11%	11%	5%	0%	0%	0%
HS	7	14%	29%	14%	14%	14%	0%	14%	0%

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

Table 6 shows the years of work experience for participants in each grade-level group.

Table 6. Years of Work Experience by Grade-Level Group

Grades	N	1-5	6-10	11-15	16-20	21+
All	69	13%	38%	17%	17%	14%
K	6	0%	17%	33%	0%	50%
1	8	25%	38%	0%	25%	13%
2-3	15	20%	27%	20%	27%	7%
4-5	13	15%	46%	23%	8%	8%
6, 7, 8	20	5%	45%	15%	25%	10%
HS	7	14%	43%	14%	0%	29%

Table 7 shows participants' experience teaching special education (SPED), English language learners (ESL/ELD), vocational education, alternative education, and adult education.

Table 7. Participant Teaching Experience with Diverse Populations by Grade-Level Group

Grades	N	SPED	ESL/ELD	Vocational Ed	Alternative Ed	Adult Ed
All	69	17%	94%	4%	9%	46%
K	6	17%	83%	17%	17%	67%
1	8	13%	100%	0%	0%	13%
2-3	15	7%	87%	0%	0%	33%
4-5	13	15%	100%	8%	8%	54%
6, 7, 8	20	25%	95%	0%	15%	50%
HS	7	29%	100%	14%	14%	57%

Note. Participants may have self-reported representation in more than one category

Participant Roles and Responsibilities

Workshop participants included the following:

- ODE staff
- ELPA consultants
- Grade-level group leads
- Table team leads (A/B)
- EPIC evaluators

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.

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

Table team leaders facilitated discussions, 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. Each table team also selected a timekeeper, 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. Two representatives from the American Institute of Research were present to describe the results of the three studies described in section 3.2.1, and to support use of the computer-based technology. In addition, a professor from Oregon State University was present as a non-participatory observer.

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, ELPA 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 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. 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.

Laptops were provided to each participant to use to review digital Ordered Item Booklets. In addition, one laptop was provided to each table for participants to use to listen to recordings of student responses to ELPA items.

3.2.3. Training
ODE staff, including Oregon's Manager of Test Design and Implementation and Manager of Psychometrics and Validity, provided training.

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 *Advanced* performance level (and *Beginning, Early Intermediate, Intermediate,* and *Early Advanced 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 ELPA consultants. At the end of the workshop each day, the ODE staff met with the grade-level group leaders and ELPA 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) review of 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 sample test items, the purpose of the ELPA, the standards setting process, and the Performance Level Descriptors (PLDs) for each performance standard.

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

- Grade level assignment for the workshop
- The Performance Level Descriptors
- The English Language Proficiency Standards
- An article summarizing best practices in performance level descriptor development (Perie, 2008).

The November workshop began with a day-long orientation and training 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.

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 English Language Proficiency Assessment, test items, and Performance Level Descriptors
- Key concepts and materials, including the Ordered Item Booklet (OIB), Ordered Item Map (OIM), scoring rubrics, and the Performance Level Descriptors (PLDs)
- The role of table leaders and group norms
- Protocols for working with secure materials
- The agenda for each day

At the end of the training, participants engaged in a brief, mock standard-setting exercise using released items from the ELPA to ensure task understanding. During this mock standard-setting exercise, participants reviewed and used sample materials including a sample Ordered Item Booklet (OIB), which can be viewed in Appendix G; Ordered Item Map (OIM) and Polytomous Item Rubric, which can be viewed in Appendix H; and the preliminary Performance Level Descriptors (PLDs), which can be viewed in Appendix I.

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

General Overview of English Language Proficiency Assessment
During the first day of the workshop, participants were provided an overview of the ELPA and a description of how assessment scores are used and how changes to cut scores determined throughout the verification process may affect Oregon students and educators. Workshop leaders described the task and the reasons for reviewing the achievement standards. 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 English Language Proficiency Standards and Achievement Standards
During the training, workshop participants reviewed materials including sample Ordered Item Booklets (OIBs), Ordered Item Maps (OIMs), Performance Level Descriptors (PLDs), and the English Language Proficiency Standards. Participants revised 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, Ordered Item Maps, and Polytomous Item Rubrics. 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 current cut scores. Item difficulty was based on operational 2011–12 data. Workshop participants

were not provided the RIT values for items, as the focus was on content and the ordered difficulty.

There was one sample OIB per grade-level group. Each item was presented with an item ID, the item prompt, response options, and the correct response. For polytomous items, a computerized sample student response was included at the score level. Within each OIB, the current cut points for each Proficiency Level were noted on items. Participants had access to both digital and hard copy OIBs. The digital and hard copy OIBs included the same content. Polytomous items (items scored on a two-point rubric) may be shown at two difficulty points, thus appearing up to two times in the OIB. The digital OIBs were available on laptops that were individually assigned to each participant. In addition, one laptop was provided to each table for participants to use to listen to recordings of student responses to ELPA items.

For open-ended items, sample writing and speaking student responses were available. Writing response student samples are provided in the print version; speaking response samples are provided via laptop for the digital OIB and transcribed in the hard-copy OIB. Because recorded speaking student responses were only available on one laptop, all participants from each table listened to student responses as a group before beginning their review of the digital OIB.

The Ordered Item Maps contained the page number of each item in the OIB, the current Oregon cut scores, the Oregon item ID, the stimulus ID (if applicable), the Domain/Item type, the answer key, and a column for participant notes.

The Polytomous Item Rubric contained the scoring criteria for speaking and writing extended response ELPA items. These items were scored on grammar and illocution. Participants were expected to use the Polytomous Item Rubric alongside their OIBs to inform their understanding of the expectations for each test item and to aid them in making decisions about where to place their bookmark.

Appendices G and H include sample Ordered Item Booklets, Ordered Item Maps, and the Polytomous Item Rubric.

Performance Level Descriptors. Prior to the standard setting workshop, ODE convened a panel of experts to update the Performance Level Descriptors (PLDs) for each of the following proficiency levels at each grade/grade band: *Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced.*

The PLDs were updated such that each of the five proficiency levels differentiated student performance in terms of increasing cognitive demand and task complexity. During the November workshop, participants provided revisions to the provided PLDs based on the newly recommended cut scores.

Appendix I contains the preliminary Performance Level Descriptors for each grade level

provided to workshop participants. Appendix A contains the revised PLDs recommended at the end of the November workshop.

Target Student Descriptions. After lunch on the first day, ODE led the participants in an exercise to revise Target Student Descriptions (TSD). The Target Student Descriptions depict the minimum knowledge and skills that a student must demonstrate on the ELPA in order to "just barely" reach each Proficiency Level.

Prior to the November workshop, ODE drafted one set of Target Student Descriptions for all of the groups to modify and use. After ODE trained participants, the table team leader facilitated a discussion to help participants articulate what a Target Student could demonstrate at each of the proficiency levels. Participants visualized Target Students for each proficiency level using the appropriate PLDs and their own expert judgment. Once adjustments were made to the draft Target Student Descriptions at the *Advanced* proficiency level, participants made modifications to the *Intermediate* and *Beginning* levels. Participants then made modifications to the *Early Advanced* and *Early Intermediate* Target Student Descriptions. ELPA consultants provided content expertise as participants developed the Target Student Descriptions.

Participants were asked to refer to the Target Student Descriptions throughout the standards verification process. Once finalized, characteristics of Target Students at each proficiency level were recorded and posted near each table. These Target Student Descriptions served as a basis for establishing a common understanding of the type of student that should be considered when setting each cut score.

Appendix K contains the presentation and instructions for developing Target Student Descriptions. Appendix L contains each grade-level group's Target Student Descriptions.

Bookmark Placement

Each participant 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 reading consultants answered questions.

Participants were instructed to use the following tools when placing their bookmarks: the English Language Proficiency Standards, their group's Target Student Descriptions (TSDs), the Performance Level Descriptors (PLDs), the content as represented by the items in the Ordered Item Booklets (OIBs) and Ordered Item Maps (OIMs), sample student responses, and current cut scores.

Workshop participants were instructed to place their bookmarks with the understanding 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 between the last item in one level and in front of the first item

in the higher level, such that their placement identified the point at which students minimally should know and be able to do the item at each proficiency level. After the cut score, students then would fall into the level defined by that cut score. Participants were instructed to begin by placing the *Advanced* bookmark, then the *Early Advanced*, followed by the *Intermediate* and *Early Intermediate* bookmarks.

ELPA Consultant and Facilitator Training
Prior to the Standards Verification Workshop, ODE staff leading the workshop provided a full-day training for the ELPA 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, Target Student Descriptions, and Ordered Item Maps; discussed the technology used at the workshop; and summarized the workshop goals. The ELPA 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). Workshop participants placed the bookmarks at the location in the OIB where the Target Student defined for that level had a 67% chance of correctly responding to the item at that location.

Prior to Round One, participants reviewed the instructions for the bookmarking process, the PLDs, and the OIBs in order 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. Because recorded student responses were only available on one laptop, all participants from each table listened to student responses as a group before beginning their review of the digital Ordered Item Booklet (OIB).

In Round One, participants worked independently to place bookmarks for the *Advanced*, *Early Advanced*, *Intermediate,* and *Early Intermediate* Proficiency Levels for all grades (kindergarten through 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-level panels completed Round Three, ODE psychometric staff analyzed the longitudinal student progression from kindergarten through high school. To ensure internal validity, the methods were consistent through all rounds so that ratings indicate increased internal consistency across rounds and panelists (NAGB, 2010).

Round One
During Round One, participants worked independently to review the OIB (either on the laptops or the paper version) and determine individual bookmarks for their grade level. Upon completion of the task, ODE summarized and presented the Round One median OIB page numbers. After this round, and each subsequent round, ODE psychometricians provided the impact data to participants, which is the percent of students who took the ELPA in 2011–2012 who would fall into each performance level category if the Round One bookmarks were adopted.

Results of Round One are summarized in Tables 8 and 9 and provided in Appendix M.

Table 8. Round One Median Bookmark Placement by Grade-Level Group

					Grade Level					
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	7	8	12	14	14	9	5	7	12	11
I	16	22	28	34	26	24	18	22	24	19
EA	24	37	38	48	38	35	35	41	35	32
A	37	46	49	57	49	47	48	52	48	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 9. Round One Impact Data by Grade-Level Group

					Grade Level					
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
B	13.2%	10.4%	13.2%	6.7%	6.9%	3.2%	2.7%	3.9%	3.9%	4.9%
EI	24.1%	17.1%	40.2%	30.7%	15.2%	9.2%	5.7%	7.8%	6.8%	4.0%
I	25.7%	36.1%	18.8%	35.1%	24.3%	14.9%	30.4%	36.9%	12.9%	15.4%
EA	24.1%	21.3%	22.6%	23.4%	38.0%	34.9%	34.6%	32.8%	27.3%	44.0%
A	12.9%	15.1%	5.1%	4.1%	15.6%	37.7%	26.6%	18.6%	49.1%	31.6%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

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. 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.

Results of Round Two are summarized in Tables 10 and 11 and provided in Appendix N.

Table 10. Round Two Median Bookmark Placement by Grade-Level Group

					Grade Level					
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	8	6	13	15	12	10	6	8	12	11
I	18	19	28	30	25	24	22	26	24	18
EA	28	31	40	44	33	35	35	44	35	32
A	34	47	49	54	49	47	48	56	48	42

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 11. Round Two Impact Data by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
B	13%	8%	13%	8%	6%	4%	3%	4%	4%	5%
EI	31%	14%	40%	24%	17%	9%	9%	11%	7%	3%
I	26%	26%	22%	32%	16%	15%	26%	44%	13%	16%
EA	13%	39%	19%	27%	46%	35%	35%	36%	27%	36%
A	17%	14%	5%	10%	16%	38%	27%	6%	49%	40%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Round Three

The workshop participants worked in grade-level groups for Round Three to reach a group consensus on bookmark placement. Participants reported increased confidence in their bookmarks after Round Three. ODE analysts presented the impact data from the Round Two bookmarks. 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 12 and 13 and provided in Appendix O.

Table 12. Round Three Median Bookmark Placement by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	10	8	14	15	14	10	6	8	10	11
I	16	21	24	30	27	24	20	28	22	17
EA	27	33	40	47	40	36	35	39	35	32
A	37	47	44	55	49	47	43	50	48	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 13. Round Three Impact Data by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
B	19%	10%	15%	8%	7%	4%	3%	4%	4%	5%
EI	18%	17%	24%	24%	18%	9%	7%	13%	5%	3%
I	29%	28%	37%	42%	26%	15%	29%	26%	15%	17%
EA	21%	31%	11%	20%	34%	35%	26%	31%	27%	44%
A	13%	14%	13%	7%	16%	38%	36%	26%	49%	32%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Round Four: Cross-Grade Articulation (Smoothing)

The Round Three bookmark placement and associated impact data were presented to the participants during the cross-grade articulation, or "smoothing," discussion on Day 4. The purpose of this smoothing discussion was to establish a system of bookmarks that would result in cut scores that were 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 bookmark placements, responding to the prompt: "If there is more than a 5% difference in impact with that of the adjacent grades, please interpret this/these "larger" differences."

Table 14 shows the difference in impact for adjacent grades with differences greater than 5% in bold font.

Table 14. Difference in Impact for Adjacent Grades After Round Three

	Grade Level								
	K-1	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**	**7-8**	**8-HS**
B	**9.0%**	4.2%	**7.1%**	0.6%	3.3%	0.4%	1.0%	0.3%	1.0%
EI	0.8%	**6.9%**	0.5%	**5.7%**	**8.9%**	2.2%	**6.3%**	**8.1%**	2.2%
I	1.4%	**9.3%**	4.3%	**15.7%**	**10.9%**	**14.0%**	2.5%	**11.5%**	1.8%
EA	**10.4%**	**19.8%**	**8.9%**	**13.8%**	1.0%	**9.3%**	**5.2%**	3.5%	**16.7%**
A	0.6%	0.5%	**5.6%**	**8.2%**	**22.1%**	2.1%	**10.0%**	**23.5%**	**17.5%**

Note. Differences greater than 5% are in bold font.

The largest differences in impact across adjacent grades were:
- The 7–8 *Advanced* scores with a 23.5% difference,
- The 4–5 *Advanced* scores with a 22.1% difference, and
- The 1–2 *Early Advanced* scores with a 19.8% difference.

Despite these large differences, several groups responded that they wished to make no revisions but would like to discuss the larger differences with the participants in the adjacent grade levels. Due to time constraints, ODE offered participants the opportunity to reconvene at a later date to continue the smoothing discussion and set final bookmarks. A description of the process and results from the follow-up workshop are presented in section 3.3.

Variability
As panelists discussed their reasons for placing bookmarks and impact data, variability across tables and individuals often decreased over the rounds of decision making. Taking the standard deviations across bookmark placements for individuals within a grade level provides a measure of variability across participants at each round. Variability does decrease with each round, to zero in Round Three for all but two grade-level groups (grade 4 and HS). In other words, the participants in all panels other than grade four and high school were able to reach consensus on a final bookmark placement by the end of Round Three.

Individual bookmarks for each panelist are presented in Appendix P and are summarized in Table 15.

Table 15. Standard Deviations and Ranges for Individual *Advanced* Bookmark Placement in Each Round

	Round 1		Round 2		Round 3	
	SD	Page Range	*SD*	Page Range	*SD*	Page Range
Kindergarten	4.03	29-40	2.67	32-37	0	37-37
Grade 1	6.22	32-47	0	47-47	0	47-47
Grade 2	3.01	43-50	2.33	43-50	0	44-44
Grade 3	3.30	51-62	2.05	51-58	0	55-55
Grade 4	3.46	45-53	2.70	47-53	1.51	49-53
Grade 5	1.21	45-49	0.38	46-47	0	47-47
Grade 6	6.21	42-57	4.71	47-57	0	43-43
Grade 7	3.67	47-57	1.64	54-57	0	50-50
Grade 8	0.76	47-49	0.35	47-48	0	48-48
HS	2.30	39-46	3.74	39-46	3.78	34-44

Revision of the Performance Level Descriptors

Throughout the workshop, participants were given time to refine the Performance Level Descriptors (PLDs). During this revision, workshop participants were encouraged to review the PLDs to be consistent with their recommended cut points and the content of the OIB. Revised PLDs are presented in Appendix A.

November Workshop Conclusion

The workshop concluded with recommendations from ODE regarding how participants could 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 recommendations that were made using the grade-level bookmarks and impact data. 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 new 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; NAGB, 2010).

All of the above were used 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, materials, or secure test items, and some participants used the cards to note issues or questions that may be important but were not directly relevant to the standards verification task.

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

November Training Evaluation Forms

At the completion of training and prior to beginning Round One, participants completed a training evaluation comprising 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 J.

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

- 92% of participants agreed or strongly agreed with the statement, "The training materials were helpful."
- 86% of participants agreed or strongly agreed with the statement, "I am confident I understand my role in the standards verification process."
- 83% 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 initially feel confident in being prepared for the task, later evaluations and interviews indicate that participants felt much more confident and prepared following engagement in the tasks during Round One.

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

November Workshop Participant Interviews

On days 2, 3, and 4, 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 the formal interviews are provided in Appendix Q.

Formal Interviews

Twelve 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.

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

- Interviewed participants reported that the training prepared them for the task. About half of interviewed participants expressed that they felt the training was too repetitive, while the other half reported that they appreciated the repetition.
- Interviewed participants reported that they felt confident with the process of the workshop.
- Interviewed participants reported that their opinions were considered and valued by their groups and that the groups worked very well together.

Informal Interviews

Informal interviews were conducted with workshop participants throughout the workshop to elicit feedback from participants about their progress in reaching consensus. Interviews were conducted individually; four participants were randomly selected and approached during break times.

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

- Participants interviewed on day three reported that their groups were very close to reaching consensus.
- Participants reported that they felt comfortable with how to place their bookmarks.
- Participants reported that the groups were respecting their established norms.

November Workshop Evaluation Forms

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

Generally, feedback was positive and included the following:
- 96% of participants agreed or strongly agreed with the statement, "I understood how to place my bookmarks."
- 90% of participants agreed or strongly agreed with the statement, "Overall, I am satisfied with my group's final bookmarks."
- 99% of participants agreed or strongly agreed with the statement, "Overall, I think my table's discussions were open and honest."

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 R and the results are summarized below.

Overall, participants placed importance on the following:

- Participants placed the most importance on visualizing a Target Student, with 90% responding that it was important or very important to their bookmark placement.
- Participants rated the importance of panel discussions next, with 88% responding that they were important or very important to their bookmark placement.
- Participants placed the least importance on their perceptions of the quality of the sample student responses, with 52% responding that they were important or very important to their bookmark placement.

Table 16. Importance of Factors Used to Place Bookmarks.

Factor	N	N/A	Not Important	Somewhat Important	Important	Very Important	Important + Very Important
The Performance Level Descriptors (PLDs) of *Beginning, Early Intermediate, Intermediate, Early Advanced,* and *Advanced.*	68	1%	0%	11%	20%	68%	87%
Your perceptions of the difficulty of the items in the Ordered Item Booklet.	68	3%	3%	10%	37%	48%	85%
Your perceptions of the quality of the sample student responses.	68	3%	18%	27%	34%	18%	52%
Your own classroom experience.	68	11%	0%	3%	18%	68%	86%
Visualizing a Target Student.	68	6%	0%	4%	17%	73%	90%
The impact data.	68	6%	10%	20%	31%	34%	65%
The three research studies that were presented.	68	0%	11%	18%	34%	37%	70%
Your initial classification of student performance in Round One.	68	3%	0%	20%	43%	33%	77%
Panel discussions.	68	7%	0%	4%	25%	64%	88%
The initial classifications of other panelists.	68	3%	6%	13%	36%	42%	78%

Process Check-Ins

At the end of each day, ODE staff met with the ELPA 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.

3.3 February 2013 English Language Proficiency Assessment Performance Standards Verification Process Summary

On February 12 and 13, ODE reconvened a smaller group of twenty educators and stakeholders to recommend a final set of ELPA performance standards in grades kindergarten through high school. The purpose of this second workshop was to continue the smoothing discussion and establish a system of grade-level bookmarks that were well articulated and, at the same time, considerate of the participants' original recommendations. All participants reviewed the data recommended in the November workshop within grade-level groups, with participants from adjacent grade levels, and as a whole group. They also considered impact data, an analysis which forecasts the potential percentages of students falling into each proficiency level at each grade based on the prior year's test results.

As in the first workshop, performance standards were set for one performance level at a time within each grade-level group. First, participants bookmarked the *Advanced* performance level, then the *Early Advanced*, followed by the *Intermediate* and *Early Intermediate* performance levels for each grade. In order to set the performance levels, panelists participated in three review rounds: in the first (Round Four) they reviewed their recommended bookmark placements that were set during the first workshop; in the second, they worked with participants from adjacent grades (Round Five (a)); in the third, they worked as grade bands (Round Five (b)), and reached consensus as a larger group by the end of the workshop (Round Six).

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

3.3.1 Workshop Agenda

During the morning of the first day of the February workshop, ODE reminded participants of the goals of the workshops and the steps involved with the bookmarking process. ODE stated that the goal of this follow-up workshop was to encourage collaboration across grades and reiterated that standards verification is a K–12 collaborative process.

The February workshop agenda is provided in Appendix S and the February workshop presentations are provided in Appendix T.

3.3.2 Panel Participants
ODE Staff and English Language Proficiency Consultants

Five of the English language proficiency (ELPA) consultants returned to assist ODE with leading and providing content expertise in the February Standards Verification Workshop. These ELPA consultants were external experts who had participated in pre-verification

training, the November workshop, and assisted with drafting the Performance Level Descriptors (PLDs).

Standards Verification Workshop Participants

Twenty participants from the November Standards Verification Workshop returned in February. The panel was carefully selected from the November participants from each grade-level group who had responded that they were willing to reconvene. This panel included ELL teachers, specialists, and coordinators (90%); school administrators (5%); and business people (5%). 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 U and below.

Grade-Level Group Composition

The twenty February workshop participants were divided into ten grade-level groups. Each group was assigned an ELPA consultant and an ODE representative who facilitated the discussion but had no input in bookmark placement.

Appendix U 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 17 shows the educational background of participants in each grade-level group.

Table 17. Participant Educational Background by Grade-Level Group

Grades	N	HSD or GED	Bachelor's	Master's	Doctorate
All	19	0%	11%	80%	11%
K	2	0%	0%	50%	50%
1	2	0%	0%	100%	0%
2-3	4	0%	0%	100%	0%
4-5	4	0%	25%	75%	0%
6, 7, 8	5	0%	0%	80%	20%
HS	2	0%	50%	50%	0%

Note. Some participants did not respond to this question.

Table 18 shows the occupation of participants in each grade-level group.

Table 18. Participant Occupation by Grade-Level Group

	N	ELL Teacher	ELL Specialist	ELL Coordinator	Community College/ University Faculty	Administrator	Parent of ELL Student	Busi-ness Member	Other
All	19	79%	11%	0%	0%	5%	0%	5%	0%
K	2	100%	0%	0%	0%	0%	0%	0%	0%
1	2	100%	0%	0%	0%	0%	0%	0%	0%
2-3	4	75%	25%	0%	0%	0%	0%	0%	0%
4-5	4	75%	0%	0%	0%	25%	0%	0%	0%
6, 7, 8	5	80%	20%	0%	0%	0%	0%	0%	0%
HS	2	50%	0%	0%	0%	0%	0%	50%	0%

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

Table 19 shows the years of work experience for each grade-level group.

Table 19. Years of Work Experience by Grade-Level Group

Grades	N	1-5	6-10	11-15	16-20	21+
All	19	16%	37%	16%	26%	5%
K	2	50%	50%	0%	0%	0%
1	2	0%	50%	0%	50%	0%
2-3	4	25%	0%	25%	50%	0%
4-5	4	0%	25%	50%	0%	25%
6, 7, 8	5	10%	40%	0%	40%	0%
HS	2	0%	100%	0%	0%	0%

Note. Some participants did not respond to this item.

Table 20 shows participants' experience teaching special education (SPED), English language learners (ESL/ELD), vocational education, alternative education, and adult education.

Table 20. Participant Teaching Experience with Diverse Populations by Grade-Level Group

Grades	N	SPED	ESL/ELD	Vocational Ed	Alternative Ed	Adult Ed
All	19	5%	100%	5%	11%	42%
K	2	0%	100%	0%	0%	0%
1	2	0%	100%	0%	0%	0%
2-3	4	0%	100%	0%	0%	50%
4-5	4	0%	100%	25%	25%	50%
6, 7, 8	5	20%	100%	0%	20%	60%
HS	2	0%	100%	0%	0%	50%

Note. Some participants may have self-represented in more than one category.

Participant Roles and Responsibilities
Workshop participants included the following:

- ODE staff
- ELPA consultants
- Grade-level group participants

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.

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

Two 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 were given their previously established norms and identified rules to follow to facilitate collaboration and efficiency. These November norms for each table team were posted on the wall near each table and remained visible throughout the workshop. As needed, ELPA consultants and ODE staff reminded table teams of the norms.

The February grade-level group norms are provided in Appendix V.

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 room near the meeting room during breaks.

Unlike the November workshop, no laptops were provided to each participant to use to review digital Ordered Item Booklets or to use to listen to recordings of student responses to ELPA items. Additionally, no sample student responses were included in the OIBs, based on feedback from participants during the November workshop.

3.3.3 Training

ODE staff, including Oregon's Manager of Test Design and Implementation and Manager of Psychometrics and Validity, provided training.

ODE staff reminded the panelists on how to use the bookmark method, the English Language Proficiency Standards and Assessment, and the other materials necessary for recommending performance standards. ODE reminded panelists of the concept of Target Students, who are just barely able to complete the work at the *Advanced* Performance Level (and *Beginning, Early Intermediate, Intermediate,* and *Early Advanced levels*).

February training presentations are included in Appendix T.

Workshop Participant Training Overview

The February workshop began with a refresher that included a review of the purpose for reviewing the cut scores and the workshop agenda.

The training covered the following topics:

- The purpose and goals of the Standards Verification Follow-Up Workshop
- A general overview of standard setting and refresher training on the bookmark procedure
- Reorientation to Oregon's English Language Proficiency Assessment, test items, and Performance Level Descriptors
- Key concepts and materials, including the Ordered Item Booklet (OIB), Ordered Item Map (OIM), and the Performance Level Descriptors (PLDs)
- The group norms
- Protocols for working with secure materials
- The agenda for each day

General Overview of English Language Proficiency Standards and Achievement Standards
During the training, workshop participants reviewed materials, including sample Ordered Item Booklets (OIBs), Ordered Item Maps (OIMs), updated and revised Performance Level Descriptors (PLDs), and the English Language Proficiency Standards. Participants reviewed Target Student Descriptions and were trained on bookmark placement. The OIBs, OIMs, Target Student Descriptions, and English Language Proficiency Standards were the same as those used in the November workshop. Those materials are discussed in section 3.2.3. Materials that differed from the November to the February workshops are discussed in more detail below.

Ordered Item Booklets
A few items in the OIBs were replaced between the November and February workshops in order to better represent the content and level of difficulty of the ELPA. These item replacements were based on feedback from participants and the ELPA consultants during the November workshop. The revised OIBs were equivalent to those used in the original OIBs in terms of psychometric properties and the ELPA content measured by the items in each booklet.

Performance Level Descriptors

The revisions that participants made to the PLDs during the November workshop were concatenated and the updated drafts were given to the February participants. Participants were instructed that they could make small edits to the PLDs during the February workshop; however, ODE elected not to allocate time on the agenda to continue to edit the PLDs. ODE reminded participants that the participants' edits are suggestions for changes that would be approved by ODE prior to Board review and adoption.

3.3.4 Placing the Bookmarks

The panel followed the bookmarking standard-setting method (Kane, 1994; Mitzel, Lewis, Patz, & Green, 2001). Workshop participants placed the bookmarks at the location in the OIB where the Target Student defined for that level had a 67% chance of correctly responding to the item at that location.

Prior to bookmarking, participants reviewed the instructions for the bookmarking process, the PLDs, 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.

First, ODE presented participants' Round Three recommendations and impact data from the November workshop. Participants then discussed their bookmark placement in their table teams. Participants then engaged in subsequent rounds of bookmarking, Rounds Four through Six.

Round Four

In Round Four, participants worked in grade-level groups to review their Round Three bookmark placement through the lens of the refined OIBs and were asked to place individual and grade-level consensus bookmarks. These Round Four bookmarks and the associated impact data were discussed as a large group to help participants understand the cross-grade impact of their bookmark placement.

Results of Round Four are summarized in Tables 21 and 22 and provided in Appendix W.

Table 21. Round Four Median Bookmark Placement by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	10	8	13	17	10	9	6	8	11	13
I	23	23	25	30	19	25	20	24	27	20
EA	32	33	40	47	33	37	35	38	39	34
A	41	47	44	55	47	47	43	48	49	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 22. Round Four Impact Data by Grade-Level Group

	K	1	2	3	4	5	6	7	8	HS
					Grade Level					
B	19%	10%	13%	9%	5%	3%	3%	4%	4%	6%
EI	40%	17%	29%	22%	6%	11%	7%	9%	9%	4%
I	18%	28%	33%	33%	28%	17%	29%	30%	30%	20%
EA	17%	31%	11%	11%	39%	31%	26%	20%	20%	39%
A	6%	14%	13%	13%	23%	38%	36%	37%	37%	32%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Round Five

In Round Five (a), participants worked with members of adjacent grade-level groups (e.g., one member of the first-grade team worked with a member of the kindergarten team, the other first-grade team member worked with a member of the second-grade team, etc.). These adjacent-grade teams reviewed the Round Four bookmarks in the context of ensuring a smooth longitudinal student progression from kindergarten through high school. After meeting with the adjacent-grade team member, participants returned to their grade-level groups and placed Round Five (a) Bookmarks.

Results of Round Five (a) are summarized in Table 23 and provided in Appendix X.

Table 23. Round Five (a) Median Bookmark Placement by Grade-Level Group

	K	1	2	3	4	5	6	7	8	HS
					Grade Level					
EI	10	8	10	17	9	12	6	8	11	13
I	23	26	25	30	25	25	20	24	27	20
EA	32	38	40	47	36	37	35	36	41	34
A	41	47	44	54	47	49	47	48	53	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 24. Round Five (a) Impact Data by Grade-Level Group

	K	1	2	3	4	5	6	7	8	HS
					Grade Level					
B	19%	10%	11%	9%	5%	4%	3%	4%	4%	6%
EI	40%	27%	32%	22%	18%	10%	7%	9%	8%	4%
I	18%	26%	33%	41%	20%	17%	29%	21%	25%	20%
EA	17%	23%	11%	17%	35%	41%	35%	28%	38%	39%
A	6%	14%	13%	10%	23%	28%	27%	37%	25%	32%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

In Round Five (b), participants worked as grade bands (K–1, 2–3, 4–5, 6–7–8–HS[2]) to discuss the impact data from Round Five (a) and move toward better articulated bookmarks. Participants then set Round Five (b) bookmarks as a grade band. The Round Five collaboration was effective—using their professional judgment and the OIBs, participants reviewed and in some cases revised their bookmarks after meeting with adjacent grades and in their grade bands.

Results of Round Five (b) are summarized in Table 25 and provided in Appendix Y.

Table 25. Round Five (b) Median Bookmark Placement by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	10	8	10	17	9	12	6	8	11	13
I	22	26	25	30	21	25	20	24	27	20
EA	32	40	40	47	36	37	35	37	40	34
A	41	47	47	54	47	49	46	49	53	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 26. Round Five (b) Impact Data by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
B	19%	10%	11%	9%	5%	4%	3%	4%	4%	6%
EI	36%	27%	32%	22%	9%	10%	7%	9%	8%	4%
I	21%	32%	33%	41%	29%	17%	29%	25%	25%	20%
EA	17%	17%	18%	17%	35%	41%	31%	29%	38%	39%
A	6%	14%	7%	10%	23%	28%	30%	32%	25%	32%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Round Six

In Round Six, all participants met as one large group to discuss and review Round Five bookmark placement. During this discussion, ODE psychometricians used an interactive tool to illustrate how changing bookmark placement would affect the vertical articulation and impact data. During this discussion, some groups revised their bookmarks in order to set final bookmarks that were well articulated across the grade levels.

Results of Round Six are summarized in Tables 27, 28, and 29 and provided in Appendix Z.

---

[2] Although not in the same grade band, high school was grouped with the 6–7–8 grade band for the purposes of this collaboration.

Table 27. Round Six Median Bookmark Placement by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	10	8	10	17	9	12	6	8	11	13
I	22	26	24	30	21	25	20	24	27	20
EA	32	40	40	47	36	37	35	37	41	34
A	41	49	47	54	47	49	46	49	52	44

Note. EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 28. Round Six Impact Data by Grade level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
B	19%	10%	11%	9%	5%	4%	3%	4%	4%	6%
EI	36%	27%	28%	22%	9%	10%	7%	9%	8%	4%
I	21%	32%	37%	41%	29%	17%	29%	25%	25%	20%
EA	17%	21%	18%	17%	35%	41%	31%	29%	34%	39%
A	6%	9%	7%	10%	23%	28%	30%	32%	29%	32%

Note. B = Beginning, EI = Early Intermediate, I = Intermediate, EA = Early Advanced, A = Advanced

Table 29. Participant-Recommended ELPA Cut Scores by Grade-Level Group

	Grade Level									
	K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**HS**
EI	481	491	492	500	494	496	493	495	497	494
I	491	503	504	511	504	508	504	508	509	500
EA	497	512	514	521	514	515	516	518	520	513
A	505	522	521	526	522	524	522	524	527	523

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 a grade level provides a measure of variability across participants at each round. For the most part, variability decreases with each round to zero for Round Six as the groups reached consensus on final bookmark placement.

Individual bookmarks for each panelist are presented in Appendix AA and are summarized in Table 30.

Table 30. Standard Deviations and Ranges for Individual *Advanced* Bookmark Placement in Each Round.

	Round 4		Round 5a		Round 5b		Round 6	
	SD	Page Range	*SD*	Page Range	*SD*	Page Range	*SD*	Page Range
Kindergarten	0	41-41	0	41-41	0	41-41	0	41-41
Grade 1	0	47-47	0	49-49	0	49-49	0	49-49
Grade 2	0	44-44	0	44-44	0	47-47	0	47-47
Grade 3	0	55-55	0	54-54	0	54-54	0	54-54
Grade 4	0	47-47	0	47-47	0	47-47	0	47-47
Grade 5	0	47-47	0	49-49	0	49-49	0	49-49
Grade 6	2.8	43-47	0	46-46	0	46-46	0	46-46
Grade 7	0	48-48	0	48-48	0	48-48	0	49-49
Grade 8	0	49-49	0	53-53	0.7	52-53	0	52-52
HS	0	44-44	0	44-44	0	44-44	0	44-44

February Workshop Conclusion

The workshop concluded with recommendations from ODE regarding how participants could 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 bookmarks or impact data. 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 data, 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).

3.3.5 Process Monitoring and Evaluation

In order to ensure procedural and internal validity, February workshop participants and leaders were provided with opportunities to evaluate the process using process check-ins, informal interviews, and workshop evaluations (recommended by Hambleton, 2001; NAGB, 2010).

All of the above were used 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, materials, or secure test

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

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

February 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 AB, and results are summarized below.

Feedback was positive and included the following:
- 100% of participants agreed or strongly agreed with the statement, "My group had enough time to consider our final bookmarks."
- 100% of participants agreed or strongly agreed with the statement, "I understood how to place my bookmarks."
- 95% of participants agreed or strongly agreed with the statement, "Overall, I am satisfied with my group's final bookmarks."

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 AB, and results are summarized below.

Overall, participants placed the most importance on the following:
- Participants placed the most importance on their perceptions of the difficulty of the items in the Ordered Item Booklet, with 90% responding that it was important or very important to their bookmark placement.
- Participants placed equivalent importance on their classroom experience, grade-level discussions, and cross-grade collaboration, with 85% responding that they were important or very important to their bookmark placement.
- Participants placed the least importance on the impact data, with 79% responding that it was important or very important to their bookmark placement.

Table 31. Importance of Factors Used to Place Bookmarks.

Factor	N	Not Applicable	Not Important	Somewhat Important	Important	Very Important	Important + Very Important
The Performance Level Descriptors.	20	5%	0%	15%	45%	35%	80%
Your perceptions of the difficulty of the items in the Ordered Item Booklet.	20	0%	5%	5%	35%	55%	90%
Your own classroom experience.	20	10%	0%	5%	20%	65%	85%
Visualizing a Target Student.	20	5%	0%	15%	25%	55%	80%
The impact data.	20	0%	0%	21%	47%	32%	79%
Your initial classification of student performance in Phase One.	20	0%	10%	10%	60%	20%	80%
Grade-level discussions.	20	10%	0%	5%	10%	75%	85%
Cross-grade collaboration.	20	10%	0%	5%	15%	70%	85%

Process Check-ins

At the end of each day, ODE staff met with the ELPA 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.

4. Formal Adoption of English Language Proficiency Standards

The State Board of Education will consider adoption of the cut scores and proposed Performance Level Descriptors (available in Appendix AC) on May 16, 2013.

References

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

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, NJ: Lawrence Earlbaum Associates.

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

Luecht, R. M., & Ackerman, T. (2007). *Oregon English Language Proficiency Examination (ELPA) Dimensionality Analysis for Blended-Domain Locator Blocks.* Greensboro, NC: Center for Assessment Research and Technology.

Mitzel, H.C., Lewis, D.D., Patz, R.J., & Green, D.R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249–281). Mahwah, NJ: 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), 15–29.

Thompson, K. D. (2012). *Time to English proficiency and relationship between English proficiency and academic achievement: Analysis of long-term linguistic and academic outcomes for English learners.* Paper presented at the meeting of the American Educational Research Association, Vancouver, BC, Canada.

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, DC: 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.