



ELPA21 Standard Setting Technical Report

Submitted by



Table of Contents

Table of Contents	2
Tables and Figures	4
1. Executive Summary	6
2. Introduction	9
2.1 Background	9
2.2 Standard Setting Design Elements	11
3. Contrasting Groups Study	15
3.1 Implementation	15
3.2 Materials and Survey Methods	15
3.3 Training and Data Collection	17
3.4 Data Analyses and Results	17
3.5 Use of Results	22
4. English Learner Expert Advisory Panel	23
4.1 Proficiency Determination Profile Rules	23
4.2 Target Student Descriptor Writing	24
4.3 Achievement Level Descriptor Writing	26
5. In-Person Bookmark Standard Setting Workshop	28
5.1 Overview of the In-Person Bookmark Standard Setting	28
5.2 Item Mapping: Locating Items on the IRT Scale	29
5.3 Preparation of Materials	30
5.4 Room Configuration	35
5.5 Workshop Roles	35
5.6 Workshop Events	39
5.7 Results	55
5.8 Reliability and Degree of Consensus of Panelists' Recommendations	56
5.9 Domain Evaluations	58
5.10 Proficiency Determination Review and Recommendations	63
6. Adopting Cut Scores, Proficiency Determination Rules, and ALDs	66
6.1 Adopted Cut Scores	66
6.2 Adopted Proficiency Determination Rules	72
6.3 Finalizing Achievement Level Descriptors	79

7. Validity	80
7.1 Adherence to Professional Standards 1.....	80
7.2 Best Practices	81
7.3 Kane’s Standard Setting Validity Criteria	82
7.4 Auditor’s Report	84
7.5 Evaluations.....	84
7.6 Conclusions.....	86
8. References	87
9. Appendices	88

Tables and Figures

Table 1. Policy Descriptors for Domain Performance.....	10
Table 2. Policy Descriptors for the Proficiency Determination.....	11
Table 3. Number of Teachers Responding to CGS and Total Number of Students Classified.....	18
Table 4. Comparison of State and CGS Impact Data	19
Table 5. Teacher Endorsements for the Number of Achievement Levels.....	20
Table 6. Seventy-Five Most Common Profiles for All Grades (Profile Order: Overall, Reading, Writing, Listening, Speaking)	21
Table 7. Policy Descriptors for Level 3 and Level 4*	26
Table 8. Grades in Which the Level 3 and Level 4 Cut Scores were Set Directly by Panelists	29
Table 9. Number of Items and Score Points in the Ordered Item Booklets.....	31
Table 10. Pacific Metrics Staffing Roles and Experience	35
Table 11. Standard Setting High-Level Agenda.....	40
Table 12. Post-Opening-Session Readiness Survey, Questions 10-8	43
Table 13. Post-Opening-Session Readiness Survey, Questions 9-10	43
Table 14. Bookmark Training Readiness Survey and Results.....	49
Table 15. Bookmark Training Readiness Survey and Results.....	49
Table 16. Round 3 Results and Associated Impact Data	55
Table 17. Standard Errors Associated with the Sampling of Panelists.....	57
Table 18. Average Range and Standard Deviation of Panelists' Scale Score Cut Score Recommendations by Domain, Achievement Level, and Round	57
Table 19. Reading: Selected Evaluation Items	59
Table 20. Writing: Selected Evaluation Items	60
Table 21. Listening: Selected Evaluation Items	61
Table 22. Speaking: Selected Evaluation Items.....	62
Table 23. ELPA21 Proficiency Determination Rules: Recommendations from In-Person Standard Setting Panelists.....	64
Table 24. Cross-Grade Proficiency Determination Evaluations and Results	65
Table 25. Adopted Cut Scores, Grade K, Grade 1, and Grade 2-3	69
Table 26. Adopted Cut Scores, Grade 4-5, Grade 6-8, and Grade 9-12	70
Table 27. Impact Data Associated with Adopted Cut Scores	71
Table 28. Impact Data Associated with Adopted Proficiency Determination Rules.....	76
Table 29. Comparison of CGS Ratings and ELPA21 Proficiency Level, Grades K-5.....	77
Table 30. Comparison of CGS Ratings and ELPA21 Proficiency Level, Grades 6-12....	78
Figure 1. Screen Shot of Contrasting Groups Study Data Entry	16
Figure 2. Screen Shot of Contrasting Groups Excel Worksheet Survey	17
Figure 3. Relationship between the Policy Descriptors, ALIs and Target Student Descriptors	25
Figure 4. Sample Item Map	34
Figure 5. Studying the Ordered Item Booklet.....	46
Figure 6. Bookmark Training Slide	48
Figure 7. Bookmarks.....	48
Figure 8. Round 1 Rating Form Data Entry Worksheet	50
Figure 9. Round 2 Display Worksheet.....	51

Figure 10. Round 3 Display Worksheet.....	52
Figure 11. Post-Round 3 Display Tab.....	54

1. Executive Summary

The English Language Proficiency Assessment for the 21st Century (ELPA21) consortium is a group of states that has developed an assessment system for English Learners (ELs) to measure performance in four domains, Reading, Writing, Speaking, and Listening, and to produce an aggregate Proficiency Determination classification of EL student proficiency. Performance in each domain is reported in terms of five achievement levels: Level 1 – Beginner, Level 2 – Early Intermediate, Level 3 – Intermediate, Level 4 – Early Advanced, and Level 5 – Advanced. The aggregate Proficiency Determination is based on profiles of performance across the four domains and is reported as *Proficient*, *Progressing*, or *Emerging*. Unique assessments for the four domains were developed for the following six grade levels or bands – K, 1, 2–3, 4–5, 6–8, and High School.

A systematic approach was taken to establish adopted cut scores for each domain, across all grades, K through high school, and in establishing the Proficiency Determination profile rules that included a Contrasting Groups Study (CGS), the formation of an English Language Expert Advisory Panel consisting of EL Experts from ELPA21 member states, and an in-person Bookmark Standard Setting workshop. The CGS study was conducted with input from over 200 teachers from ELPA21 member states who classified over 6,500 students on each domain, plus provided input, using the same 1-5 scale as the domain-rating, on a classification of the student’s overall English language skills. The results were used to both identify common profiles associated with determinations of Proficiency and, after matching the survey data with students’ operational test results, to provide cut score estimates for each domain.

The EL Expert Advisory Panel reviewed the CGS profiles, responded to a survey to refine the proficiency profile rules associated with the Proficiency Determination, wrote Target Student Descriptors to support the in-person standard setting workshop, and wrote Achievement Level Descriptors (ALDs) based on the adopted cut scores using the annotated item maps created at the in-person workshop.

An in-person standard setting workshop was conducted to recommend cut scores that define the five achievement levels using the Bookmark Standard Setting Procedure (Lewis, Mitzel, Mercado, & Schulz, 2012). The workshop was conducted July 19–22, 2016 and involved 51 panelists with eight to ten panelists per grade band. The workshop was designed to produce three outcomes: 1) recommended domain cut scores defining the thresholds of the Level 3 and Level 4 achievement levels; 2) recommended profiles and decision rules for combining individual student performance across the four domains to yield an aggregate Proficiency Determination classification; and 3) drafted ALDs for each achievement level.

The in-person standard setting workshop auditor’s report (Cizek, G. J., 2016) is provided in Appendix H. An excerpt from that report follows:

“It was observed that the standard setting activities for the ELPA21 assessments were conducted professionally, efficiently and, with few exceptions, as described in the standard setting design document/plan. There were many strengths noted; few issues arose during the standard setting workshop. Issues that did arise were

minor; they were handled deftly by the meeting facilitators and project staff; and they are unlikely to have a discernible negative effect on the results.

The procedures and processes used to derive recommended performance standards for the ELPA21 followed sound psychometric practices and no threats to the validity of the resulting performance standards were observed. One important source of validity information was not available at the time this report was written: the results of participants' evaluations.

With the assumption that the participants' evaluations do not suggest cause for concern, the available information and observations suggest that cut score recommendations produced at the workshop can be considered to be valid and reliable estimates of the cut scores for the ELPA21. Policy makers should have confidence that the recommendations from the standard setting activity were based on sound procedures, and produced trustworthy, valid, and defensible results.”

The panelist evaluations referenced in the excerpt are provided in Appendix G and support the validity of the in-person standard setting process and results.

Various approaches to establishing adopted cut scores were considered by ELPA21's governing body, the Consortium Council. The Council consists of representatives from all member states who consider evidence and recommendations presented by the Task Management Teams (TMTs) to make operational and policy decisions for the consortium. The decision-making associated with the adoption of Proficiency Determination rules by the Consortium Council are described in Section 6 and summarized below.

Following significant discussion among the stakeholders, ELPA21 adopted the recommendations of the in-person standard setting workshop panelists in each domain and the following Proficiency Determination Definitions:

Proficient: Students are Proficient when they attain a level of English language skill necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Level 4 or higher in all domains. Once Proficient on ELPA21, students can be considered for reclassification.

Progressing: Students are Progressing when, with support, they approach a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile with one or more domain scores above Level 2 that does not meet the requirements to be Proficient. Students scoring Progressing on ELPA21 are eligible for ongoing program support.

Emerging: Students are Emerging when they have not yet attained a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Levels 1 and 2 in all four domains. Students scoring Emerging on ELPA21 are eligible for ongoing program support.

Achievement Level Descriptors were developed to support score reporting following the adoption of cut scores. The process used to complete final ALDs is summarized in Section 6, and the resulting ALDs are provided in Appendix A2.

2. Introduction

2.1 Background

ELPA21 is a consortium of states that have designed and developed an assessment system for English Learners—the English Language Proficiency Assessment for the 21st Century (ELPA21). These assessments were administered operationally for the first time in spring 2016.

Unique assessments were developed for six grade levels or bands:

- K
- 1
- 2–3
- 4–5
- 6–8
- 9–12 (High School)

Assessments in each grade levels or band measure four domains:

- Reading
- Writing
- Listening
- Speaking

The ELPA21 consortium and the Council of Chief State School Officers (CCSSO) contracted with Pacific Metrics to design and implement a process to establish achievement levels for the ELPA21 assessments. The standard setting process provided the following outcomes:

1. Four recommended cut scores that define five achievement levels for each grade and domain
2. Recommendations for Proficiency Determination rules that associate profiles of performance across the four domains with three levels of proficiency—Emerging, Progressing, and Proficient
3. Achievement level descriptors (ALDs) that describe the knowledge, skills, and processes (KSPs) attributed to students at each level in each grade and domain

The five domain achievement levels are labeled:

- Level 1: Beginner
- Level 2: Early Intermediate
- Level 3: Intermediate
- Level 4: Early Advanced
- Level 5: Advanced

Table 1 provides the Policy Descriptors, adopted by consortium states prior to the standard setting process that describe students in each achievement level.

Table 1. Policy Descriptors for Domain Performance

LEVEL	A STUDENT AT THIS LEVEL...
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of developing grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits superior English language skills, as measured by ELPA21.

ELPA21 Policy Descriptors also guide the Proficiency Determination definition, as follows:

ELPA21 recommends that students be considered proficient when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once proficient on ELPA21, ELs can be considered for reclassification.

The adopted Proficiency Determination definitions are provided in Table 2. Note that the sentences beginning with “This is indicated on ELPA21 by attaining a profile of ...” were completed after the in-person workshop, upon adoption of definitions by the Consortium Council. All activities using draft Proficiency Determination definitions were as below, but included the incomplete sentence cited in this paragraph.

Table 2. Policy Descriptors for the Proficiency Determination

Policy Descriptors for the Proficiency Determination	
Proficient	Students are Proficient when they attain a level of English language skill necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Level 4 or higher in all domains. Once Proficient on ELPA21, students can be considered for reclassification.
Progressing	Students are Progressing when, with support, they approach a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile with one or more domain scores above Level 2 that does not meet the requirements to be Proficient. Students scoring Progressing on ELPA21 are eligible for ongoing program support.
Emerging	Students are Emerging when they have not yet attained a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Levels 1 and 2 in all four domains. Students scoring Emerging on ELPA21 are eligible for ongoing program support.

The following standards-referenced student level results are reported at each grade for the ELPA21 assessments:

Domain Scores. Domain scores in Reading, Writing, Listening, and Speaking are reported in terms of scale scores and achievement levels (Levels 1–5).

Proficiency Determination. The Proficiency Determination is a level of proficiency based on achievement in the four domains. The Proficiency Determination is reported in three levels—Emerging, Progressing, and Proficient.

2.2 Standard Setting Design Elements

A systematic standard setting process, including a Contrasting Groups Study, recommendations and work products from an English Learner Expert Advisory Panel, and an in-person Bookmark Standard Setting workshop, was conducted for ELPA21. Each of the three components is described briefly in this section and in detail in Sections 3, 4, and 5, respectively. The data and approaches considered by the ELPA21 Consortium Council and resulting policy decisions are documented in Section 6.

2.2.1 Contrasting Groups Study

The Contrasting Groups Study (CGS) was based on survey data from a broad sample of educators of English Learners who classified their students into achievement levels within each domain and overall using policy definitions and their knowledge of their students' performance on the four domains. The educators also provided input with

respect to the number of achievement levels for ELPA21 reporting. The survey was intended to inform and support the validity of ELPA21 decisions leading to cut scores and Proficiency Determination rules.

The results of the CGS provided ELPA21 with two useful pieces of data. First, it provided common profiles of proficiency, comprised of ratings in the four domains, that the participating teachers associated with overall English language proficiency. This information was used to support the establishment of the Proficiency Determination rules. Second, the CGS ratings provided an external measure of domain-specific and overall language proficiency to which we were able to compare the associated students' operational ELPA21 test results. These estimates were considered, in concert with cut-score recommendations from the in-person workshop, to support the adoption of cut scores for the four domains.

A detailed description of the CGS is provided in Section 3.

2.2.2 EL Expert Advisory Panel

An EL Expert Advisory Panel was formed to provide an additional opportunity for states and stakeholders to support and guide the establishment of cut scores. EL Expert Advisory Panel members supported the following activities, which are described in detail in Section 4:

1. The EL Experts reviewed CGS Proficiency Determination profiles prior to the in-person workshop and proposed recommendations for specific Proficiency Determination profile rules.
2. The EL Experts developed Target Student Descriptors—descriptors of specific knowledge, skills, and processes (KSPs) associated with students at the thresholds of Level 3 and Level 4 for each domain in grades K, 1, 3, 5, 7, and HS. The resulting Target Student Descriptors acted as “starting points” for the in-person standard setting panelists as they made their cut score recommendations.
3. A member of the EL Expert Advisory Panel participated in the in-person workshop as a resource for the panelists.
4. The EL Experts completed development of the Achievement Level Descriptors following the in-person workshop.

A detailed description of the EL Expert Advisory Panel activities and work products is provided in Section 4.

2.2.3 In-Person Bookmark Standard Setting Workshop

ELPA21 recruited 49 qualified EL educators and three alternates from the member-states to form six standard setting panels—one per grade band—to engage in the in-person standard setting workshop using the Bookmark Standard Setting Procedure. One alternate replaced a panelist who was unable to attend; the remaining alternates were assigned to appropriate grade bands to serve as panelists, resulting in a total of 51 panelists with eight to ten panelists per grade band. Each of the eight states had panelist representation in each grade band.

Panelists directly recommended Level 3 and Level 4 cut scores for the four domains in one grade of each grade level or band—K, 1, 3, 5, 7, and HS. Cut scores for grades 2, 4, 6, and 8 and for Levels 2 and 5 for all grades were established using the common within-grade band scale and linkages established across adjacent grade band scales, as described in Section 6. This approach—directly setting some cut scores and setting others through interpolation or estimation—was reviewed by the ELPA21 TAC and has been implemented in other English language development standard settings including CELDT (CTB/McGraw-Hill, 2006) and WELPA (CTB/McGraw-Hill, 2013).

Two additional activities were supported by the in-person standard setting panelists. First, panelists supported the writing of ALDs by selecting a scribe for each table to enter written descriptions of the KSPs associated with each item on the item map. These KSPs were based on table-level discussion of each item during study of ordered item booklets. Tables that finished scheduled activities early received training to refine these descriptions to support the development of ALDs.

Second, all panelists participated in a half-day systematic process to develop rule recommendations for Proficiency Determination profiles that would be used to classify students into one of three levels—Emerging, Progressing, or Proficient—based on the student’s achievement on the four domains.

A detailed description of the in-person Bookmark Standard Setting workshop is provided in Section 5.

2.2.4 Activities Subsequent to the In-Person Standard Setting Workshop

After the in-person workshop, the following information was provided by Pacific Metrics to ELPA21 to support the adoption of cut scores and Proficiency Determination rules:

1. Panelists’ recommended bookmarks and associated cut scores and impact data
2. Estimates of the standard error associated with panelist sampling for each grade band and domain
3. Panelists’ recommendations resulting from the Proficiency Determination discussions and activities

Pacific Metrics developed initial drafts of the ELPA21 ALDs, eliciting feedback from the EL Expert panel. These drafts reflected the knowledge gained from the workshop. Final ALDs were produced by ELPA21 under the guidance of a committee of state representatives. Details are provided in Section 6.

The in-person standard setting panelists recommended raw Level 3 and Level 4 cut scores (i.e., Bookmark placements) for grades K, 1, 3, 5, 7, and HS. Next, the National Center for Research, Evaluation, Standards, and Student Testing (CRESST) translated these bookmark placements to a scale score metric using parameters obtained from the item response theory (IRT) calibration of the item pool. IRT-based linking procedures were used to ensure proper ordering of the Level 3 and Level 4 cut scores across the grade bands and to interpolate these cut scores in the grades between those considered by the panelists (2, 4, 6, and 8). Finally, impact data were used to obtain recommendations for the Level 2 and Level 5 cuts, as described in Section 6.

The standard setting study plan, in-person workshop results (including auditor's evaluation), CGS survey results, and eventual recommendations concerning the rules for determining overall proficiency and the domain-specific cut scores were reviewed and discussed over multiple meetings with the ELPA21 Consortium Council and Technical Advisory Committee. In September 2016, the ELPA21 Governing Board adopted the cut scores presented in this report, defining the five achievement levels in each domain and domain, as well as the rules specifying overall Proficiency Determination from domain profile. These operational cut scores and rule are described in Section 6.

3. Contrasting Groups Study

The Contrasting Groups Study (CGS) is an examinee-centered standard setting method often conducted by survey. In this case, EL teachers recruited by ELPA21 states completed a survey by classifying their students into achievement levels on each domain and determining an aggregate classification of English language proficiency using policy definitions and knowledge of their students' overall English skills on the domains. The teachers also answered questions about the appropriateness of the proposed number of achievement levels. The survey provided information that helped inform and validate ELPA21 decisions leading to the adopted cut scores, Proficiency Determination rules, and ALDs.

The CGS yielded (a) a set of profiles on the four domains associated with various levels of aggregate proficiency that were used to support the adoption of the Proficiency Determination rules, (b) a set of estimated cut scores on each of the four domains that were considered by ELPA21 in addition to those from the in-person standard setting workshop prior to adopting cut scores, and (c) correlations between teachers' ratings and ELPA21 scores that lend validity evidence to the assessments. This section describes the CGS implementation including the materials used, training conducted, data, and results.

3.1 Implementation

Pacific Metrics designed the CGS, developed the surveys, conducted training for teacher-participants, and collected the resulting data. ELPA21 staff drafted a recruitment survey which was delivered via Formsite.com—a website designed to develop and administer online surveys. The recruitment survey collected demographic and background information that supported the selection of qualified teachers for the CGS (as well as the in-person standard setting workshop). ELPA21 and its member-states recruited teacher-participants to provide the demographic and background information and to participate in the CGS.

3.2 Materials and Survey Methods

Online Survey. The CGS survey was accessed by participating teachers through a secure website. Teachers completed the survey using an online form or a downloaded Excel workbook. The online survey required respondents to enter each student's name, state ID, and grade level and then, after considering the Policy Descriptors and associated contextual elements, to classify each student into one of five levels of proficiency for reading, writing, speaking, listening, and overall. Figure 1 provides a screenshot of the survey. Note that the assessment target appears at the top of the survey. This screenshot shows the data entry form for the overall category. The student names were entered in the first column. Teachers rated each student in one of five levels using radio buttons. In addition to overall achievement, this type of form was completed for each domain. Screenshots of the complete online survey, including the Policy Descriptors that define Levels one through five, are provided in Appendix B1.

Overall Proficiency: ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level, content-related academic tasks without on-going EL program support. Once proficient on ELPA21, ELs can be considered for reclassification.

Overall					
	Level 1	Level 2	Level 3	Level 4	Level 5
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 1. Screen Shot of Contrasting Groups Study Data Entry

Excel Workbook. As an alternative to hand-entering each student's name and ID in the online survey, teachers were provided the option to download an Excel workbook. The Excel workbook had the same requirements as the online survey but it allowed respondents to copy and paste their roster into the worksheet. Excel workbooks were submitted by the respondents to Pacific Metrics via Formsite. The Excel file contained five tabs:

- Instructions
- Policy Descriptors
- Assessment Targets
- Teacher Information
- Contrasting Groups Survey

Figure 2 shows a screenshot of the Excel worksheet version of the CGS survey. The first four columns collected student first name, last name, student ID, and grade. The remaining columns were associated with the four domains and overall proficiency. The users assigned students to one of five achievement levels for each domain and overall. Screen shots of each tab in the Excel CGS survey are provided in Appendix B2.

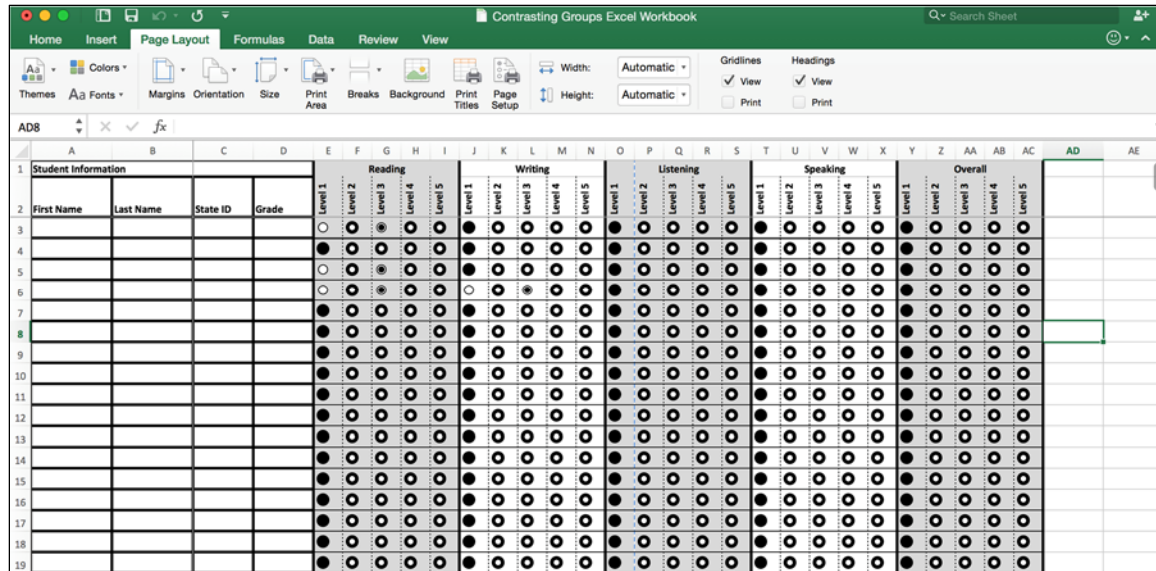


Figure 2. Screen Shot of Contrasting Groups Excel Worksheet Survey

3.3 Training and Data Collection

Pacific Metrics sent a link to the survey on February 28, 2016, and the collection of data began on February 29, 2016. The respondents were instructed to complete the survey within two weeks of their students taking the test. Additionally, the respondents were instructed to make their classifications based on their overall knowledge of student skills and performance, not based on how they think students performed on the ELPA21 assessment. Data collection ended in late April, 2016.

Pacific Metrics provided three online training webinars that were accessed by survey respondents prior to completing the online survey. The three webinars provided general information about standard setting and details about the CGS process, trained respondents to complete the survey using the online tool and provided tips for entering data, and demonstrated how to complete the survey using the Excel workbook. Screenshots of the webinar training slides are provided in Appendix B3.

3.4 Data Analyses and Results

3.4.1 Data

Over two hundred teachers from Arkansas, Iowa, Kansas, Nebraska, Ohio, Oregon, Washington, and West Virginia participated in the CGS, providing data for over 6,500 students. Of these, 4,524 records from the 2015-2016 ELPA21 summative results were subsequently matched with the CGS ratings.

Table 3 provides the number of teachers, by grade, who responded to the CGS survey and the total number of students, by grade, that were classified by teachers.

Table 3. Number of Teachers Responding to CGS and Total Number of Students Classified

Grade	Number of Teachers*	Number of Students
K	97	818
1	94	797
2	101	770
3	100	742
4	100	703
5	77	406
6	71	421
7	65	519
8	59	493
9	47	315
10	55	318
11	46	187
12	39	147

* The teachers are not unique by grade. The total number of unique teachers was 242.

Comparison of Current State Data and CGS Data. Data were requested from ELPA21 states to support the comparison of states' current percentage of students in each EL achievement level based on the 2015 ELD assessments. This comparison was conducted in order to evaluate the "reasonableness" of the CGS impact data.

Data associated with their current English Language Proficiency assessments were provided by four ELPA21 states (Ohio, Oregon, Arkansas, and Washington). Most states provided Overall Proficiency, an achievement level that is conceptually similar comparable in nature to ELPA21's overall Proficiency Determination. Most Overall Proficiency levels were reported in terms of 5 levels. The median percent of students in each level for the four states was used for comparison purposes. We note that the comparisons are useful, but are limited, due to differences between the policy definitions, assessment targets, sampling of students, and ELD standards used in these states and ELPA21.

The differences indicate that most of the CGS impact data was within 5% of current state data, as reflected in the rows labeled "CGS – State" in Table 4. Only one individual achievement level had a difference between the CGS and state median impact data of more than 10% (grade 8 Level 1). Some differences between the previous state instrument and the new ELPA21 assessments would be expected and the results appear reasonable and appropriate to support the standard setting.

Table 4. Comparison of State and CGS Impact Data

Grade	Data Source	N	Level				
			1	2	3	4	5
K	CGS	808	15%	36%	32%	15%	2%
	State Med.		16%	38%	30%	12%	3%
	CGS - State		-1%	-2%	2%	4%	-1%
1	CGS	792	6%	27%	37%	23%	7%
	State Med.		9%	32%	30%	20%	6%
	CGS - State		-3%	-5%	7%	3%	1%
2	CGS	763	5%	20%	39%	27%	8%
	State Med.		7%	28%	33%	20%	9%
	CGS - State		-2%	-8%	6%	7%	-1%
3	CGS	738	6%	19%	40%	27%	7%
	State Med.		11%	25%	34%	18%	5%
	CGS - State		-5%	-6%	6%	9%	2%
4	CGS	691	5%	15%	40%	28%	12%
	State Med.		7%	17%	32%	31%	6%
	CGS - State		-2%	-2%	8%	-3%	6%
5	CGS	402	5%	17%	35%	31%	11%
	State Med.		6%	12%	29%	31%	8%
	CGS - State		-1%	5%	6%	0%	3%
6	CGS	426	15%	19%	31%	26%	8%
	State Med.		7%	13%	29%	28%	6%
	CGS - State		8%	6%	2%	-2%	2%
7	CGS	510	12%	18%	31%	27%	12%
	State Med.		7%	11%	29%	28%	8%
	CGS - State		5%	7%	2%	-1%	4%
8	CGS	481	18%	18%	28%	23%	14%
	State Med.		7%	11%	29%	31%	10%
	CGS - State		11%	7%	-1%	-8%	4%
HS	CGS	931	15%	20%	30%	24%	11%
	State Med.		11%	16%	26%	33%	13%
	CGS - State		5%	4%	4%	-9%	-2%

3.4.2 Recommended Number of Achievement Levels

Teachers were asked to provide their opinion with respect to the appropriate number of achievement levels to support ELPA21 reporting. The results indicated a strong recommendation for five achievement levels, as indicated in

Table 5. This supports ELPA21's decision to report assessment results based on five levels.

Table 5. Teacher Endorsements for the Number of Achievement Levels

Number of Levels	Number of Endorsements	Percent of Endorsements
3	0	0.0%
4	8	3.7%
5	190	88.0%
6	18	8.3%

3.4.3 Results Supporting Proficiency Determination Profile Rules

The CGS results provided common profiles used to classify students. These classifications were used to identify common profiles that provided valuable information with respect to the adoption of Proficiency Determination rules.

The 75 most common patterns are provided in Table 6, based on all grades. While 75 profiles are provided in Table 6, note that nearly half of the students described by the CGS associate with one of the first 13 profiles; more than 85% of rated students associated within the 75 listed profiles. Appendix B4 provides frequencies for all Level 4 profiles by grade and the 25 most common profiles by grade.

Table 6. Seventy-Five Most Common Profiles for All Grades (Profile Order: Overall, Reading, Writing, Listening, Speaking)

Profile	N	%	Cum %	Profile	N	%	Cum %	Profile	N	%	Cum %
33333	460	7.0%	7.0%	11122	60	0.9%	65.8%	44354	24	0.4%	78.6%
11111	405	6.2%	13.2%	32244	52	0.8%	66.6%	32343	23	0.4%	79.0%
22222	368	5.6%	18.8%	43444	50	0.8%	67.4%	43343	23	0.4%	79.3%
44444	352	5.4%	24.2%	44355	50	0.8%	68.1%	43454	23	0.4%	79.7%
33344	275	4.2%	28.4%	32243	46	0.7%	68.8%	21133	21	0.3%	80.0%
55555	249	3.8%	32.2%	43354	42	0.6%	69.5%	23222	20	0.3%	80.3%
44455	185	2.8%	35.1%	33354	41	0.6%	70.1%	34334	20	0.3%	80.6%
22233	167	2.6%	37.6%	34344	39	0.6%	70.7%	45444	20	0.3%	80.9%
44344	165	2.5%	40.1%	43355	37	0.6%	71.3%	22133	19	0.3%	81.2%
33233	153	2.3%	42.5%	54555	34	0.5%	71.8%	23223	19	0.3%	81.5%
32233	149	2.3%	44.8%	34333	33	0.5%	72.3%	34343	19	0.3%	81.8%
43344	145	2.2%	47.0%	45455	33	0.5%	72.8%	22243	18	0.3%	82.0%
22223	131	2.0%	49.0%	21222	32	0.5%	73.3%	32234	18	0.3%	82.3%
33343	123	1.9%	50.9%	33243	30	0.5%	73.7%	44345	18	0.3%	82.6%
33334	115	1.8%	52.6%	43455	30	0.5%	74.2%	33444	17	0.3%	82.8%
55455	113	1.7%	54.3%	21132	29	0.4%	74.6%	21112	16	0.2%	83.1%
22232	92	1.4%	55.7%	32344	28	0.4%	75.1%	22112	16	0.2%	83.3%
54455	83	1.3%	57.0%	11121	27	0.4%	75.5%	22244	16	0.2%	83.6%
11112	81	1.2%	58.3%	44434	27	0.4%	75.9%	33223	16	0.2%	83.8%
21122	81	1.2%	59.5%	21123	26	0.4%	76.3%	22123	15	0.2%	84.1%
44454	75	1.1%	60.6%	32232	26	0.4%	76.7%	22212	15	0.2%	84.3%
33244	74	1.1%	61.8%	33234	26	0.4%	77.1%	22234	15	0.2%	84.5%
22122	71	1.1%	62.9%	33323	26	0.4%	77.5%	23233	14	0.2%	84.7%
32333	71	1.1%	63.9%	32223	25	0.4%	77.9%	33332	14	0.2%	84.9%
44445	62	0.9%	64.9%	33355	24	0.4%	78.2%	33345	14	0.2%	85.2%

An analysis of the CGS profiles can be summarized as follows:

- The common CGS profiles associated with an overall Proficiency Determination of Proficient (defined as Overall Level 4 or Level 5 in the CGS results) are defined by the following 3 rules.
 - (Writing is greater than or equal to three) and (Reading, Listening, and Speaking are greater than or equal to four)
 - (Reading and Writing are greater than or equal to three) and (Listening and Speaking are greater than or equal to four)
 - Reading, Writing, Listening and Speaking are all greater than or equal to four

- If students were assigned a 1 or a 2 in any domain then they are not Proficient
- Two common profiles conflicted at most grades.
 - Some teachers thought the profile 3344 (Reading, Writing, Listening, Speaking) should be associated with an overall Proficiency Determination of three while others thought it should be associated with four (using the five category Proficiency Determination scheme that was in place for the CGS).
 - Some teachers thought the profile 2233 (Reading, Writing, Listening, Speaking) should be associated with an overall Proficiency Determination of two while others thought it should be associated with an overall three.

3.5 Use of Results

The CGS results were used for multiple purposes. First, the survey responses provided a basis for judging the appropriateness of the proposed number of levels. The vast majority of teachers recommending the use of five achievement levels, confirming the validity of the planned, and subsequently adopted, number of levels of achievement within each language domain.

Second, the observed patterns of domain profiles and their association with the ratings of overall student proficiency were used to inform the specification and adoption of Proficiency Determination profile rules used to classify students as Emerging, Progressing, and Proficient. The common profiles associated with an overall proficiency rating were presented to the EL Expert Advisory Panel and to the in-person standard setting panelists for consideration and discussion as part of the process used to make their respective recommendations for Proficiency Determination profile rules.

Third, the positive correlations and generally high levels of agreement between the teachers' ratings and the matched students' ELPA21 Proficiency determinations provides support for the validity of the ELPA21 assessments (see Tables 29 and 30). That is, the teachers' conceptualization of the construct associated with English Language Proficiency is positively associated with the construct measured by the ELPA21 assessments.

4. English Learner Expert Advisory Panel

ELPA21 members were asked to nominate experts who could participate throughout the standard setting process to represent their state's stakeholders and interests. Eight states nominated eleven experts who were engaged in EL education at various levels. These experts included contributors from the State Departments of Education, district, and school staff who provided expertise in education equity, language instruction for limited English proficient and immigrant students, and dual language immersion.

The following activities were conducted by various groupings of the EL experts in support of the standard setting:

1. A review of the CGS profiles and the subsequent recommendation of rules associating domain profiles with Proficiency Determination achievement levels (May and June 2016).
2. Development of Level 3 and Level 4 Target Student Descriptors to support the in-person standard setting panelists. These Target Student Descriptors are descriptions of the knowledge, skills, and processes expected of students who are just at the entry of the associated achievement level. (June and July 2016).
3. Provision of subject matter expertise to support panelists at the in-person workshop. One member of the EL Advisory Panel attended the in-person workshop to support panelists across all grade-groups. (July 2016).
4. Development of achievement level descriptors. The resources to develop final ALDs were created at the in-person workshop. That is, a scribe at each table took notes summarizing panelists' review and discussion of the KSPs measured by items in the OIBs. Summarizing and refining these notes resulted in grade-level ALDs that communicate the English language skills typically held by students in each achievement level to parents, teachers, and other stakeholders (July-September 2016).

4.1 Proficiency Determination Profile Rules

The EL Experts were convened in an online meeting on May 11, 2016. Dr. Daniel Lewis from Pacific Metrics provided an overview of ELPA21 and the CGS results, and presented and led a discussion of issues related to the establishment of profile rules for the Proficiency Determination. The slides presented at this meeting are provided in Appendix C1.

A survey was distributed to the EL Experts via email on May 18, 2016, that supported the EL Expert's recommendations for the following:

- Rules to crosswalk the five-level Overall Proficiency classification used by the CGS teachers to the three-level Proficiency Determination applied during the in-person meeting and used for reporting
- Profiles associated with an overall classification to Level 4 (using the five-level Overall Proficiency classification scheme used by the CGS teachers)

- Profiles associated with an overall classification to Level 3 (using the five-level Overall Proficiency classification scheme used by the CGS teachers)

A copy of the survey is provided in Appendix C2.

Survey Results. Six of the eleven panelists responded in the one-week survey window. An additional week was provided, however no additional responses were received. The following results were endorsed by the majority of responders:

1. Rules to crosswalk the five-level Overall Proficiency classification used by the CGS teachers to the three-level Proficiency Determination used for reporting
 - a. Overall Level 1 and Level 2 from the CGS are associated with a Proficiency Determination of Not Proficient (reabeled Emerging).
 - b. Overall Level 3 from the CGS is associated with a Proficiency Determination of Nearly Proficient (reabeled Progressing).
 - c. Overall Level 4 and Level 5 from the CGS are associated with a Proficiency Determination of Proficient.
2. Profiles associated with an overall classification to Level 4 (using the five-level Overall Proficiency classification scheme used by the CGS teachers)
 - a. All domains are Level 4 or Level 5.
 - b. Reading and Writing are Level 4 and Listening and Speaking are at least Level 3.
3. Profiles associated with an overall classification to Level 3 (using the five-level Overall Proficiency classification scheme used by the CGS teachers)
 - a. All domains are Level 3, Level 4, or Level 5, but not enough to be classified as Level 4 or Level 5.
 - b. One or more Level 2 domains can be compensated by an equal number of Level 4 or Level 5 domains (with no Level 1 domains).

4.2 Target Student Descriptor Writing

Target Student Descriptors define the knowledge, skills, and processes (KSPs) of the student just entering the achievement level. So the Target Student Descriptor for Level 4 represents a student entering (crossing the threshold into) Level 4. Target Student Descriptors were created to support the in-person standard setting workshop, and they provided the workshop panelists with a common framework for the KSPs expected of the student entering the achievement level. The Target Student Descriptors were not meant to be prescriptive—the in-person standard setting panelists were free to make cut score recommendations that diverged from the Target Student Descriptors based on their expert judgment—however they were intended to provide a common starting point.

The EL experts created Target Student Descriptors for Level 3 and Level 4 by modifying the ELPA21 achievement level indicators (ALIs). The ALIs describe the range of performance within each achievement level. Students who demonstrate all of the KSPs described in an ALI for a given achievement level are expected to be at the top of the achievement level, not just entering the level. So the ALIs describing Level 3 were used

to develop the Target Student Descriptors for Level 4. The ALIs were written for grade bands, while the cut scores established at the in-person standard setting were set for specific grade levels. Thus, the ALIs needed to be modified by the EL Experts to support the grade-specific work at the in-person workshop.

Figure 3 depicts the relationship between the ALIs, policy descriptors, and Target Student Descriptors for Level 4 (an analogous relationship was used to develop the Level 3 Target Student Descriptors). The figure shows that the policy descriptors may not align perfectly with the expectations articulated in the ALIs. In the figure, the Level 4 target student is described by most of the skills found in the Level 3 ALIs (which cover the full range of ability, from low to high, for Level 3 students). These are the KSPs that the target student must master to just make it into Level 4.

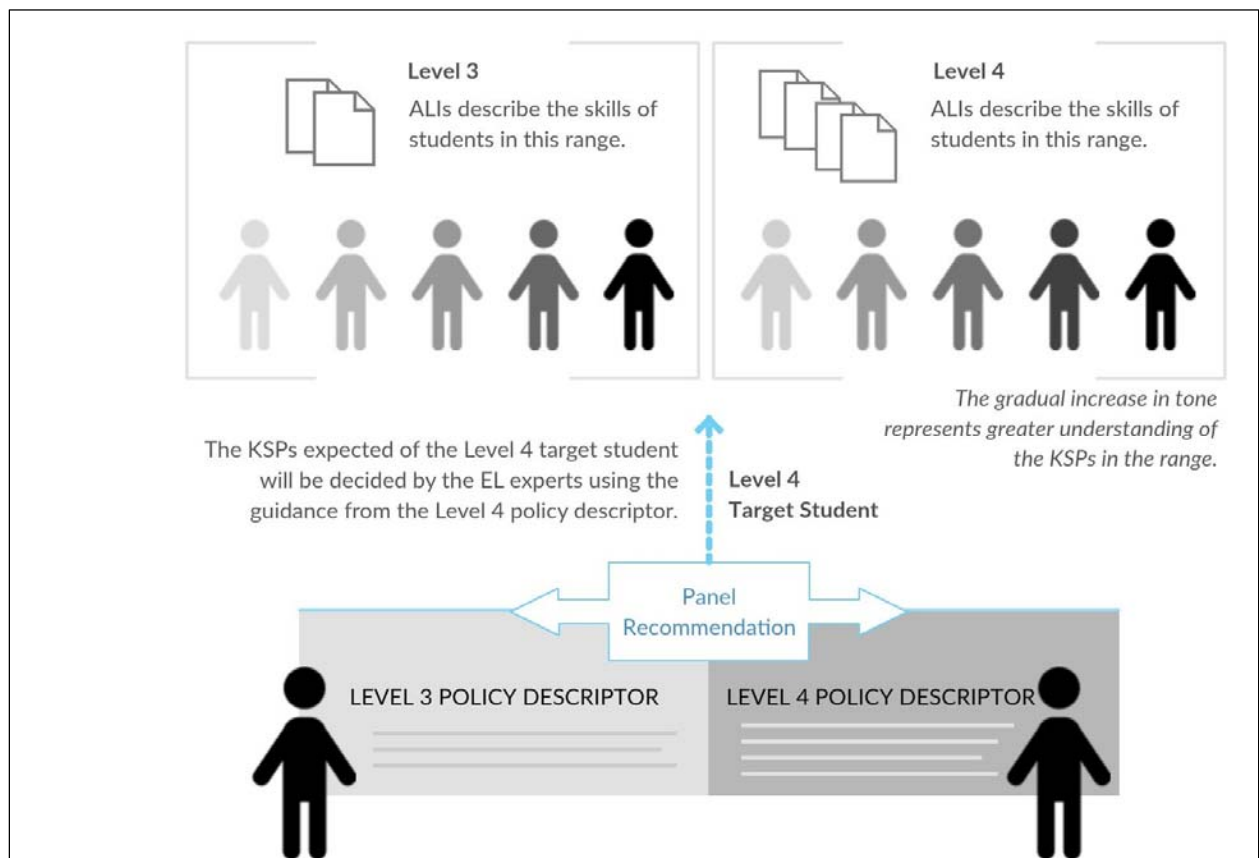


Figure 3. Relationship between the Policy Descriptors, ALIs and Target Student Descriptors

Process for Creating Target Student Descriptors. EL Experts were trained on June 21, 2016, in an online meeting. Dr. Karla Egan, a Pacific Metrics consultant, led the training session, which included an overview of the standard setting process, guidance on how the Target Student Descriptors support the in-person standard setting, and instructions on how to create the Target Student Descriptors.

The EL Experts developed Level 3 and Level 4 Target Student Descriptors for grades K, 1, 3, 5, 7, and High School for each domain. In this process, the EL experts reviewed the

Level 3 and Level 4 ALIs, with the understanding that the ALIs were written for grade bands and described the full range of ability for each level, while the Target Student Descriptors described students at the threshold of the level at specific grades. The EL Experts were guided by the Level 3 and Level 4 policy descriptors, provided in Table 7, the ALIs, and their expert judgment.

The resulting Target Student Descriptors are provided in Appendix A1.

Table 7. Policy Descriptors for Level 3 and Level 4*

Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.

* Subsequently revised

4.3 Achievement Level Descriptor Writing

After cut scores were adopted by ELPA21, EL Experts wrote achievement level descriptors (ALDs) that, like the Target Student Descriptors, reflect the KSPs of the student just entering the achievement level. The difference between ALDs and Target Student Descriptors is that ALDs are based on student performance on the test items they are expected to master to reach the cut score while target student descriptors are expectations of how students should perform. The ALDs reflect the KSPs that students who have just met the cut score *can* do while Target Student Descriptors reflect the KSPs students at the cut score *should* be able to do. If the EL Experts were able to perfectly predict what target students could do on the test, then there would be a one-to-one correspondence between target student descriptors and ALDs.

To begin developing the ALDs, Dr. Egan trained the in-person standard setting panelists to synthesize useful information derived during the workshop. That is, as part of the in-person Bookmark process, panelists reviewed ELPA21 test items in order of difficulty and answered the following two questions:

What does this item measure?

Why is it more difficult than the preceding items?

On August 4, 2016, Dr. Egan led the EL Experts through training on how to write ALDs. Because the final cut scores had not yet been determined, the EL Experts first task was to finish the task of synthesizing information from the two questions answered by panelists during the in-person workshop.

After ELPA21 adopted final cut scores, the EL Experts completed a draft of the ALDs. This task was straightforward. Dr. Egan associated each item on the annotated item maps that resulted from the in-person standard setting workshop with an achievement level based on the ELPA21-adopted cut scores. The EL Experts compiled the synthesized information into the draft ALDs. The EL Experts compiled the synthesized information

into the draft ALDs. ELPA21's Item Development TMT used these drafts to develop final ALDs. This process is described in more detail in Section 6.3 Finalizing Achievement Level Descriptors.

5. In-Person Bookmark Standard Setting Workshop

The Bookmark Procedure is well suited to support the establishment of cut scores for the ELPA21 assessments and, properly implemented, adheres to best practices (see Section 7, Validity). The Bookmark Standard Setting Procedure:

- is an item-mapping procedure, which is appropriate for tests with multiple item formats that are scaled using item response theory,
- allows panelists to engage in the same judgment task whether considering selected-response items, constructed-response, or other partial credit items, and
- is well documented and has been successfully implemented in most state testing programs and for other English language development tests.

The Bookmark Procedure is also appropriate for assessment programs in which detailed grade-specific ALDs have not been developed prior to the in-person standard setting workshop but are desired. Bookmark panelists use their expertise to specify the skills associated with each achievement level as part of their cut score recommendation (placing of bookmarks). ALDs can then be drafted after the final round of Bookmark ratings using the materials and expertise developed by panelists in their review of the ordered item booklets.

5.1 Overview of the In-Person Bookmark Standard Setting

The in-person standard setting workshop was conducted July 19–22, 2016, at the St. Louis Westin Hotel. The Bookmark Procedure is an item-mapping standard setting method; the fundamental tools used for Bookmark Procedure are the ordered item booklet (OIB) and item map, which are described in detail later in this section.

The Bookmark Standard Setting Procedure derives its name from the primary task required of panelists—the placement of a bookmark in the OIB to represent a cut score recommendation. For example, the bookmark for the Level 4 cut score is placed at the first point in the OIB at which panelists determine that a student who demonstrates mastery of the skills reflected by the items before the bookmark has demonstrated English language proficiency sufficient to be classified in Level 4.

Over the course of the three rounds of judgments, panelists recommended content-based cut scores using information from the Policy Descriptors, Target Student Descriptors, test content viewed in the OIBS, panelist discussions, and impact data.

Panelists recommended the Level 3 and Level 4 cut scores—which separate Level 2 from Level 3, and Level 3 from Level 4, respectively—for one grade in each grade band (K, 1, 3, 5, 7, and HS; see Table 8). IRT-based linkages between the scales of adjacent grades were used to ensure that the cuts increased in rigor with grade level (e.g., the Level 4 cut in grade 5 corresponds to a higher level of ability in the domain than the Level 4 cut in grade 3) and to interpolate Level 3 and Level 4 cut scores for the remaining grades (2, 4, 6, and 8). This linking was necessary due to the fact that ELPA21 grade band tests are not on a common vertical scale. Finally, cut scores for Level 2 and Level 5 were established using impact data. Specifically, these cuts were placed at locations that would produce equal proportions of students in Level 1 and 2, and in Levels 4 and 5.

Table 8. Grades in Which the Level 3 and Level 4 Cut Scores were Set Directly by Panelists

Grade Band	Grades in Which Level 3 and Level 4 Cut Scores were Set Directly by Panelists
K	K
1	1
2–3	3
4–5	5
6–8	7
HS	HS

5.2 Item Mapping: Locating Items on the IRT Scale

The Bookmark Procedure is an item response theory (IRT)-based standard setting procedure, which requires the location of items on the IRT scale using a specified response probability. A detailed discussion of issues associated with the selection of a response probability is available in Lewis, Mitzel, Mercado, & Schulz (2012). RP67 is the most commonly used response probability for standard setting purposes and was selected for the ELPA21 standard setting upon review by ELPA21 and their TAC.

Selected response and other dichotomous items are located on the test scale where an examinee would have a .67 likelihood of knowing the correct response. Polytomous items, such as multi-point constructed response items and partial credit items, are located at multiple points on the scale—once for each positive score point. Thus, a constructed-response item scored 0–2 would have two RP67 locations. The first location is associated with a score of 1 at the scale score where an examinee would have a .67 likelihood of scoring at least 1 on the item. The second location is associated with the scale score at which an examinee would have a .67 likelihood of achieving a score of 2 on the item.

The use of RP67 is an implicit operationalization of the degree to which examinees in an achievement level hold the skills attributed to them by the descriptors. Thus, when the Target Student Descriptor indicates that a first grade Level 4 EL “can identify reasons an author gives to support the main point,” the use of RP67 indicates that we expect that such a student, when presented with an item that measures this skill, will achieve a certain score point (or greater) with at least .67 probability.

The RP value provides a means of displaying the distribution of items in the OIB across the range of ability. In an item-mapping procedure, it is important that items are well distributed. Pacific Metrics examined the distribution of items based on RP67 using the field test data that were available prior to the standard setting to support the development of ordered item booklets that had relatively few scale gaps and met content requirements.

The RP67 values were calculated by CRESST and confirmed by Pacific Metrics. All test items were calibrated using a logistic graded response model, as described in the ELPA21 Scoring Specification. The RP67 location, $\theta_{j,k}$, of a positive score point k for an item j is provided by the formula

$$\theta_{j,k} = b_{j,k} + (1/a_j)\log(0.67/0.33),$$

where $b_{j,k}$ is the difficulty (or location) parameter of score point k for partial credit items and a_j is the slope (or discrimination) parameter of item j . Note that for dichotomous items, $k = 1$. Preliminary item calibrations using data from the 2015-2016 operational summative test administration provided the initial RP67 values that were used in the standard setting meeting. Upon completion of final item calibration, RP67 values were updated and used in the translation of raw scores to the ELPA21 scale within each grade band.

The values $\theta_{j,k}$ were linearly transformed from their native logit metric (established in the item calibration to have a mean of 0 and standard deviation of 1) to an interim scale for standard setting purposes. Transformation constants were used to establish an interim scale with a mean of 500 and standard deviation of 50 within the grade band. The purpose of establishing the interim scale was to put the item locations on a scale that would more closely resemble the final reporting scale, than the untransformed logit scale (with mean of zero and standard deviation of one).

Note that transformation constants used in the final reporting scale (resulting in a scale mean of 550 and standard deviation of 80) were selected after standard setting.

The use of an interim scale allowed item locations to be placed on the item map in a scale score metric that was more appropriate for use by panelists (i.e., a scale that had no negative numbers and with scale scores and RP67 values rounded to whole numbers).

5.3 Preparation of Materials

Numerous sets of materials were developed to support the in-person standard setting. A description of the various materials and the methods used to develop those follows.

5.3.1 Ordered Item Booklets

The ordered item booklet (OIB) is the central tool in the Bookmark Procedure. It allows panelists to study each test item or score point in order of difficulty. Note that the term “Ordered Item Booklet” is commonly associated with the Bookmark Standard Setting Procedure and is used here. ELPA21 uses the term “task” to refer to the stimulus prompting students’ responses. Some tasks include multiple components with multiple scores assigned, other tasks have multiple components with a single score assigned, and still other tasks have only a single component (with a single score assigned). It is to the scored units to which we refer when using the term “item” in this report.

The information in the OIB served to prompt discussion among the panelists, in order to gain an understanding of the items’ response demands—the skills that the items measure and what students who respond successfully to the items or score points must know and be able to do. Following discussion, panelists use the OIB to guide their selection of cut scores.

Because the information in the OIB provides important validity evidence, it is important that the OIBs consist of a well-distributed set of items. In other words, the difficulty of the items should increase gradually and there should not be large gaps between the difficulties of adjacent items.

The OIB is domain-specific and is composed of a set of items selected using field test data such that the content coverage is representative of the domain and covers the scale appropriately. This was accomplished by selecting one complete operational test form for each OIB and supplementing the OIB with additional items that (a) met the overall test form blueprint with respect to Task Type and Task Subtype as closely as possible and (b) had item locations that supported minimal gaps in the test scale. Maintaining the test blueprint supported the maintenance of the test construct in the OIBs. The complete form was selected by reviewing all operational test forms and identifying the one with items that provided the best scale coverage.

OIBs were constructed with a target length of 45 to 60 score points. The lower end of this target range was deemed sufficient if two or more intact test forms were included in the OIB. The OIB length targets are based on Pacific Metrics' extensive experience conducting Bookmark Standard Settings with respect to the amount of time panelists take to review items, the cognitive load that the review of the OIB places on panelists, and the number of items necessary to represent the construct. Specifically, fewer than 45 items would likely not provide panelists with sufficient evidence to support their decision making; that is, panelists need a sufficient body of evidence to support their judgments. Also, more than 60 items, especially in the higher grades, tends to be cognitively draining on panelists and takes too much time to review. In this case, the timing was very important because each panelist had to study OIBs for four domains. Table 9 provides the number of pages (items and score points) in the OIB for each grade and domain.

Table 9. Number of Items and Score Points in the Ordered Item Booklets

Grade	Number of Score Points			
	Listening	Reading	Speaking	Writing
Kindergarten	45	45	54	51
Grade 1	50	48	60	57
Grade 3	51	48	57	48
Grade 5	51	55	60	51
Grade 7	50	48	57	56
High School	55	45	54	51

Each page of the OIB represented a single dichotomous item or polytomous item score point. Each dichotomous item was represented once in the OIB—at the item's location. Each polytomous item was located multiple times in the OIB—once for each positive score point. For example, a constructed-response item scored from 0–2 was located twice in the OIB. The first time this constructed-response item appeared in the OIB, it was associated with the skills necessary to achieve a partial credit score of 1. The second time it appeared, later in the OIB, it was associated with the skills necessary to achieve the full score of 2.

The first page/item in the OIB represented the item with the lowest IRT location, or the “easiest” item, and the items were ordered by difficulty such that the last item or score point was the “hardest” item and thus, had the highest IRT RP67 location.

Creating OIBs. PDFs of the items and stimuli in a domain and grade band were split into individual item PDFs (one or multiple pages) and stimulus pages. PDFs of items selected to be in the OIB were ordered and numbered with the order of difficulty in the upper right-hand corner of the pages. In addition, if an item had score points, that was indicated at the top of the page (e.g., Score Point 1 of 2, Score Point 2 of 2).

Each item in the OIB began with item information at the top of the page, including Grade level/band, Domain, Item ID, Stimulus ID, ELP Standard(s), Sub-Claim(s), Common Core State Standard (CCSS)/Next Generation Science Standards (NGSS) Language Practice(s), Item Type, and Correct Answer.

All OIBs underwent quality assurance reviews to make sure that the Item and Stimulus IDs on each page matched those on the item map and each PDF page was examined to confirm the quality of the item rendering.

5.3.2 Item Maps

Item maps accompany and provide information about the items in the OIBs. The columns of the item maps indicated the following:

- (a) Order of Difficulty (OIB Page Number)
- (b) Location (IRT RP67 scale location)
- (c) Score Point (1 for selected-response and other dichotomous items and the associated score point for constructed-response and other polytomous items, e.g., 1 of 2)
- (d) Item ID__AIR ITS Item ID (the item ID used by ELPA21 with the AIR item ID following the underscore)
- (e) Set Leader__AIR Set Leader ID (the ELPA21 set leader ID with the AIR set leader ID following the underscore), which indexes a common stimulus associated with multiple items
- (f) Task Type
- (g) Task Sub Type

The final two columns in the item maps posed questions for panelists to complete as they studied each item in the OIB as part of the Bookmark Procedure:

1. What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?
2. Why is this item or score point more difficult than the items that precede it?

The panelists discussed the answers to these questions in small groups in order to develop a solid understanding of the construct measured by the test. This is considered skill development that is a prerequisite to their making cut score recommendations. Figure 4 shows the sample item map that was used to train participants. The item IDs have been edited for security purposes.

5.3.3 Stimulus Booklets

Stimulus booklets were developed to display the common stimulus that accompanied items. The stimuli were provided in a separate book because seeing the same stimulus repeated throughout the OIB would break the flow of the OIB and can make it unwieldy. Panelists were able to identify items' associated stimuli by the identifying information on the item map in the column labeled "Set Leader__AIR Set Leader ID" or in the item information at the top of each items' page in the OIB. All grades and domains had stimulus booklets except Writing Grades 1 and 3 because the stimuli were relatively short in these grades.

5.3.4 Rubrics

Scoring rules were provided for machine-scored items for Listening in grades 3, 5, 7, and High School and for Reading grades 3, 5, and 7.

5.3.5 Scoring Guides

Writing scoring guides included the scoring rubrics and exemplars of student responses for Writing constructed response items that were included in the OIBs. Speaking scoring guides included the rubrics and links to .mp3s of student exemplars for Speaking constructed response items that were included in the OIBs. The .mp3 links were accessed using the .PDF copy of the Speaking scoring guide on panelists' computers.

5.3.6 Cluster-Scored Set Booklets

Cluster-scored sets were included in the Speaking OIBs in Kindergarten and grades 1, 3, and 5. Cluster-scored sets are multiple dichotomous items that are calibrated and scored as a single partial-credit item. The cluster-scored sets were provided in a separate book because seeing the same cluster-scored set repeated throughout the OIB, once for each positive score point, would break the flow of the OIB and make it unwieldy.

5.3.7 Target Student Descriptors

Target Student Descriptors were developed to support panelist decision making as described in Section 4.2 Target Student Descriptor Writing.



ELPA21 Standard Setting July 2016

Grade 3 Reading Item Map

Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID_AIR ITS Item ID	Set Leader__ AIR Set Leader ID	Task Type	Task Sub Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	381	1	VHxxxxxx _xxxx		Read-Along Sentence			
2	440	1 of 2	VHxxxxx_ _xxxx	VHxxxxx _xxx	Procedural Text			
3	453	1	VHxxxxx_ _xxxx		Read and Match	Sentence		
4	481	2 of 2	VHxxxxx_ _xxxx	VHxxxxx _xxx	Procedural Text			
5	485	1	VHxxxxx_ _xxxx	VHxxxxx _xxx	Short Correspondence			
6	510	1	VHxxxxx_ _xxxx		Read and Match	Word		
7	539	1	VHxxxxx_ _xxxx	VHxxxxx _xxx	Short Correspondence			
8	552	1	VHxxxxx_ _xxxx	VHxxxxx _xxx	Short Correspondence			

Figure 4. Sample Item Map

5.4 Room Configuration

Four rooms were used to support the in-person standard setting workshop.

General Assembly Room. All panelists worked in the St. Louis Westin’s Cupples Ballroom. The ballroom was of sufficient size to hold 12 tables (2 tables per grade band × 6 grade bands). Each table was a “round of 6” seating 4–5 panelists with room for a laptop and a 24” monitor. Headsets and a headphone splitter and amplifier were provided so that panelists could listen to the Listening items and Speaking .mp3 exemplars without disturbing panelists at other tables.

A podium with a microphone was at the center front of the room. Two projectors were set up so that all panelists could easily view presentations.

Tables for observers were located against the walls where they would not distract panelists.

Breakout Room. A dedicated breakout room—the St. Louis Westin’s Plaza—was maintained for meetings and training of panelists. The breakout room was used for Bookmark training and for achievement level descriptor training that occurred for panelists who finished specific tasks early.

Operations Room. A secure workspace was dedicated for the storage of confidential materials and for Pacific Metrics staff to conduct analyses and work in support of the workshop. A limited set of keys were available to this secure room.

Dining Room. A separate room was used each day for lunch.

5.5 Workshop Roles

5.5.1 Pacific Metrics Staff

Table 10 summarizes Pacific Metrics staffing for the workshop.

Table 10. Pacific Metrics Staffing Roles and Experience

Name	Workshop Role	Responsibility and Experience
Dr. Daniel Lewis, Pacific Metric	Senior Advisor and Workshop Facilitator	Dr. Lewis co-developed the Bookmark Standard Setting Procedure and was primary author of the Bookmark Chapter in Cizek’s seminal book on Setting Performance Standards (2 nd Ed.). Dr. Lewis has designed and facilitated numerous standard settings including those for the Smarter Balanced Assessment Consortium and the California English Language Development Test.
Dr. Karla Egan, Pacific Metrics Consultant, EdMetric LLC	Workshop Facilitator	Dr. Egan designed and led over 40 standard setting workshops, including multiple workshops for English language assessments. Dr. Egan co-created a framework for achievement level descriptors, and she was the lead author of the ALD chapter in Cizek’s seminal book on Setting Performance Standards (2 nd Ed.). She designed and facilitated the Smarter Balanced Assessment Consortium Achievement Level Descriptor Writing Workshop.

Adele Brandstrom	Operations Room Manager	Ms. Brandstrom has more than 15 years of standard setting experience and has created and managed materials for over 100 standard setting workshops. Ms. Brandstrom was responsible for developing standard setting materials (e.g., ordered item booklets, item maps, etc.) and the setup and tracking of secure materials during the in-person standard setting workshop.
Kyle Beekman	Technology Specialist and Runner	Kyle Beekman is a Data Analyst at Pacific Metrics. He is well versed in technology requirements and was responsible for computer setup, equipment testing, sound quality, and assuring that panelists are able to access online tools for viewing test content. Mr. Beekman was available to support communication and materials movement between the general assembly room, break out room, and operations room and supported panelists as needed.
Leslie Mugan	Program Manager	Leslie Mugan is Vice President of Customer Experience & Program Management for Pacific Metrics and supported meeting (rooms, meals) and travel logistics. She also supported panelist travel and hotel issues.
Holly Garner	Support and Facilitation	Holly Garner is Director of Sales for Pacific Metrics and provided support as needed.

5.5.2 ELPA21, EL Expert, State Representatives, and Observer Roles

ELPA21 staff, an EL Expert, ELPA21 state representatives, an auditor, and other ELPA21 vendors attended the in-person workshop. A description of their roles follows.

ELPA21 and Sponsoring Agency Staff. Mike Middleton, a representative from ELPA21 and the State of Washington—welcomed the panelists to the in-person workshop. Throughout the workshop, the ELPA21 staff—Ms. Cathryn Still, Program Director of ELPA21 and Dr. Mary Seburn, Task Management Team Lead for Performance Standard Setting, Data, and Reporting, provided an overview of ELPA21, the purpose of the in-person workshop, and the place the in-person workshop held in the multi-step process implemented to adopt cut scores for the ELPA21 assessments. Ms. Still and Dr. Seburn also responded to panelists’ questions associated with policy throughout the workshop.

EL Expert. One EL Expert from the EL Expert Advisory Panel—Carolyn Stearns, District EL Teacher, Perry Local Schools, Ohio—attended the in-person workshop as a subject matter expert. The EL Expert supported panelists from all grade bands and was not a voting member of any panel. During the workshop, Ms. Stearns supported discussion of the ALIs, ELP Standards and Target Student Descriptors and provided guidance to panelists on interpreting scoring rubrics.

State Representatives. ELPA21 State Representatives were silent observers when panelists were working and attended daily debriefs to provide advice in response to issues that arose that day. Their attendance was also important in that state representatives are able to communicate the nature of the in-person standard setting workshop and the validity of the process to other stakeholders in their respective states and to their fellow state department staff members.

The following State Representatives attended the workshop:

Colleen Anderson, Iowa
 Jenny Choi, Washington
 Kim Hayes, Washington
 Mike Middleton, Washington
 Terri Schuster, Nebraska
 Steve Slater, Oregon
 Kurt Taube, Ohio

Auditor. An external auditor, Dr. Gregory J. Cizek (Professor of Educational Measurement and Evaluation, University of North Carolina at Chapel Hill), was present as a methodological subject matter expert and an objective evaluator to assess the validity of the in-person workshop processes and results. Dr. Cizek submitted a written evaluation of the in-person standard setting to support ELPA21 in their decision-making with respect to the adoption of cut scores. The auditor’s report is provided in Appendix H. The auditor was a silent observer of the process and attended, to the degree possible, all workshop activities.

Additional Observers. ELPA21 invited representatives from the states’ administration, scoring, and reporting vendors, AIR and Questar, to observe the standard setting. Representatives from these vendors were provided the opportunity to attend the in-person activity to establish transparency in process and to support a shared understanding of ELPA21 practices across states. The vendor-representatives observed quietly from the perimeter of the room, did not move about or distract panelists, and did not attend daily debriefs.

5.5.3 Panelist Roles, Recruitment, and Composition

Panelists and Alternates. The 49 initial panelists were selected to represent the population of eligible panelists and served as subject matter experts with respect to the KSPs English Learners need to place in each achievement level. Panelists actively participated in training, studied test materials, and provided recommendations.

Three alternate panelists—one each for elementary, middle, and high school—attended in case one of the scheduled panelists was unable to attend or needed to leave before the process was complete. In this case, one panelist was unable to attend the workshop and the remaining alternates were asked to join as panelists. Thus, a total of 51 panelists participated.

Table Facilitators. Each grade band panel was seated at one of two tables with four to five panelists per table. The Table Facilitators received special training the week prior to the workshop and daily, as needed, to help facilitate workshop activities at their tables. The Table Facilitators (a) took prompts from the Pacific Metrics Workshop Facilitators to begin scheduled activities at their table, (b) monitored table activities with respect to time, (c) facilitated discussion and supported the participation of each panelist at their table, (d) transferred panelists’ bookmark recommendations into Excel spreadsheets on their tables’ computers, and (e) supported secure materials collection at the end of each day. Table

Facilitators, unlike other panelists, attended the first part of daily debriefs to report on the “climate of the standard setting” or to raise concerns.

Table Facilitator training was provided during the week prior to the standard setting workshop via an online meeting that was recorded. Table Facilitators that were not able to attend the scheduled training were provided access to the .mp4 recording for viewing on their own schedule. The Table Facilitators also met prior to the opening session on the first day of the in-person workshop to become familiar with the materials, computers, and file set up. Training included an introduction to the standard setting process and round-by-round activities. Table Facilitators received a detailed, step-by-step annotated agenda to support their work. A copy of the Table Facilitator agenda is provided in Appendix E1.

Scribes. One or more panelist(s) at each table served as scribe and entered a summary of the table responses to questions on the item map on the table’s laptop. Scribes were self-selected.

Panelist Recruitment and Composition

ELPA21 recruited eight to nine panelists per grade band plus three alternates for a total of 52 panelists. The panelists were recruited from the ELPA21 states. Each state was invited to contribute panelists.

ELPA21 selected 49 panelists (plus 3 alternates) from 80 nominees based on the following:

- Grade-level EL experience
- Additional experience (e.g., working with students with disabilities)
- Familiarity with the ELP Standards
- Knowledge of and prior involvement in ELPA21 (e.g., item writing or reviewing)
- Availability for travel and the workshop
- State representation across grade-level tables and panels
- Willingness to facilitate during the workshop

The composition of the final workshop panels mirrored the composition of other consortium workshops and is representative of ELPA21 educators and stakeholders. A summary of panelist composition follows:

- Represented the eight member states: AR (13%), IA (12%), KS (15%), NE (6%), OH (21%), OR (13%), WA (12%), WV (8%)
- Mostly female
- Were White (80%), Hispanic (10%), and Asian/Pacific Islander, Decline to Answer, Other (10%)
- 35% were from rural areas, 40% were suburban, and 24% were urban
- 58% were from Title I schools
- 48% had experience with elementary ELs, 14% had experience with middle school ELs, and 22% had experience with high school ELs
- Collectively experienced in teaching students from over 40 different languages

Participant backgrounds provided multiple areas of expertise; participant titles included District EL Teacher, EL Teacher, EL Coordinator, EL Specialist, District EL Support, EL Facilitator, Instructional Coach, Principal, Language Academy I teacher, Lead EL Teacher, and Title III EL Specialist.

5.6 Workshop Events

The following sections provide a detailed description of the in-person workshop activities. A more detailed step-by-step summary of the in-person workshop events is provided as directly observed and described in the auditor's report in Appendix H.

5.6.1 High-Level Workshop Agenda

The high-level agenda in Table 11 was provided to panelists, ELPA21, and all observers.

Table 11. Standard Setting High-Level Agenda

Note: Times are approximate and will be adjusted as needed. Appropriate breaks will be provided throughout.	
Tuesday, July 19: Day 1	
Tuesday Morning	<p>7:30 AM: Table Facilitator Meeting in Cupples Ballroom</p> <p>8:00 AM: Opening Session: Welcome and Training</p> <ul style="list-style-type: none"> • ELPA21 welcomes panelists, provides background information associated with test, and highlights relevant policy information • Dan Lewis provides overview of standard setting and describes relevance of Contrasting Groups Study (CGS) and EL Expert work to date • Karla Egan provides overview of Bookmark Standard Setting and prepares panelists for their first activities <p>9:30 AM: Table-level Introductions and Secure Materials Sign-out</p> <p>9:45 AM: Q & A for the ELP Standards and ALLs for Reading</p> <p>10:00 AM: Table review of Reading online operational form</p> <p>10:30 AM: Study Reading ordered item booklet (OIB)</p> <ul style="list-style-type: none"> • Review items one-by-one, together as a table, in order of difficulty <p>Noon: Lunch</p>
Tuesday Afternoon	<p>1:00 PM: Reading Round 1 (R1)</p> <ul style="list-style-type: none"> • Discuss the Reading Target Student Descriptors, which describe the knowledge, skills, and processes (KSPs) associated with students at the threshold of an achievement level • Bookmark Training: How to express recommendations for the KSPs needed for threshold Level 3 and Level 4 students • Bookmarking Readiness Survey • Panelists independently place R1 bookmarks <p>2:30 PM: Reading Round 2 (R2)</p> <ul style="list-style-type: none"> • Review R1 results and CGS bookmarks at each table • Table-level discussion: Difference in panelists' R1 ratings • Panelists independently place R2 bookmarks <p>3:30 PM: Reading Round 3 (R3)</p> <ul style="list-style-type: none"> • Presentation of Table 1, Table 2, and Overall R2 median ratings and bookmarks associated with CGS cut scores • Presentation of impact data based on Overall R2 medians • Group discussion • Panelists independently place R3 bookmarks <p>4:45 PM: Session Close (time approximate; the day may be extended if necessary to complete a task)</p> <ul style="list-style-type: none"> • Collect and audit secure materials • Review final Reading results • Complete Reading evaluation

Tuesday End of Day	5:15 PM: Daily Debrief in Breakout Room (CCSSO, Pacific Metrics, Table Facilitators) <ul style="list-style-type: none"> Table Facilitators attend first part of debrief
Wednesday, July 20: Day 2	
Wednesday	7:30 AM: Table Facilitator Meeting in Cupples Ballroom 8:00 AM: Listening <ul style="list-style-type: none"> Review the ELP Standards and ALLs for Listening Table review of Listening online operational form Study Listening OIB Study Listening Target Student Descriptors Listening R1, R2, and R3 Noon: Lunch 1:00 PM: Writing <ul style="list-style-type: none"> Review the ELP Standards and ALLs for Writing Table review of Writing online operational form Study Writing OIB 4:45 PM: Session Close (time approximate; the day may be extended if necessary to complete a task) <ul style="list-style-type: none"> Collect and audit secure materials
Wednesday End of Day	5:15 PM: Daily Debrief in Breakout Room (CCSSO, Pacific Metrics, Table Facilitators) <ul style="list-style-type: none"> Table Facilitators attend first part of debrief
Thursday, July 21: Day 3	
Thursday	7:30 AM: Table Facilitator Meeting in Cupples Ballroom 8:00 AM: Writing (continued) <ul style="list-style-type: none"> Study Writing Target Student Descriptors Writing R1, R2, and R3 Noon: Lunch 1:00 PM: Speaking <ul style="list-style-type: none"> Review the ELP Standards and ALLs for Speaking Table review of Speaking online operational form Study Speaking OIB Study Speaking Target Student Descriptors Speaking R1, R2, and R3 (as time allows) 4:45 PM: Session Close (time approximate; the day may be extended if necessary to complete a task) <ul style="list-style-type: none"> Collect and audit secure materials
Thursday End of Day	5:15 PM: Daily Debrief in Breakout Room (CCSSO, Pacific Metrics, Table Facilitators) <ul style="list-style-type: none"> Table Facilitators attend first part of debrief

Friday, July 22: Day 4	
Friday	<p>7:30 AM: Table Facilitator Meeting in Cupples Ballroom</p> <p>8:00 AM: Speaking (continue to complete)</p> <ul style="list-style-type: none"> • Collect and audit secure materials <p>Noon: Lunch</p> <p>1:00 PM: Cross-Grade Activities—Table Facilitators and panelists divided among the following three groups:</p> <ul style="list-style-type: none"> • Achievement Level Descriptor Writing Group • Vertical Articulation Group • Proficiency Determination Group <p>4:45 PM: Session Close</p> <ul style="list-style-type: none"> • Collect and audit secure materials • Complete evaluations
Friday End of Day	<p>5:15 PM: Daily Debrief in Breakout Room (CCSSO, Pacific Metrics, Table Facilitators)</p> <ul style="list-style-type: none"> • Table Facilitators attend first part of debrief (as flight times allow)

5.6.2 Opening Session

The opening session began at 8 AM, July 19, 2016. The following presentations provided an introduction to the workshop activities, after which the panelists began their work on the first domain—Reading.

Welcome Address. Michael Middleton, ELPA21 Consortium Council member from Washington, welcomed panelists to the workshop and provided the context for the panelists’ work.

Introduction to Standard Setting. Dr. Daniel Lewis introduced the concept of standard setting to panelists and provided an introduction to the specifics of the ELPA21 standard setting including the number and names of the achievement levels, the Policy Descriptors, and the approach for setting the Levels 3 and 4 cut scores directly for some grades and using the CGS and interpolation for other grades and cut scores. Dr. Lewis also provided an overview of the CGS.

Introduction to the Bookmark Standard Setting Procedure. Dr. Karla Egan overviewed the Bookmark Procedure, and described the materials that would be used by panelists including OIBs, Item Maps, stimulus booklets, rubrics and exemplars. Drs. Lewis and Egan’s opening session slides are provided in Appendix D2.

Readiness Survey. A readiness survey was administered after the opening session to determine whether panelists felt prepared to begin working on the first domain. The readiness survey and a summary of the responses are provided in Table 12 and Appendix G1. As indicated in Table 12, positive responses with respect to readiness to begin the next activity were endorsed by approximately 98% of panelists. This indicated that the vast majority of panelists understood the goals of, and their role in, the workshop and how to proceed with the subsequent activities. Panelists who indicated that they needed additional training or had questions were supported.

Table 12. Post-Opening-Session Readiness Survey, Questions 10-8

Item No.	Question	SD		D		A		SA		Total
		N	%	N	%	N	%	N	%	
1	The orientation session provided a clear overview of the standard setting process	0	0%	0	0%	30	59%	21	41%	51
2	I understand the goals of the standard setting workshop	0	0%	0	0%	26	51%	25	49%	51
3	I understand my role in the standard setting workshop	0	0%	0	0%	24	47%	27	53%	51
4	The orientation session provided a clear overview of the Contrasting Groups Study	0	0%	1	2%	30	59%	20	39%	51
5	I understand how the Contrasting Groups Study data will be used in the standard setting process	1	2%	2	4%	27	53%	21	41%	51
6	The orientation session provided a clear explanation of the development of ELPA21	0	0%	5	10%	28	55%	18	35%	51
7	I understand how the results of the standard setting will be used to support the reporting of ELPA21 results	0	0%	1	2%	28	55%	22	43%	51
8	I understand how to study the items in the ordered item booklet	0	0%	0	0%	27	53%	24	47%	51
Total Responses		1	0%	9	2%	220	54%	178	44%	408

Table 13. Post-Opening-Session Readiness Survey, Questions 9-10

Item No.	Question	No		Yes		Total
		N	%	N	%	
If you answered Disagree or Strongly Disagree to any of questions 1–8, then please answer the next two Yes/No questions.						
9	I would like additional training on studying the ordered item booklet.	28	90%	3	10%	31
10	I have additional questions on material presented during the opening session that I would like answered before I begin the next task.	29	94%	2	6%	31
Total Responses		57	92%	5	8%	62

5.6.3 Process for Recommending Cut Scores for Each Domain

The following set of activities was conducted similarly for each domain. Any variance for specific domains is described in the section in which the variance occurred. Where no

variance is indicated all domains followed the process described. The process repeated for each domain can be conceptually grouped into the following three sets of activities:

1. Provide an opportunity for panelists to develop a comprehensive understanding of what each domain's test measures.
2. Develop a common understanding of the Level 3 and Level 4 Target Student—the students at the Level 3 and Level 4 cut score, respectively. Target Students are students just at the threshold of the given achievement level.
3. Engage in three rounds of Bookmarking—the task in which panelists make their cut score recommendations.

1. Provide an opportunity for panelists to develop a comprehensive understanding of what each domain's test measures.

Panelists were provided with the opportunity to develop a comprehensive understanding of what each domain's test measures by completing the following three activities—review the ELP standards and ALIs for the given domain, review a complete online form of the ELPA21 test for the given domain, and study the ordered item booklet for the given domain. These are described as follows:

Review the English Language Proficiency (ELP) Standards and ALIs. Panelists were provided with a link to the ELP standards and ALIs prior to standard setting and were asked to review materials before they came to the in-person workshop. A copy of the grade band ALIs for each domain was provided for each panelist and an electronic copy of the ELP Standards was provided on the computer at each table. The on-site EL expert—Ms. Carolyn Stearns—was introduced after the opening session; panelists were provided time to review and discuss the ALIs and ELP standards at their tables and ask questions of Ms. Stearns in the group setting.

Review one form of the operational test. One form of the operational ELPA21 assessment was reviewed jointly by panelists at each table using the same secure browser used by students. Panelists reviewed the complete form that composed, with supplemental items, their OIBs. Panelists reviewed the test for two domains at a time; Listening and Reading were reviewed first, followed by the setting of cut scores for Reading and Listening (in that order). Writing and Speaking were reviewed two days later, followed by the setting of cut scores for Writing and Speaking (in that order). The order of form review was established by the test delivery vendor and was the order the forms were presented to students.

Headsets were provided for each panelist and an audio amplifier with multiple headphone outputs was provided to support sound quality. Panelists with questions about items posed them to the EL Expert, Ms. Stearns.

Study the ordered item booklet (OIB). Panelists studied OIBs in a structured, table-level activity led by their Table Facilitator. The OIB is a set of test items selected to be representative of the construct measured by each form of the test, ordered by difficulty. A detailed description of the design and development of the ordered item booklet (OIB), is provided in Section 5.3.1.

The following materials were used when studying the OIB:

1. **Ordered Item Booklet.** The OIB had one item per page with the easiest item first and most difficult item last (with difficulty defined as IRT location, as specified in Section 5.2 Item Mapping: Locating Items on the IRT Scale Each page of the OIB was associated with an item and includes a rendering of the item and item information such as ELP Standards, Sub-Claims, Task Type and Task Subtype, and correct response.
2. **Stimulus Booklet.** Longer stimuli common to multiple items were presented in a separate booklet to maintain the flow of the OIB. This booklet was ordered by Stimulus ID, which could be found on each associated page in the OIB.
3. **Item Map.** Item maps supported the review of the OIB. A sample item map is provided in Figure 4. The item map included the following information:
 - Order of Difficulty (OIB page number)
 - Location: Item IRT RP67 location on an interim scale (not the final reporting scale). The location is the scale score at which a student would have a .67 likelihood of success on the item.
 - Score Point: 1 for dichotomous items or the score point being considered for polytomous/partial credit items, e.g., 1 of 2, 2 of 2
 - Item and Set Leader ID
 - Task Type and Task Sub Type
 - The last two columns of the item map pose the following two questions that were to be answered by the panelists in table discussion. The discussions were intended to support deeper understanding of the test items.
 1. What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?
 2. Why is this item or score point more difficult than the items that precede it?

The panelists discussed each item in the OIB. In particular, they discussed the KSPs measured by each item as well as why the item was more difficult than the items that preceded it, as indicated in Figure 5.

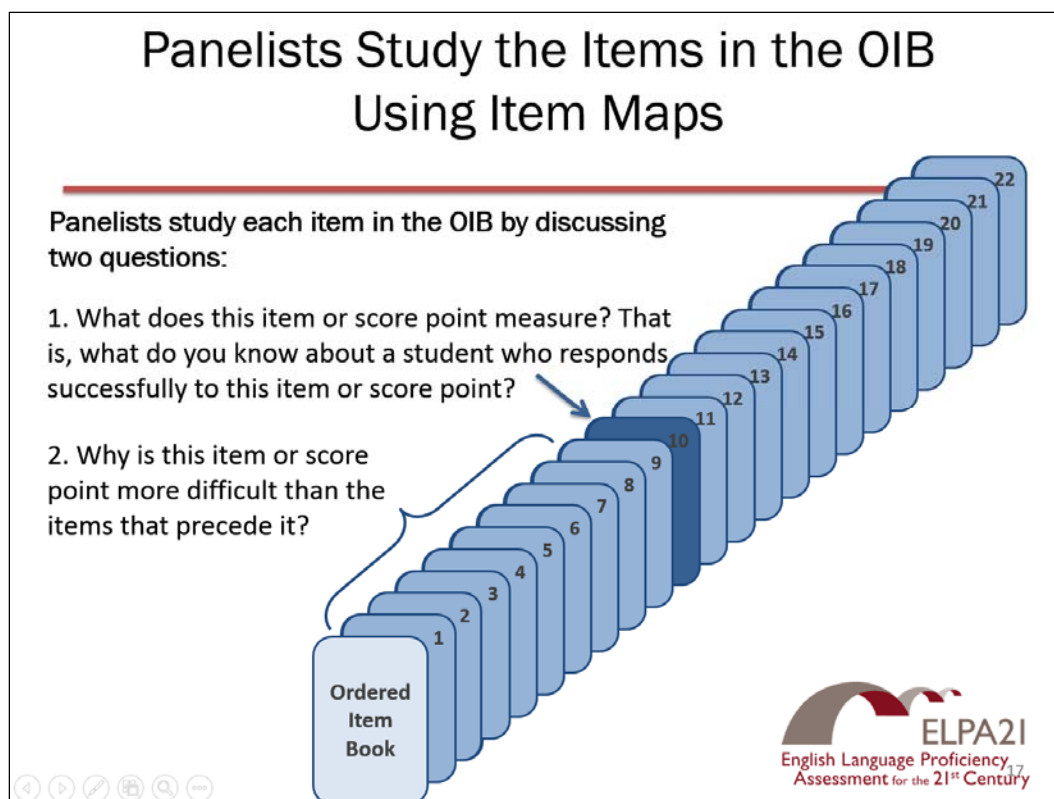


Figure 5. Studying the Ordered Item Booklet

Studying the OIB: Procedural Details (As described in Table Facilitator training and in the Table Facilitator agenda). The Table Facilitator supports this activity by making sure that each panelist at the table is reviewing the same item.

- The Table Facilitator asks the scribe to display the first item in the OIB on the monitor.
 - Panelists turn to that item in their OIBs and locate it on their item maps.
 - After viewing the item, the panelists locate the item's stimulus (if necessary) in the Stimulus Booklet, review the rubric in the Scoring Guide (if necessary), and read/listen to the exemplar responses (if necessary; speaking exemplars are mp3 files that play when the link in the Scoring Guide is clicked).
 - The Table Facilitator asks the first question and facilitates discussion at the table: What does this item measure? That is, what do you know about a student who responds successfully to this item?
 - The Scribe takes notes in the item map that is displayed in Excel on the monitor.
 - Table facilitator asks the second question and facilitates discussion at the table: Why is this item more difficult than the preceding items?
 - The Scribe takes notes.

- The Table facilitator monitors time, supports discussion by trying to have all panelists contribute appropriately, and moves to the next item when discussion is complete.
- This process repeats for all items in the OIB.

2. Develop a common understanding of the Level 3 and Level 4 Target Student—the students at the Level 3 and Level 4 cut score, respectively. Target Students are students just at the threshold of the given achievement level.

Panelists set cut scores for the Target Student—the student who is just at the threshold of the level. It is important for panelists to develop a common understanding of the Target Student. That is, although panelists may differ on the skills that students should hold to enter Level 4, for example, they must develop a clear conceptual understanding of who that student is—that the Target Student is not in the middle of the achievement level, but just entering the level. Understanding the Level 4 Target Student was intended to help them make their judgments, which are intended to differentiate the skills associated with just leaving Level 3, and just entering Level 4.

Following best practice, the panelists also begin with a Target Student Descriptor to provide a common starting point to help reinforce their conceptualization of the Target Student. Panelists may place bookmarks that are based on the skills described by the Target Student Descriptor or are free to place their bookmarks such that fewer, or more, skills are required to enter Level 4 based on their expert judgment. However, they all begin with the common understanding established by the Level 4 Target Student Descriptor developed by the EL Expert Advisory Panel prior to the workshop.

Panelists reviewed the Target Student Descriptors at their table and were provided an opportunity to ask questions of the EL Expert for clarification, if necessary. After their introduction to, and review of the Target Student, the panelists were ready to begin training on how to make their bookmark recommendations.

3. Engage in three rounds of Bookmarking—the task in which panelists make their cut score recommendations.

Training panelists on the judgment task: Bookmarking. Dr. Karla Egan made a presentation on the task of placing Bookmarks to make their cut score recommendations. Training was conducted in two sessions in the breakout room to allow panelists to complete studying the OIBs and reviewing Target Student Descriptors at their own pace.

Figure 6 describes the fundamental task for placing the Level 4 bookmark. Level 4 is the anchor level—Early Advanced. The Bookmark training slides used by Dr. Egan are provided in Appendix D3.

Panelists were provided with physical bookmarks that reflect their meaning, as illustrated in Figure 7.

Bookmark training included ample time for questions, answers, and clarifications.

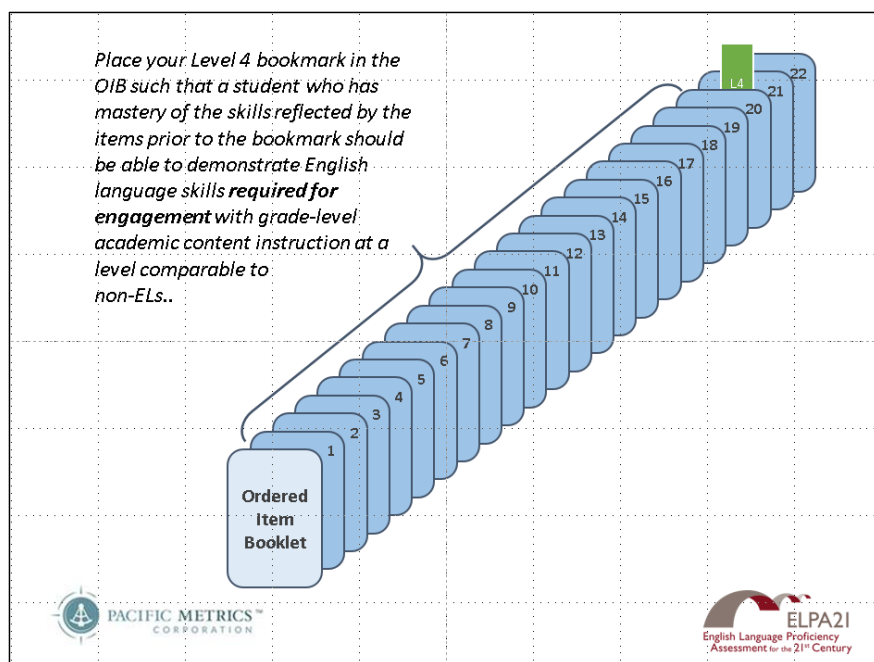


Figure 6. Bookmark Training Slide



Level 3 Bookmark	Level 4 Bookmark
 <p>ELPA21 Standard Setting</p> <p>Level 3 Bookmark</p> <p>I am placing my Level 3 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark applies some grade-level English language skills and will benefit from EL Program support.</p>	 <p>ELPA21 Standard Setting</p> <p>Level 4 Bookmark</p> <p>I am placing my Level 4 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark should be able to demonstrate English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.</p>
Level 3 Bookmark	Level 4 Bookmark

Figure 7. Bookmarks

Bookmark Training Readiness Survey. A Bookmark training readiness survey was administered to determine whether panelists felt prepared to begin their next task—bookmarking. The Bookmark training readiness survey and a summary of the responses are provided in Table 14 and Appendix G2. As indicated in Table 14, 99.5% of responses were positive with respect to readiness to begin the next activity, indicating that the vast majority of panelists understood how to place their bookmarks and were prepared to engage in the fundamental task—bookmarking. As shown in Table 15, no panelists indicated that they needed additional training or had questions.

Table 14. Bookmark Training Readiness Survey and Results

Item No.	Question	SD		D		A		SA		Total
		N	%	N	%	N	%	N	%	
1	I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors	0	0%	1	2%	17	33%	33	65%	51
2	I participated in bookmark training and had an opportunity to ask questions and discuss the meaning of the Level 3 and Level 4 bookmarks	0	0%	0	0%	19	37%	32	63%	51
3	I understand how to place my bookmarks	0	0%	0	0%	19	37%	32	63%	51
4	I understand I will have opportunities to change my bookmarks in Rounds 2 and 3	0	0%	0	0%	17	33%	34	67%	51
Total Responses		0	0%	1	1%	72	35%	131	64%	204

Table 15. Bookmark Training Readiness Survey and Results

Item No.	Question	No		Yes		Total
		N	%	N	%	
If you answered Disagree or Strongly Disagree to any of questions 1–4, then please answer the next two Yes/No questions.						
5	I would like additional training on placing my bookmarks for Round 1.	19	100%	0	0%	31
6	I have additional questions that I would like to ask before placing my Round 1 bookmarks.	19	100%	0	0%	31
Total Responses		38	100%	0	0%	38

Round 1

Panelists placed bookmarks that reflected their recommendations. Panelists were instructed to begin by making judgments associated with Level 4 (the target level), as described in Figure 6, followed by the Level 3 bookmark. This activity was done independently and silently. Participants recorded their bookmarks on a rating form. A sample rating form is provided in Appendix E2.

Documenting Panelist Ratings: Procedural Details (As described in Table Facilitator training and in the Table Facilitator agenda). After panelists entered their ratings on the paper rating form they were instructed to give them to the Table Facilitator. The Table Facilitator was trained to check to see that each panelist made two ratings and that the Level 4 bookmark was higher than the Level 3 bookmark (validity check).

The Table Facilitator entered each panelist's bookmarks into an Excel spreadsheet on the computer. The Facilitator was trained to save the Excel file after entering the bookmark values and to have another panel member at the table QA the entries.

A screenshot of a Round 1 Rating Form Data Entry Worksheet is provided in Figure 8.

	A	B	C	D	E	F	G	H	I	J
2							Enter Panelists' Level 3 Bookmarks Below	Enter Panelists' Level 4 Bookmarks Below		
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

Figure 8. Round 1 Rating Form Data Entry Worksheet

The Excel files were stored in Dropbox, a cloud-based file storage and synchronization service. As panelists entered and saved data into their excel files the entries were stored and synchronized in Dropbox so that they could be accessed by the Pacific Metrics staff. Thus, an instance of the saved file was available in the Dropbox folder on the Operations Room and Workshop Facilitators' computers. The data was reviewed by the Operations Room Manager as a final validity check and saved in a separate location as a backup and for technical documentation.

The display materials needed for Round 2 were automatically populated in the next tab in the Excel file. A screenshot of a Round 2 Display Worksheet is provided in Figure 9.

	A	B	C	D	E	F	G	H	I	J	K	L	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
			Results from Round 1 Displayed at the Beginning of Round 2																									
			Review and Discuss Round 1 Level 4 Bookmarks													Review and Discuss Round 1 Level 3 Bookmarks												
1			1Place a green post-it in your ordered item booklet on the following pages to represent panelists' Round 1 Level 4 Bookmarks: Pages41, 21, 41, 33,													5Place a blue post-it in your ordered item booklet on the following pages to represent panelists' Round 1 Level 3 Bookmarks: Pages29, 11, 28, 18,												
3																												
4																												
6			2Skip this step													6Skip this step												
8			3Discuss your rationales for any differences in your Round 1 Level 4 bookmark ratings.													7Discuss your rationales for any differences in your Round 1 Level 3 bookmark ratings.												
10			4Panelists record their Round 2 Level 4 Bookmark Ratings													8Panelists record their Round 2 Level 3 Bookmark Ratings												
12																9Record panelists' new Level 3 and Level 4 bookmarks in "Enter R2 Ratings" tab, QA the entries, and save the file.												
14																												
15																												
16																												
			R2 Display	Enter R2 Ratings	R3 Display	Enter R3 Ratings	Post R3 Display	domain grade table	G3 RD Item Map	FDs																		

Figure 9. Round 2 Display Worksheet

Round 2

Round 2 began with panelists considering their table's Round 1 bookmarks. The original standard setting design called for panelist review of the CGS results in Round 2 with discussion about its relevance and relationship to the panelists' Round 1 results. However, the CGS results were not yet available at the time of the in-person workshop. Thus, Steps 2 and 6 in Figure 9, where panelists were expected to place bookmarks representing the Level 4 and Level 3 CGS cut score estimates were eliminated and the excel files were updated to state "Skip this step" where instructions for interpreting the CGS results were displayed previously.

Not incorporating the CGS results as an external benchmark does not reduce the validity of the standard setting. It was expected to be a positive source of information that would elicit discussion, but it was not a required component of the standard setting design.

Panelist placed their Round 1 bookmarks independently. Seeing the diversity of other panelists' bookmark recommendation, sharing the rationales for their own bookmarks, and hearing the rationales for their fellow panelists' differing recommendations in the Round 2 phase either reinforces panelist's commitment to their Round 1 ratings or provides a perspective that results in modification of their original ratings.

The panelists discussed the information provided. They were instructed to ground their conversations in terms of skills reflected by items in the OIB, specifically with respect to the skills students should have to be in each level. Panelists were not required to reach consensus during this process. Following discussion, the panelists placed their Round 2 bookmarks independently and recorded their ratings.

In Round 3, the two tables in each grade join together to view (a) each panelists' bookmarks, (b) the median bookmark for each table and for all grade level panelists, and impact data associated with the grade level median bookmark.

Dr. Lewis' presentation provided the context for the information that panelists would see when the Round 3 display tab was opened (Table Facilitators were asked not to open it until after the presentation) and to prepare them for moving into the next domain when Round 3 was completed for the current domain. Dr. Lewis cautioned panelists not to "anticipate" where bookmarks might fall as they discussed the ordered item booklet for the next domain. That is, because panelists "knew the routine," they might be tempted to work together on bookmarking while studying the OIB for the next domain. Because independence of ratings is important for the first round, panelists were cautioned not to discuss bookmarks prematurely.

To prepare for Round 3, Dr. Lewis stated that having the two grade band tables work independently through Round 2 was important to the process, but that in Round 3 the two tables would come together to work as a single grade group. Four key pieces of information were conveyed in the presentation. First, panelists were asked to do a visual check to be certain that the links between the two tables' Excel files worked—that is, a visual check of the accuracy of each table's bookmarks was necessary to confirm the technology was working correctly. No problems occurred, however some table's Excel file links were required to be updated manually by the Workshop Facilitators because of the lag time between data entry and Dropbox updating of linked files.

Second, Dr. Lewis explained how the medians for each table and the grade were calculated and their relevance. Third, Dr. Lewis explained how the impact data associated with the grade level median was obtained—in this case it was based on data from some of the member states, but not all, and thus, caution in over-interpreting the impact data was urged.

Fourth, Dr. Lewis provided the context for the appropriate use of the impact data. That is, he explained that the Bookmark Procedure is considered a criterion-referenced exercise; however, it is considered best practice to bring norm-referenced information into consideration prior to the final round of ratings to provide a reality check to panelists. A version of the following script was conveyed to panelists.

"If the data seem reasonable based on the median bookmark location, then it supports the validity of that bookmark. If the data seem unusual to you—too rigorous or not rigorous enough—then it should prompt you to look again at the bookmarks, and it provides a direction for further discussion that may result in reconsidering your judgments for Round 3. In either case, you should make decisions based on skills and not on impact data. We do not want you to 'chase numbers.' However, the impact data should also not be a surprise—you know the students and what they can do based on your experience. You placed bookmarks based on your expectations and should have some idea of how well students will meet them. Of course this data is from multiple states and none of you are familiar with students across all the states.

If the impact data suggest your bookmarks may be too rigorous, then review the items before the grade band median bookmark to see if you were asking too much to be in the level. If the impact data suggest your bookmarks may not be rigorous enough, then review the items after the grade band median bookmark to see if you were asking too little to be in the level. You may decide that your recommendations were

appropriate; in that case do not let the impact data drive your decision making—it should be skills-based. You are not expected to change your bookmarks to achieve a specific percentage of students in each achievement level.

As your Table Facilitator shows you the impact data for your grade band, if you are surprised by the data, call a Workshop Facilitator to your table to help frame the discussion. Otherwise, discuss the differences between the table’s median bookmarks much as you discussed differences between panelists’ bookmark in the previous round. Begin with Level 4, indicate your Round 3 ratings on the rating form, and then repeat the process for Level 3.”

Workshop Facilitators floated between the tables to support panelists’ discussions. Following table discussions, the panelists independently placed their Round 3 bookmarks, Table Facilitators entered them into the Excel Round 3 Ratings worksheet, and the Excel Post-Round 3 Display tab provided final results to panelists. A Post-Round 3 Display tab is provided in Figure 11.

	A	B	C	D	E	F
1						
2	Results from Round 3				Level 3 Bookmark	Level 4 Bookmark
3			Grade Band (T1 & T2) Median		21	36.5
4						
5					Estimated percent of students based on R2 Results	
6					Levels 1 + 2	Level 3
7			Grade Band (T1 & T2)	32.1%	36.4%	31.6%
8						
9	Table Facilitator Visual Check					
10	Table 1 Level 3 Bookmarks:	21	20	21	20	
11	Table 1 Level 4 Bookmarks:	38	37	38	33	
12						
13	Table 2 Level 3 Bookmarks:	21	21	21	21	
14	Table 2 Level 4 Bookmarks:	38	33	33	36	
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

Figure 11. Post-Round 3 Display Tab

Bookmark Process for the Remaining Domains. This process was repeated for each of the domains, however, the Bookmark training and Round 3 presentations were not repeated for each domain. That is, once panelists understood and had engaged in the process for one domain (Reading) they followed the agenda and the same methods for the remaining domains. The Workshop Facilitators and EL Expert monitored table activities to confirm that

panelists had an accurate interpretation of their bookmark placements and were following the appropriate methodology.

Daily Secure Materials Collection

Materials were collected in a systematic manner after Round 3 of each domain and at the end of each day to support the tracking of secure materials. That is, Table Facilitators were provided with a list of all secure materials and they observed for completeness as their table's panelists' stacked each piece of the color-coded (secure) materials in the specified order called out by the Table Facilitator. All secure materials were accounted for during, and at the conclusion of, the workshop. Materials were collected by Pacific Metrics staff and stored in the secure Operations Room nightly.

Daily Debrief

A daily debrief was conducted with ELPA21 staff, Pacific Metrics, the Auditor, the Table Facilitators, and state observers. Debriefs began with brief statements from each Table Facilitator with respect to the perceived validity and organization of the workshop and to report any issues that they were aware of. Panelists were generally very positive about the workshop; a few issues, mostly personality issues between panelists were raised, and support for Table Facilitators with such conflicts was arranged (i.e., the table(s) in question were closely monitored by the Workshop Facilitators and the EL Expert). No issues persisted through the workshop. Table Facilitators were dismissed after their comments and discussion of the issues that they raised.

ELPA21, state observers, and Pacific Metrics staff used the daily debrief for discussion and decision making with respect to issues that arose. For example, the decision to not provide CGS results as feedback to panelists was made after discussion in the daily debrief. In this case, it was observed that a representative set of student ELPA21 impact data was not yet available to match to CGS teachers' ratings; rather than present data that might not be reliable, a decision was made to proceed without bringing the CGS data into the in-person workshop.

5.7 Results

Table 16. Round 3 Results and Associated Impact Data provides the Round 3 results for all grades and domains, including the median bookmark for the grade and domain, associated scaled scores, and associated impact data. Appendix F1 provides the same information for each round by domain, grade, and table as well as the minimum and maximum bookmarks and associated scaled scores for each domain by grade and table. Recall that the scale provided is the interim scale developed specifically for standard setting purposes and is not the same as the operational reporting scale, which was established subsequent to the standard setting. The interim standard setting scale had a mean of 500 and standard deviation of 50 within each grade band. Recall also, that the impact data is based on a partial sample of the states' ELPA21 data and may not be representative of the ELPA21 census data (which is provided in Section 6 along with the cut scores on the ELPA21 reporting metric).

Table 16. Round 3 Results and Associated Impact Data

Domain	Grade	Median Bookmarks		Associated Cut Scores		Associated Impact Data*		
		L3	L4	L3	L4	%L1+L2	%L3	%L4+L5
Listening	K	24	38	482	544	0.32	0.48	0.19
	1	34	45	444	527	0.11	0.44	0.45
	3	21	43	442	521	0.05	0.31	0.64
	5	24	39	431	487	0.07	0.19	0.74
	7	26	45	442	511	0.15	0.26	0.59
	HS	19	36	448	507	0.18	0.47	0.35
Reading	K	23	37	469	541	0.23	0.51	0.26
	1	23	35	472	521	0.28	0.40	0.32
	3	21	36.5	495	543	0.35	0.35	0.30
	5	21	35	476	515	0.27	0.37	0.36
	7	17	34	475	516	0.42	0.33	0.25
	HS	16	31	484	535	0.49	0.35	0.15
Speaking	K	29	43	493	541	0.36	0.41	0.23
	1	49.5	57	525	600	0.68	0.30	0.02
	3	34.5	45	474	519	0.18	0.36	0.46
	5	33.5	43	455	507	0.11	0.38	0.50
	7	36	48	492	548	0.46	0.46	0.09
	HS	30	43	482	524	0.25	0.45	0.30
Writing	K	33	48	503	573	0.60	0.30	0.10
	1	48	53	519	540	0.42	0.41	0.17
	3	26.5	37	493	533	0.32	0.37	0.31
	5	27	39.5	455	517	0.14	0.49	0.37
	7	36	49	470	549	0.32	0.62	0.06
	HS	28	38	470	534	0.30	0.54	0.16

5.8 Reliability and Degree of Consensus of Panelists' Recommendations

The validity of a standard setting is supported by empirical evidence of the reliability of panelists' cut score recommendations and an increase in the degree of consensus over rounds.

Reliability. The reliability of panelists' cut score recommendations associated with the sampling of panelists was estimated using the methods described by Lewis, et. al. (1998); the approach and formulae are provided in Appendix F2. The standard errors in Table 17 are estimates of the standard error of the cut score (in the scale score metric) associated with the sampling of panelists. That is, assuming that the panelists were selected randomly from the population of qualified panelists, there is a two-thirds likelihood that other qualified panels would recommend bookmarks within one standard error of the scale score cut scores reported in

Table 16. Thus, smaller standard errors reflect more reliable cut score recommendations than larger standard errors. The magnitude of the standard errors can be interpreted in relation to the interim standard setting scale standard deviation of 50.

The standard errors for Level 4 range from 2, in Grade 1 Listening and Grade 5 Writing, to 35 in Grade 1 Writing with an average of 12.2 and a median of 12.5. The standard errors for Level 3 range from 0, in Grade 1 Listening and Grade 7 Reading, to 55 in Grade 1 Writing with an average of 10.5 and a median of 6.5. With few exceptions, these standard errors are small relative to the standard deviation, and none would be considered unusual.

Table 17. Standard Errors Associated with the Sampling of Panelists

	Listening		Reading		Speaking		Writing	
Grade	Standard Error		Standard Error		Standard Error		Standard Error	
	<u>L3</u>	<u>L4</u>	<u>L3</u>	<u>L4</u>	<u>L3</u>	<u>L4</u>	<u>L3</u>	<u>L4</u>
K	3	16	2	10	3	21	4	5
1	0	2	8	11	19	5	55	35
3	3	16	15	14	10	22	13	14
5	1	8	14	16	1	5	5	2
7	17	16	0	3	24	11	16	14
HS	14	15	4	6	3	20	17	6

Degree of Consensus of Panelists' Recommendations. Consensus among panelists' recommendations is not expected; however, an increase in agreement from round to round is evidence that panelists have considered the diversity of opinions in their grade groups and the discussion they engaged in to understand their differences.

Table 18 provides the cross-grade average range and standard deviation of scale score cut scores for each achievement level and domain. The data supports the validity of the panelists' interactions. That is, the average range and standard deviation uniformly decrease from round to round as would be expected in a consensus-building activity.

The average range and standard deviation of panelists' scale score cut score recommendations for each achievement level, domain, grade, table, and round are provided in Table 18. The consensus data in Table 18 shows that the range and standard deviation decreased from Round 1 to Round 3 in all cases at the grade level for each domain for Level 3 and Level 4. However, at the table level there were some increases in the range and standard deviation from Round 1 to Round 3. This occurred when a high degree of agreement in cut score recommendations occurred in Round 1. For the seven instances in which this occurred, the average Round 1 range was 5 scale score units and the average standard deviation was 2.5 scale score units, indicating that these tables had a high degree of consensus in the first round.

This consensus data indicates that panelists were influenced by feedback and discussions from Round 1 to Round 3 and supports the validity of panelists' interactions and the standard setting results.

Table 18. Average Range and Standard Deviation of Panelists' Scale Score Cut Score Recommendations by Domain, Achievement Level, and Round

Domain	Average Scale Score Cut Score Range				Average Cut Score Standard Deviation		
	Round 1	Round 2	Round 3		Round 1	Round 2	Round 3
Listening Level 3	22.67	5.50	4.17		5.75	2.39	1.35
Listening Level 4	12.83	6.00	3.67		4.26	2.76	1.63
Reading Level 3	15.83	5.67	2.33		5.63	2.24	0.84
Reading Level 4	22.67	8.17	3.50		6.83	3.57	1.36
Speaking Level 3	14.00	10.50	8.33		5.93	4.72	3.05
Speaking Level 4	9.50	4.83	3.67		3.34	1.98	1.33
Writing Level 3	14.83	7.00	5.50		5.21	3.23	2.12
Writing Level 4	11.00	8.50	5.50		3.98	3.07	2.02
All	15.42	7.02	4.58		5.12	2.99	1.71

5.9 Domain Evaluations

Panelists completed evaluations after Round 3 of each domain. The evaluations are provided and the results summarized for each domain by grade band in Appendices G3–G6.

The evaluation results indicated favorable endorsements by 98%, 99%, 98%, and 96% of panelists for Reading, Listening, Writing, and Speaking, respectively. The Speaking evaluations were likely lower, at 96% positive endorsement, because the Speaking OIBs were composed of polytomous and partial credit items, which are more challenging to review than domains with a mix of dichotomous and polytomous items. Further, grades K, 1, 3, and 5 Speaking OIBs had cluster item sets—sets of multiple dichotomous items scored as a single polytomous partial credit item. These more complex items were observed to be more cognitively challenging for panelists to engage with and may have resulted in the modest decrease in positive evaluations.

Selected cross-grade evaluation items are provided in Table 19, Table 20, Table 21, and Table 22 for Reading, Writing, Listening, and Speaking, respectively. The items were selected because they directly reflect panelists’ opinions with respect to validity.

The results indicate that the average percentage of positive endorsements (Agree [A] or Strongly Agree [SA]) for the selected evaluation items across all domains was 98%. These results support the validity of the standard setting and indicate that the vast majority of panelists (a) believed the process to be fair, (b) had a common understanding of the Target Student and Policy Definitions, (c) indicated they understood the bookmark task, (d) had sufficient time to complete their tasks, and (e) would defend the rigor of the recommended cut scores.

Table 19. Reading: Selected Evaluation Items

Question	SD		D		A		SA		Total
	N	%	N	%	N	%	N	%	
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0%	1	2%	17	33%	33	65%	51
My group shared a common understanding of the Target Students	0	0%	2	4%	23	45%	26	51%	51
The policy definitions were clearly communicated	0	0%	0	0%	24	47%	27	53%	51
I understood how to place my bookmarks	0	0%	0	0%	11	22%	40	78%	51
I had enough time to consider my bookmark placement	0	0%	0	0%	10	20%	41	80%	51
I feel the recommended standards that resulted from this process are reasonable	0	0%	0	0%	16	31%	35	69%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	1	2%	0	0%	22	43%	28	55%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0%	2	4%	21	41%	28	55%	51
Total Responses	1	0%	5	1%	144	35%	258	63%	408

Table 20. Writing: Selected Evaluation Items

Question	SD		D		A		SA		Total
	N	%	N	%	N	%	N	%	
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0%	0	0%	11	22%	40	78%	51
My group shared a common understanding of the Target Students	0	0%	1	2%	22	44%	27	54%	50
The policy definitions were clearly communicated	0	0%	0	0%	18	35%	33	65%	51
I understood how to place my bookmarks	0	0%	1	2%	13	25%	37	73%	51
I had enough time to consider my bookmark placement	0	0%	0	0%	10	20%	41	80%	51
I feel the recommended standards that resulted from this process are reasonable	0	0%	0	0%	14	27%	37	73%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0%	1	2%	15	29%	35	69%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0%	2	4%	14	27%	35	69%	51
Total Responses	0	0%	5	1%	117	29%	285	70%	407

Table 21. Listening: Selected Evaluation Items

Question	SD		D		A		SA		Total
	N	%	N	%	N	%	N	%	
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0%	0	0%	15	29%	36	71%	51
My group shared a common understanding of the Target Students	0	0%	1	2%	18	35%	32	63%	51
The policy definitions were clearly communicated	0	0%	0	0%	17	33%	34	67%	51
I understood how to place my bookmarks	0	0%	0	0%	11	22%	40	78%	51
I had enough time to consider my bookmark placement	0	0%	0	0%	11	22%	40	78%	51
I feel the recommended standards that resulted from this process are reasonable	0	0%	0	0%	25	49%	26	51%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0%	0	0%	26	51%	25	49%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0%	1	2%	25	49%	25	49%	51
Total Responses	0	0%	2	0%	148	36%	258	63%	408

Table 22. Speaking: Selected Evaluation Items

Question	SD		D		A		SA		Total
	N	%	N	%	N	%	N	%	
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0%	0	0%	14	27%	37	73%	51
My group shared a common understanding of the Target Students	0	0%	4	8%	22	44%	24	48%	50
The policy definitions were clearly communicated	0	0%	1	2%	18	35%	32	63%	51
I understood how to place my bookmarks	0	0%	0	0%	8	16%	43	84%	51
I had enough time to consider my bookmark placement	0	0%	2	4%	6	12%	43	84%	51
I feel the recommended standards that resulted from this process are reasonable	1	2%	1	2%	22	43%	27	53%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0%	2	4%	16	31%	33	65%	51
I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0%	1	2%	18	35%	32	63%	51
Total Responses	1	0%	11	3%	124	30%	271	67%	407

5.10 Proficiency Determination Review and Recommendations

The panelists worked on the final day, July 22, from 11 AM to the close of the workshop (approximately 5 PM) on the Proficiency Determination rules. The activity began with a presentation by Dr. Daniel Lewis that summarized the profile rules that emerged from the CGS and the EL Expert Advisory Panel Recommendations. These slides are provided in Appendix D5. Dr. Mary Seburn provided a summary of cross-grade differences in profiles observed in the CGS results.

A series of “thought questions” were posed to panelists to help them determine the profiles that separated (a) Proficient from Progressing and (b) Progressing from Emerging. The thought questions are included in the slides in Appendix D5.

One of the thought questions was “Should the profiles associated with Proficient, Progressing, and Emerging be the same for all grade bands?” An informal poll of panelists indicated that they did not agree on the answer to this question—some felt differences across grade bands should be permitted and others thought they should be the consistent across all grade bands.

Following Dr. Lewis’ presentation, panelists worked in grade groups to review and discuss the thought questions that had been placed on each table’s computer. Panelists were provided with a worksheet that allowed them to visualize specific profiles and were instructed to formulate descriptions of the profiles associated with Proficient and Emerging that were interpretable and exhaustive. They needed to be interpretable so that the meaning of Proficient, Progressing, and Emerging could be easily communicated to stakeholders. They needed to be exhaustive so that all profiles were associated with one and only one Proficiency Determination level.

The results for each grade band are provided in Table 23. A review of the results in Table 23 indicate that somewhat different profiles across grades emerged naturally, but the profiles do have commonalities. For example five of the six grade bands (the exception being grade band 4–5) argued that being Level 3 in one domain and Levels 4 or 5 in the other three domains was sufficient for a Proficiency Determination of Proficient; however, the domain in which the Level 3 could be varied across the grades. Similarly, most grade bands were focused on profiles including only Level 1 and Level 2 to be Emerging, with some allowing a Level 3 in one domain. This provided valuable information for the ELPA21 Consortium Council to consider as they adopted Proficiency Determination profile rules.

Table 23. ELPA21 Proficiency Determination Rules: Recommendations from In-Person Standard Setting Panelists

	Proficient	Progressing	Emerging
Grade	ELPA21 recommends that students be considered Proficient when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once Proficient on ELPA21, ELs can be considered for reclassification.	ELPA21 recommends that students be considered Progressing when they are approaching a level of English language necessary to produce, interpret, and collaborate, with support, on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... Students scoring Progressing on ELPA21 are recommended to receive program support.	ELPA21 recommends that students be considered Emerging when they have not yet attained a level of English language necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... Students scoring Emerging on ELPA21 are eligible for ongoing program support.
High School	• All 4s or 5s with one 3	• Two 3s or higher with one 1	• All 1s and 2s with no more than one 3
Grade 7	• All 4s or 5s with a 3 in Speaking ^a	• 2 in Reading and Writing	• A 1 in Reading or Writing
Grade 5	• All 4s or 5s	• More than two 2s	• No more than two 2s
Grade 3	• All 4s or 5s with one 3	• Two 3s or higher	• All 1s and 2s with no more than one 3
Grade 1	• All 4s or 5s, with one 5 compensating for one 3 ^b	• One 3 or higher	• Only 1s and 2s
Kindergarten	• All 4s or 5s with a 3 in Writing ^c	• One 3	• All 1s or 2s with a 1 in any domain ^d

^a Students can listen, read and write, but sometimes have trouble speaking. If they can score a 4 or 5 in Writing and Reading, then they should be able to engage in class. Speaking should not hold an otherwise proficient student back. Higher scores are **not** allowed to compensate for lower scores.

^b Compensatory (a Level 3 compensates for a Level 5 five; no double compensation (two 3s cannot be compensated for by two 5s)). For Proficiency, there can be at most one 3.

^c The demands of writing in Kindergarten are less weighty than they are in the upper grades.

^d A deficiency this low warrants attention

Proficiency Determination Evaluations. Panelists completed a survey evaluating the Proficiency Determination activity (responses were obtained from 48 of the 51 panelists). The Proficiency Determination evaluation items reflect panelists' opinions with respect to validity including panelists' perceived understanding and fairness of the process and the defensibility of the cut scores. Table 24 provides the evaluation items and cross-grade results. These results indicate favorable endorsements (Agree [A] and Strongly Agree [SA] by approximately 98% of panelists, which supports the validity of the process. That is, panelists' endorsements indicate the vast majority of panelists believed the Proficiency Determination process was fair and that the resulting recommendations were defensible. The Proficiency Determination evaluation and results, disaggregated by grade, are provided in Appendix G7.

Table 24. Cross-Grade Proficiency Determination Evaluations and Results

Question	SD		D		A		SA		Total
	N	%	N	%	N	%	N	%	
I understood the process used for the Proficiency Determination	0	0%	0	0%	11	23%	37	77%	48
I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0%	0	0%	17	35%	31	65%	48
I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	1	2%	2	4%	16	33%	29	60%	48
I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0%	0	0%	11	23%	37	77%	48
I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	0%	2	4%	10	21%	36	75%	48
I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0%	0	0%	12	25%	36	75%	48
Total Responses	1	0%	4	1%	77	27%	206	72%	288

6. Adopting Cut Scores, Proficiency Determination Rules, and ALDs

6.1 Adopted Cut Scores

Multiple sources of data were available to the ELPA21 Consortium Council in consideration of the adoption of cut scores, including the recommendations resulting from the in-person standard setting workshop, described in Section 5.7. The adoption of cut scores can be viewed in four steps: First, cut scores for Level 3 and Level 4 were obtained for those grades examined during the in-person standard setting workshop (grades K, 1, 3, 5, 7, HS). Second, the Level 3 and Level 4 cut scores were for the grades not directly considered by the in-person standard setting panelists (grades 2, 4, 6, and 8). Third, cuts were set for Level 2 and Level 5 for all grades. Finally, cut scores were transformed in order to be expressed on the ELPA21 reporting scales. These steps are described below.

6.1.1. Cut Score Recommendations from the In-Person Workshop: Level 3 and Level 4 for Kindergarten, Grades 1, 3, 5, and 7, and High School

The ELPA21 Consortium Council considered two issues associated with the adoption of the cut scores directly set by panelists at the in-person workshop (Levels 3 and 4 for K, grades 1, 3, 5, and 7, and HS). First, the Council considered whether the CGS and in-person workshop cut scores should be used to compute cut scores independently of the panel judgements (i.e., by identifying the cut score placements that would maximize consistency of ELPA21 achievement level placements with the ratings of domain proficiency provided by educators in the CGS survey) and then combined with the workshop-derived cuts (e.g., as a weighted average of the two methods). The Consortium Council decided to utilize the CGS results as validity evidence but not as a weighted component of the adopted cut scores. The Council, in making this decision, cited (a) the significant training for and (b) the greater commitment of time to, the tasks the in-person panelists engaged in compared to the CGS participants, who may have had a less uniform interpretation of the language skills that should be represented in the ELPA21 achievement levels.

Second, the Consortium Council considered whether the cut scores should be smoothed for each domain to reduce variability across adjacent grades in the proportions of students falling into each achievement level. The Council, in considering this question, observed that unlike academic content areas in which some stability from grade to grade might be expected, the EL population has a changing population from grade to grade (with the highest performing students each year exiting the population and new students entering the population at all grade levels). As a consequence, the tested population is not necessarily comparable across grades. Moreover, Council members noted that current state ELP assessment results do not always show consistent patterns from grade to grade.

Thus, the Consortium Council voted to adopt the in-person panelist-recommended cut scores for Levels 3 and 4 without adjustments to improve the stability of the sizes of the achievement levels across grade levels.

6.1.2 Translating In-Person Workshop Cut Scores from the Interim Scale to the final ELPA21 Scale

As previously mentioned, the in-person standard setting workshop panelists worked with items that were located on an interim scale using parameter estimates from a preliminary calibration of the item pool. After the final item calibration was performed, the recommended Level 3 and Level 4 cut scores on the raw score point scale (i.e., the bookmark placements) were translated to the scale score metric for each grade band.

Based on the in-person workshop, each item appearing in the OIBs was associated with one of three levels, based on item's position and the bookmark locations: (a) below the Level 3, (B) within Level 3, or (c) above Level 3. After final item calibration, the RP67 values were updated based on the final item parameters (those used in operational scoring), and the cuts on the scale metric implied by the bookmark placements were identified. In some cases, the updated item parameters resulted in a reordering of items such that any cut score placement would result in at least one item falling within a different level from the one with which it had been previously associated (given the earlier item parameters and RP67 value). In such cases, cut scores were selected that minimized the misclassification of items to achievement levels (i.e., preserving as much as possible the panelists' judgements about the level into which each item fell).

A total 48 cuts were translated from the raw score (bookmark position) to the reporting scale: two levels (Level 3 and Level 4); six grade levels (K, 1, 3, 5, 7, HS), and four domains (Listening, Reading, Speaking, and Writing). Across these translations, the average percentage item items consistently classified (falling into the same level in the original and final calibrations) was 97%; an average of 1.5 items per OIB.

As a final step in establishing the Level 3 and Level 4 cut scores, CRESST examined the extent to which the cut scores from the in-person workshops (after translation to the reporting scale, as described above) represented increasing levels of rigor across adjacent grade levels. Although the ELPA21 tests do not utilize a vertical scale, item response theory (IRT) modeling was used to characterize the relationships between scales in the adjacent grade bands.

In the summative test forms for grade bands above grade K, a sample of items from the band below were administered. These items were chosen to span the content from their respective summative test blueprints and were embedded in randomly assigned, non-operational (i.e., unscored) blocks. A series of two-dimensional IRT models were fit to the combined response data including both the on- and off-grade band items. The two dimensions were defined by grade band of the items. In grade 1, one dimension measured grade 1 language skills, and the second dimension measured the grade K skills; grade 2-3, one dimension measured grade 2-3 language skills, and the second dimension measured the grade 1 skills. In these analyses, item parameters were fixed (based on the prior calibration), and the off-grade band mean, off-grade band variance, and the covariance between the grade bands were estimated (because the analyses were performed on the calibration sample, the on-grade band mean and variance were fixed).

Using the estimated structural parameters, the Level 3 and Level 4 cut scores were projected onto the adjacent grade band (Thissen, Liu, Magnus, & Quinn, 2015). This amounted to describing the score on the grade band below that would be expected given a score at the Level 4 and Level 5 in the grade band above. Of the 48 Level 3 and Level 4 cuts, 4 were adjusted to

achieve proper ordering. Adjustments were made by finding the midpoint between the projected cut from the grade band above and the grade band below.

The procedures above resulted in the establishment of the Level 3 and Level 4 cut scores for Kindergarten, grades 1, 3, 5, and 7, and high school. The Level 3 and Level 4 cut scores for the even grades and the Level 2 and Level 5 cut scores for all grades were subsequently established as described in the sections below.

6.1.3 Grades 2, 4, 6, and 8 Level 3 and Level 4 Cut Scores

The Level 3 and Level 4 cut scores for the grades not directly considered by the in-person standard setting panelists were estimated using item response theory (IRT) linking procedures. The IRT linking procedure described in the previous section was again used. The following method was used to establish the cut scores for the even grades—2, 4, 6, and 8. Each of these grade levels share the scale with the grade level that was considered by the panelists. The cuts were selected to represent midpoint between the corresponding cuts from the grade below and the grade above, projecting those cuts across the grade bands, where needed. For example, the grade 1 Level 4 cut score was projected onto the grade 2 scale (using linkages established by common items on the grade 1 and grade 2-3 ELPA21 assessments). The grade 3 Level 4 cut was already on the grade 2-3, so the grade 2 Level 4 cut score was obtained by averaging the projected Level 4 cut (from grade 1) and the Level 4 cut from grade 3. This method of interpolation produced the remaining candidate Level 3 and Level 4 cuts.

6.1.4 Level 2 and Level 5 Cut Scores.

Three approaches toward establishing the Level 2 and Level 5 cut scores were considered, as follows:

1. Use the CGS results for Level 2 and Level 5.
2. Use the relationship of cut scores across grades. For example, set the Level 3 cut score for grade g equal to the Level 2 cut score for grade $g + 1$. This can be interpreted as follows: A student at the Level 3 cut score in grade g who achieved zero growth for one year would be at the Level 2 cut score in grade $g + 1$. Similarly, set the Level 5 cut score for grade g equal to the Level 4 cut score For grade $g + 1$. This can be interpreted as follows: A student who achieved a scale score equal to the grade $g + 1$ Level 4 cut score one year early (in grade g) would be at the grade g Level 5 cut score.
3. Divide the interval from the bottom of the scale to the Level 3 cut score into two equal sized groups based on impact data. That is, set the Level 2 cut score such that an approximately equal percentage of students are classified in Level 1 and Level 2.

After reviewing cut scores and impact data based on these three approaches, the Consortium Council approved the third option, resulting in an approximately equal percentage of EL students in Levels 1 and 2. The Council, in making this decision, noted (a) the relatively large number of unarticulated cut scores—cut scores in a lower grade that were higher than the associated cut score in the next higher grade—resulting from the CGS results and (b) the extremely narrow interval associated with the Level 2 and Level 4 cut scores that frequently resulted from Option 2, above. Thus, the Level 2 (Level 5) cut scores were adopted to produce approximately equal sized percentages of students in Level 1 and Level 2 (Level 4

and Level 5). A few adjustments were required to ensure these cut scores were appropriately across grades within the grade band.

6.1.5 The ELPA21 reporting scale.

The reporting scale utilized by ELPA21 for domain scale scores uses a linear transformation of the IRT logit scale score, which are obtained by expected *a posteriori* (EAP) scoring. For all grade bands and domains, the logit scale scores are multiplied by 80 and added to 550. These values were chosen to provide a possible score range that would not fall below 100 or exceed 999. After applying the transformation, the scale scores are rounded to the nearest integer. (Standard errors are also transformed, by multiplying by 80 and rounding to the nearest integer.) Because test items were calibrated with an assumed standard normal distribution within grade band, the transformation constants can be interpreted as the mean and standard deviation with the domain and grade band on the reporting scale. The observed range in the 2015-2016 summative test data was about 200 to 800 across all grades and domains.

The same transformations applied to the scale scores were applied to the cut score scores onto the reporting scale. Tables 25 and 26 show the cut scores, by grade band, on both the IRT logit scale and ELPA21 reporting scale metrics.

Table 25. Adopted Cut Scores, Grade K, Grade 1, and Grade 2-3

Cut Scores for Grade K									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
K	Listening	-1.032	-0.539	0.790	1.186	467	507	613	645
K	Reading	-0.966	-0.447	0.519	0.964	473	514	592	627
K	Speaking	-0.786	-0.189	0.600	0.942	487	535	598	625
K	Writing	-0.666	0.147	1.259	1.535	497	562	651	673
Cut Scores for Grade 1									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
1	Listening	-1.436	-1.036	-0.009	0.548	435	467	549	594
1	Reading	-0.890	-0.436	0.421	0.989	479	515	584	629
1	Speaking	-0.281	0.337	0.533	0.864	528	577	593	619
1	Writing	-0.648	-0.019	0.784	1.135	498	548	613	641
Cut Scores for the 2–3 Grade Band									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
2	Listening	-1.779	-1.405	-0.475	0.179	408	438	512	564
2	Reading	-1.160	-0.767	0.060	0.559	457	489	555	595
2	Speaking	-0.751	-0.258	0.064	0.481	490	529	555	588
2	Writing	-1.229	-0.716	0.066	0.515	452	493	555	591
3	Listening	-1.763	-1.270	-0.176	0.602	409	448	536	598
3	Reading	-0.682	-0.107	0.756	1.173	495	541	610	644
3	Speaking	-0.631	-0.155	0.279	0.777	500	538	572	612
3	Writing	-0.644	-0.096	0.659	1.074	498	542	603	636

Table 26. Adopted Cut Scores, Grade 4-5, Grade 6-8, and Grade 9-12

Cut Scores for the 4–5 Grade Band									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
4	Listening	-1.898	-1.487	-0.726	0.162	398	431	492	563
4	Reading	-1.208	-0.781	-0.006	0.546	453	488	550	594
4	Speaking	-1.097	-0.552	-0.077	0.431	462	506	544	584
4	Writing	-1.413	-0.864	0.230	0.628	437	481	568	600
5	Listening	-1.711	-1.190	-0.649	0.386	413	455	498	581
5	Reading	-1.024	-0.489	0.471	0.966	468	511	588	627
5	Speaking	-0.840	-0.296	0.291	0.707	483	526	573	607
5	Writing	-1.400	-0.802	0.602	0.978	438	486	598	628
Cut Scores for the 6–8 Grade Band									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
6	Listening	-1.744	-1.373	-0.648	0.182	410	440	498	565
6	Reading	-1.115	-0.673	0.191	0.674	461	496	565	604
6	Speaking	-1.063	-0.485	0.151	0.566	465	511	562	595
6	Writing	-1.563	-0.969	0.173	0.554	425	472	564	594
7	Listening	-1.502	-0.963	0.033	0.593	430	473	553	597
7	Reading	-0.798	-0.194	0.734	1.152	486	534	609	642
7	Speaking	-0.938	-0.285	0.399	0.760	475	527	582	611
7	Writing	-0.956	-0.369	0.590	0.939	474	520	597	625
8	Listening	-1.472	-0.896	0.190	0.783	432	478	565	613
8	Reading	-0.704	-0.038	1.122	1.489	494	547	640	669
8	Speaking	-0.925	-0.279	0.495	0.863	476	528	590	619
8	Writing	-0.828	-0.216	0.867	1.209	484	533	619	647
Cut Scores for HS									
Grade	Domain	IRT Logit Scale				ELPA21 Reporting Scale			
		L2	L3	L4	L5	L2	L3	L4	L5
HS	Listening	-1.232	-0.732	0.268	0.787	451	491	571	613
HS	Reading	-0.774	-0.141	1.012	1.405	488	539	631	662
HS	Speaking	-0.868	-0.171	0.539	0.858	481	536	593	619
HS	Writing	-0.807	-0.211	0.815	1.133	485	533	615	641

6.1.5 Impact data

Table 27 shows the distribution of students by grade and achievement level, given the final cut scores presented in the previous section. Scoring of paper-pencil and Blind/Low Vision test forms was not complete at the time these were generated. Thus, the impact data presented for consideration of the proposed cut scores was limited to online tests. Consortium Council members received these consortium-wide tables, as well as versions specific to their state.

Table 27. Impact Data Associated with Adopted Cut Scores

Grade	Total N	M	SD	Level 1		Level 2		Level 3		Level 4		Level 5	
				N	P	N	P	N	P	N	P	N	P
Listening Domain													
K	44501	-.016	.913	6518	.147	6522	.147	22537	.506	4454	.100	4470	.100
1	45318	-.027	.852	3003	.066	3003	.066	15232	.336	12036	.266	12044	.266
2	44929	-.226	.829	1888	.042	1885	.042	12172	.271	14505	.323	14479	.322
3	41106	.223	.870	1226	.030	1226	.030	9319	.227	14666	.357	14669	.357
4	36252	-.125	.858	1253	.035	1253	.035	5312	.147	14209	.392	14225	.392
5	30272	.108	.902	1298	.043	1298	.043	2733	.090	12466	.412	12477	.412
6	24374	-.118	.830	962	.040	960	.039	3889	.160	9273	.380	9290	.381
7	20440	-.030	.918	1596	.078	1599	.078	6746	.330	5250	.257	5249	.257
8	18269	.102	.966	1454	.080	1451	.079	6020	.330	4673	.256	4671	.256
HS	54019	-.033	.933	6195	.115	6192	.115	20626	.382	10508	.195	10498	.194
Reading Domain													
K	44501	-.014	.905	6987	.157	6982	.157	17555	.395	6491	.146	6486	.146
1	45318	-.069	.923	8355	.184	8366	.185	15535	.343	6522	.144	6540	.144
2	44929	-.297	.855	7020	.156	7013	.156	15610	.347	7634	.170	7652	.170
3	41106	.249	.921	6827	.166	6821	.166	14496	.353	6483	.158	6479	.158
4	36252	-.168	.881	4451	.123	4455	.123	11436	.316	7949	.219	7961	.220
5	30272	.115	.935	3783	.125	3780	.125	11122	.367	5795	.191	5792	.191
6	24374	-.177	.846	3485	.143	3474	.143	9419	.386	4001	.164	3995	.164
7	20440	-.040	.932	4446	.218	4445	.218	7180	.351	2186	.107	2183	.107
8	18269	.147	.978	3792	.208	3785	.207	7590	.416	1550	.085	1552	.085
HS	54019	-.037	.938	12790	.237	12790	.237	20536	.380	3956	.073	3947	.073

Speaking Domain													
K	44501	-.013	.936	8099	.182	8093	.182	16624	.374	5839	.131	5846	.131
1	45318	-.026	.891	14600	.322	14609	.322	4358	.096	5867	.130	5884	.130
2	44929	-.215	.859	9881	.220	9883	.220	7672	.171	8749	.195	8744	.195
3	41106	.198	.898	5884	.143	5872	.143	8255	.201	10553	.257	10542	.257
4	36252	-.093	.896	4156	.115	4163	.115	7410	.204	10268	.283	10255	.283
5	30272	.076	.921	3895	.129	3887	.128	8410	.278	7043	.233	7037	.233
6	24374	-.038	.872	2773	.114	2767	.114	7402	.304	5707	.234	5725	.235
7	20440	-.035	.943	3127	.153	3126	.153	7027	.344	3583	.175	3577	.175
8	18269	.055	.968	2552	.140	2553	.140	6612	.362	3274	.179	3278	.179
HS	54019	-.003	.936	9200	.170	9200	.170	19249	.356	8181	.151	8189	.152
Writing Domain													
K	44501	-.059	.921	13082	.294	13079	.294	14262	.321	2039	.046	2039	.046
1	45318	-.084	.938	11246	.248	11255	.248	15035	.332	3893	.086	3889	.086
2	44929	-.298	.888	6768	.151	6771	.151	15179	.338	8108	.181	8103	.180
3	41106	.248	.920	6380	.155	6381	.155	14021	.341	7164	.174	7160	.174
4	36252	-.157	.897	3406	.094	3404	.094	16578	.457	6426	.177	6438	.178
5	30272	.121	.934	2157	.071	2159	.071	16376	.541	4786	.158	4794	.158
6	24374	-.138	.860	1720	.071	1718	.071	11790	.484	4572	.188	4574	.188
7	20440	-.032	.943	2933	.144	2932	.143	9529	.466	2516	.123	2530	.124
8	18269	.128	.983	2692	.147	2696	.148	8952	.490	1964	.108	1965	.108
HS	54019	-.026	.943	9907	.183	9911	.184	24875	.461	4657	.086	4669	.086

6.2 Adopted Proficiency Determination Rules

A systematic process was used to support the establishment of appropriate Proficiency Determination rules. The process included the following:

1. CGS Results: The CGS results were reviewed to identify the most common profiles that emerged from teacher classifications of students into achievement levels for reading, writing, listening, speaking, and aggregate student performance across all four domains.
2. EL Expert Advisory Panel Review: EL Experts reviewed and discussed the CGS profiles and made recommendations for Proficiency Determination rules.
3. In-Person Standard Setting Workshop: Panelists reviewed the CGS and EL Expert Panel results and recommended Proficiency Determination rules.
4. ELPA21 Discussion and Adoption of Proficiency Determination Rules: The ELPA21 Consortium Council reviewed the various sources of information, discussed various options, and adopted rules that associated specific profiles with the three Proficiency Determination levels.

We describe the process and results of each of these steps, in the following summary.

6.2.1 CGS Results Supporting the Proficiency Determination

The CGS was reported in detail in Section 3. Table 6 in Section 3 provided the seventy-five most common profiles for all grades and Appendix B4 provides a summary of all profiles associated with Level 4 and provides the 25 most common profiles by grade. The results are summarized as follows.

- The most common CGS profiles associated with an overall Proficiency Determination of Proficient are defined by the following 3 rules.
 - (Writing is greater than or equal to three) and (Reading, Listening, and Speaking are greater than or equal to four)
 - (Reading and Writing are greater than or equal to three) and (Listening and Speaking are greater than or equal to four)
 - Reading, Writing, Listening and Speaking are all greater than or equal to four
- If students were assigned a 1 or a 2 in any domain, then they are not Proficient.
- Two common profiles conflicted at most grades.
 - Some teachers thought the profile 3344 (Reading, Writing, Listening, Speaking) should be associated with an overall Proficiency Determination of three while others thought it should be associated with four (using the five-level Proficiency rating used in the CGS); similarly, some teachers thought the profile 2233 (Reading, Writing, Listening, Speaking) should be associated with an overall Proficiency rating of two while others thought it should be associated with a rating three.

These results were shared with the EL Expert Advisory Panel for their review and recommendations.

6.2.2 EL Expert Advisory Panel Review Supporting the Proficiency Determination

The CGS results were presented to the EL Expert Advisory Panel, as described in Section 4. The slides presented to the experts are provided in Appendix C1. The EL Experts were emailed a survey that provided a summary of the profiles that emerged from the CGS and asked the experts to endorse agreement or disagreement with the rules that emerged from the CGS. The survey is provided in Appendix C2, and the results are described in Section 4.

The results are summarized as follows:

- An overall Proficiency Determination of Proficient is defined by the following rules:
 - (Listening and Speaking are greater than or equal to Level 3) and (Reading and Writing are greater than or equal to Level 4)
 - Reading, Writing, Listening and Speaking are all greater than or equal to Level 4
 - Compensatory models should not be implemented. That is, a Level 3 in one domain cannot be compensated with a Level 5 in another domain to meet the requirements to be Proficient.
- An overall Proficiency Determination of Progressing (Overall Level 3 using the CGS five-level overall classification scheme) is defined by the following rules:
 - Reading, Writing, Listening, and Speaking are all greater than or equal to Level 3, but the profile is not sufficient to be classified as Level 4 or Level 5).
 - One more Level 2s *can* be compensated by an equal number of domains at Level 4 or Level 5.
 - There can be no domains at Level 1.

The EL Expert recommendations, and the CGS profiles, were presented to the in-person workshop panelists.

6.2.3 In-Person Standard Setting Workshop Activities Supporting the Proficiency Determination

A summary of the in-person standard setting workshop Proficiency Determination activity presentation and discussion is summarized in Section 5.10 Proficiency Determination Review and Recommendations and the resulting recommendations are provided in Table 23.

A review of the results in Table 23 indicate that somewhat different profiles across grades emerged naturally, but the profiles do have commonalities. For example, five of the six grade bands argued that to be classified Proficient, students should be at Level 4 or Level 5 in at least three domains and at least Level 3 in the fourth domain; however, the domain that the Level 3 could be in varied across the grades. The exception was grade band 4–5 which required all domains to be Level 4 or Level 5 to be Proficient.

To be classified as Emerging, most grade bands focused on profiles dominated by Level 1 and Level 2 with some grade bands including one Level 3.

6.2.4 ELPA21 Profile Determination Adoption

The ELPA21 Consortium Council considered each of the above cited sources of information. A set of questions was posed to the ELPA21 Consortium Council, TMT Leads, and EL Experts to prompt discussion with respect to whether the grade specific or common Proficiency Determination profiles should be adopted. The responses were nearly unanimous in recommending a common Proficiency Determination across grades. The following points summarize the responses to the questions and provide the rationale for the adoption of a common set of profiles across grade bands.

- a. Most EL Expert and state representative comments simply asserted that the expectation is a pattern of domain profiles associated with Proficiency Determination that is the same across grade bands.
- b. Inconsistency would send the wrong message: The domain cut scores were predicated on policy definitions (descriptors) that reflect the proficiency demands of the academic curriculum at the given grade. It would be difficult to explain why a student must have Level 4 or Level 5 in all domains to be proficient in one grade band but could be Proficient with a Level 3 in another grade band. Inconsistency would imply a departure from the policy definitions, which sends a confusing message.
- c. The following statements about consistency were discussed:
 - i. The ELP standards were written to be consistent across the grade bands.
 - ii. All grade bands are testing students with a level of rigor on par with their respective native English speakers.
 - iii. ELs are learning content and language (with ESL support) from one grade level to the next and expectations should flow seamlessly from one grade band to the next.
 - iv. If the tests are consistent and reliable and the grade level expectations were constructed consistently then consistent Proficiency Determination rules would be expected across the grade bands.
 - v. If the cut scores were set consistently for each grade band then it makes sense to have the Proficiency Determination rules consistent as well.

The primary rationale supporting different rules for different grade bands was the fact that the in-person standard setting panelists arrived at different profile rules across the grade bands. However, as can be observed in Table 23, there were commonalities across the grade bands and these commonalities are reflected in the adopted Proficiency Determination rules.

ELPA21 adopted common Proficiency Determination profile rules for all grade bands, as follows:

Proficient: Students are Proficient when they attain a level of English language skill necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21

by attaining a profile of Level 4 or higher in all domains. Once Proficient on ELPA21, students can be considered for reclassification.

Progressing: Students are Progressing when, with support, they approach a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile with one or more domain scores above Level 2 that does not meet the requirements to be Proficient. Students scoring Progressing on ELPA21 are eligible for ongoing program support.

Emerging: Students are Emerging when they have not yet attained a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Levels 1 and 2 in all four domains. Students scoring Emerging on ELPA21 are eligible for ongoing program support.

The decision was arrived at after consideration of the CGS, the EL Expert recommendations, the in-person workshop panelists' recommendations, as summarized in Table 23, and discussion. Impact data resulting from the application of the adopted Proficiency Determination rules and cut scores are provided for online tests across all consortium states in Table 28.

Table 28. Impact Data Associated with Adopted Proficiency Determination Rules

Grade	Total N	Emerging		Progressing		Proficient	
		N	P	N	P	N	P
K	44501	7743	.174	34275	.770	2483	.056
1	45318	4612	.102	35518	.784	5188	.115
2	44929	3286	.073	31137	.693	10506	.234
3	41106	2275	.055	28339	.689	10492	.255
4	36252	2337	.065	22909	.632	11006	.304
5	30272	2321	.077	20406	.674	7545	.249
6	24374	1723	.071	16783	.689	5868	.241
7	20440	2830	.139	14532	.711	3078	.151
8	18269	2571	.141	13404	.734	2294	.126
HS	54019	10816	.200	37762	.699	5441	.101

Educator ratings of students' overall English proficiency obtained in the Contrasting Groups Study (CGS) were matched with the ELPA21 2015-2016 summative test results. Results for the 4,524 matched students in Table 29 (grades K-5) and Table 30 (grades 6-12) below. The tables show the cross-classifications according to CGS ratings and ELPA21 Proficiency Level. For the purpose of evaluating agreement, CGS overall ratings of 1 and 2 were collapsed, as were ratings of 4 and 5. Within each grade level, the polychoric correlation (r), percent of close agreement (within one level), percent exact agreement, and a weighted kappa (κ) coefficient (quadratic weights) are reported.

Table 29. Comparison of CGS Ratings and ELPA21 Proficiency Level, Grades K-5

Grade	N	ELPA21 Level	CGS Overall Proficiency Rating						r	close	exact	κ
			L1	L2	L3	L4	L5	Total				
KG	542	Emerging	.098	.063	.015	.002	.000	.177	.678	.998	.489	.317
		Progressing	.054	.323	.295	.094	.006	.771				
		Proficient	.000	.002	.017	.031	.002	.052				
		Total	.151	.387	.327	.127	.007					
1	576	Emerging	.023	.028	.012	.003	.002	.068	.561	.995	.462	.209
		Progressing	.033	.264	.309	.151	.023	.780				
		Proficient	.000	.010	.040	.068	.035	.153				
		Total	.056	.302	.361	.222	.059					
2	531	Emerging	.030	.028	.011	.002	.000	.072	.520	.998	.559	.289
		Progressing	.023	.179	.333	.119	.019	.672				
		Proficient	.000	.006	.083	.119	.049	.256				
		Total	.053	.213	.427	.239	.068					
3	482	Emerging	.041	.019	.004	.000	.000	.064	.553	1.000	.556	.313
		Progressing	.012	.187	.311	.127	.019	.656				
		Proficient	.000	.006	.089	.135	.050	.280				
		Total	.054	.212	.405	.261	.068					
4	452	Emerging	.035	.013	.002	.000	.000	.051	.633	1.000	.558	.277
		Progressing	.007	.146	.288	.139	.027	.606				
		Proficient	.000	.009	.113	.159	.062	.343				
		Total	.042	.168	.403	.299	.088					
5	313	Emerging	.048	.038	.006	.000	.000	.093	.681	1.000	.575	.281
		Progressing	.013	.147	.300	.185	.038	.684				
		Proficient	.000	.000	.035	.121	.067	.224				
		Total	.061	.185	.342	.307	.105					

The polychoric correlations between the collapsed CGS ratings were strong, ranging from 0.443 to 0.820. In addition, nearly all students (0.988 or greater in each grade level) had CGS ratings within one step of their ELPA21 Proficiency determination, while exact agreement ranged from 0.453 to 0.618. The weighted kappa coefficient describes the level of agreement

above and beyond what would be expected due to chance (i.e., if the two measures of English language proficiency had no association but the proportions in the Total rows and columns remained the same). The kappa coefficients range from 0.097 to 0.411, indicating that the levels of agreement were consistently higher than what would be expected due to chance alone. There is some variability in the results, with grades K and 6 showing very strong agreement and grade 12 showing weaker agreement. However, the results suggest general agreement between the ELPA21 Proficiency determination and the evaluations of local educators.

Table 30. Comparison of CGS Ratings and ELPA21 Proficiency Level, Grades 6-12

Grade	N	ELPA21 Level	CGS Overall Proficiency Rating						r	close	exact	kappa
			L1	L2	L3	L4	L5	Total				
6	344	Emerging	.137	.061	.003	.000	.000	.201	.820	1.000	.608	.411
		Progressing	.015	.145	.273	.163	.015	.610				
		Proficient	.003	.009	.041	.087	.049	.189				
		Total	.154	.215	.317	.250	.064					
7	425	Emerging	.087	.078	.038	.002	.000	.205	.706	.998	.480	.223
		Progressing	.045	.101	.247	.242	.075	.711				
		Proficient	.007	.000	.009	.033	.035	.085				
		Total	.139	.179	.294	.278	.111					
8	324	Emerging	.102	.093	.031	.006	.006	.238	.497	.988	.478	.207
		Progressing	.019	.096	.241	.235	.123	.713				
		Proficient	.000	.000	.006	.028	.015	.049				
		Total	.120	.188	.278	.269	.145					
9	178	Emerging	.213	.157	.051	.006	.006	.433	.706	.989	.618	.311
		Progressing	.011	.056	.219	.163	.084	.534				
		Proficient	.000	.000	.006	.017	.011	.034				
		Total	.225	.213	.275	.185	.101					
10	186	Emerging	.086	.108	.081	.000	.000	.274	.742	1.000	.527	.249
		Progressing	.011	.059	.263	.263	.059	.656				
		Proficient	.000	.000	.000	.038	.032	.070				
		Total	.097	.167	.344	.301	.091					
11	96	Emerging	.073	.083	.094	.010	.000	.260	.528	.990	.510	.247
		Progressing	.000	.063	.271	.302	.021	.656				
		Proficient	.000	.000	.000	.010	.073	.083				
		Total	.073	.146	.365	.323	.094					
12	75	Emerging	.000	.040	.053	.000	.000	.093	.443	1.000	.453	.095
		Progressing	.000	.013	.307	.320	.147	.787				
		Proficient	.000	.000	.013	.053	.053	.120				
		Total	.000	.053	.373	.373	.200					

6.3 Finalizing Achievement Level Descriptors

After cut scores were adopted by ELPA21, EL Experts drafted Achievement Level Descriptors (ALDs) that, like the Target Student Descriptors, reflect the KSPs of the student just entering the achievement level. The difference between ALDs and Target Student Descriptors is that ALDs are based on student performance on test items while Target Student Descriptors are expectations of how students should perform. In other words, the ALDs reflect the KSPs students *can* do while Target Student Descriptors reflect the KSPs students *should* be able to do.

To create the ALDs, Dr. Karla Egan trained the standard setting panelists to synthesize the information for the two main questions on the item map. Recall that each group answered two questions for each item: “What does this item measure?” And, “Why is this item more difficult than the preceding items?” Following training, panelists created a column that synthesized item information for use in ALD writing.

On August 4, 2016, Dr. Karla Egan led the EL Experts through training on how to write ALDs. Because the final cut scores had not yet been determined, the EL Experts’ first task was to finish synthesizing information from the questions on the item map asking about what the item measures and its difficulty. To do this, the EL Experts relied on the annotated item maps and OIBs from the standard setting.

After ELPA21 adopted final cut scores, Pacific Metrics’ Director of Content Services, Michael Baker, synthesized the EL Experts’ item-level descriptors to create draft ALDs. When there were few, or no items in achievement level descriptor ranges, language from the achievement level indicators (ALIs) that had been established for the levels was used to support the descriptor writing. ELPA21’s Item Development TMT used these draft ALDs to develop final ALDs.

7. Validity

Validity evidence for the ELPA21 Standard Setting is provided in multiple ways. First, the ELPA21 standard setting should adhere to the standards established by appropriate professional organizations, in this case the AERA/APA/NCME Standards for Educational and Psychological Testing (2014). Second, the standard setting should be consistent with the recommendations for best practices in the literature. Third, the standard setting should be supported by established validity criteria; in this case Kane's (2001) four criteria are considered—the conceptual coherence of the standard setting process, procedural evidence for the descriptive and policy assumptions, internal consistency evidence, and agreement with external criteria. We describe each of these as follows.

7.1 Adherence to Professional Standards 1

The ELPA21 Standard Setting was designed to adhere to the following AERA/APA/NCME standards related to standard setting:

Standard 5.21: When proposed score interpretation involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.

Pacific Metrics provided a detailed design document and this technical report summarizing the rationale and procedures used for the ELPA21 Standard Setting. The detailed design document was reviewed by ELPA21 and their TAC and was approved by ELPA21 prior to the standard setting. Critical elements of the detailed design document are incorporated into this technical report, which documents the design, implementation, and results of the standard setting process.

Standard 5.22: When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way.

The CGS and the Bookmark Standard Setting Procedure were designed to directly leverage the knowledge and experience of the respective participants. The CGS participants were EL educators working with specific rosters of EL students, and their task was aligned with their experience. They already had direct knowledge of their students' skills, and online training provided them with an understanding of the ELPA21 policy definitions for the new achievement levels. Thus, after training, the CGS participants were well prepared to classify their students into the ELPA21 achievement levels in each domain.

The Bookmark Procedure has been successful, according to the claims of the developers (Lewis, Mitzel, Mercado & Schulz, 2012) largely because it directly leverages the subject matter expertise of the panelists in the judgment task. That is, panelists recommend cut scores by identifying the skills associated with test items and indicate which test items students should master to qualify for each achievement level. In this case, the ELPA21 panelists were selected based on criteria such that each can be assumed to have experience with EL students and a good understanding of the ELP standards and ELPA21 assessments. Additional and extensive training on the latter—the ELPA21 assessments—was provided at the in-person standard setting workshop through study of the Bookmark Procedure's ordered item booklets.

Thus, the methods supporting the ELPA21 standard setting were designed so that the participants providing the judgments were able to bring their knowledge and experience to bear in a reasonable way.

Standard 5.23: When feasible and appropriate, cut scores defining categories and distinct substantive interpretations should be informed by sound empirical data concerning the relation of test performance to the relevant criteria.

The CGS and Bookmark Procedure both made sound use of empirical data concerning the relation of test performance to various relevant criteria. The CGS is fundamentally based on the relationship between teachers' classifications of students and their associated ELPA21 test performance, as described in Section 3. The Bookmark Procedure incorporates the use of empirical data and its relationship to student ELPA21 performance in multiple ways. First, the order of difficulty of items presented in the ordered item booklets is based on empirical data—specifically student test performance on ELPA21—and its relationship to panelists' expectations for each achievement level.

Second, the relationship of panelists' judgments to empirical student performance data is provided as feedback in the form of impact data in Round 2, as described in Section 5 and Figure 9. The vast majority of panelists indicated in their evaluations that they considered the impact data when they made their judgments. The impact data was based on ELPA21 results available at that time. Census data produced results that were not substantially different than the data reviewed by panelists.

Finally, the policy decisions made by the ELPA21 Consortium Council were based on strong consideration of the empirical data associated with the CGS and in-person standard setting workshop. Thus, the appropriate use of empirical data and its relationship to the various criteria was incorporated into multiple components of the ELPA21 standard setting and cut score adoption processes.

7.2 Best Practices

The ELPA21 in-person standard setting workshop was designed to be consistent with established standard setting practices and principles. This section overviews best practices in terms of the selection of panelists, method, and implementation (Hambleton & Pitoniak, 2006; Hambleton, Pitoniak, & Copella, 2012; Kane, 2001; Mehrens, 1995).

Panels. Panelists should be recruited such that the resulting panels are representative of important demographic groups, and are knowledgeable about English language proficiency and English Learners. The ELPA21 in-person standard setting consisted of eight to ten panelists per grade band, with each ELPA21 state represented in each grade band. Section 5.6.3 provides details about the panel demographics and recruitment.

Method. The standard setting method should be appropriate for the type of test administered. The Bookmark Standard Setting Procedure (BSSP) is appropriate for tests of this type—with multiple item formats and scaled using IRT.

Implementation. The following aspects must be considered when evaluating the implementation of a standard setting:

- (a) the orientation and training of panelists,

- (b) the use of achievement level indicators (ALIs)/Target Student Descriptors, the ELP Standards, and the Policy Definitions,
- (c) the support of panelists' understanding of students' test-taking experiences,
- (d) the use of an iterative process for panelists to provide and revise their recommendations based on feedback and discussion, and
- (e) the timing and use of student impact data and other forms of panelist feedback.

Each of these standard setting elements as they pertain to the Bookmark Procedure were directly addressed and are described in Section 5. That is,

- (f) panelists' evaluations provided evidence of a successful orientation to ELPA21 and standard setting in general and specific training on the Bookmark Procedure,
- (g) panelists' evaluations provided evidence that they used and considered the ALIs, Target Student Descriptors, ELP Standards, and Policy definitions when making their cut score recommendations,
- (h) panelists were provided the opportunity to take a form of the online ELPA21 assessments to understand the students' test-taking experience,
- (i) the Bookmark Procedure incorporates multiple, iterative rounds of ratings in which panelists modify their judgments based on feedback and discussion, and
- (j) the timing of the use of impact data was well-considered, discussed with ELPA21 and incorporated in Round 2 of the process prior to the final round of ratings.

Thus, the in-person ELPA21 Bookmark Standard Setting was designed and implemented to support best practices.

7.3 Kane's Standard Setting Validity Criteria

Kane's (2001) four standard setting validity criteria are considered and described as follows:

7.3.1 Conceptual Coherence

Kane describes conceptual coherence first, in terms of adhering to the AERA/APA/NCME Standards as described in Section 7.1. Further, the systematic design of the ELPA21 standard setting process provides evidence of a coherent set of activities that together support the validity of the process. That is, the systematic, and linear implementation of the multiple standard setting components (i.e., the CGS, the EL Expert Advisory Panel, and the in-person Bookmark Standard Setting workshop) suggests a coherent and synergistic standard setting process. For example, the CGS provided proficiency profiles that were reviewed and advanced by the EL Experts. In turn, the EL Expert recommendations were incorporated into the in-person Bookmark workshop to support panelist discussion resulting in recommendations to the ELPA21 Consortium Council that supported the adoption of well-considered Proficiency Determination rules. The use of multiple sources of information and multiple sources of input, such as that described above, supports the conceptual coherence of the ELPA21 standard setting process.

7.3.2 Procedural Evidence

Procedural validity is supported by evidence that (a) the standard setting followed a defensible methodology and was conducted as described in the standard setting design document developed by Pacific Metrics and approved by ELPA21 and (b) panelists are knowledgeable about the content area and student population with respect to the assessed skills, actively and appropriately participate in the process, and exhibit behavior that indicates an appropriate understanding of the tasks required of them.

The following observations support the procedural validity of the July 19–22, 2016 ELPA21 cut score activities and recommended cut scores.

- The Bookmark Procedure is a well-established method that has been used on numerous occasions to set cut scores for state summative and English Language Development assessments.
- The workshop was conducted in an organized manner and, with few exceptions, followed the schedule of events outlined in the standard setting agendas provided in this document (See auditor’s report in Appendix H).
- The workshop was well staffed (See auditor’s report in Appendix H). Sufficient numbers of Pacific Metrics’ staff, supported by the ELPA21 staff and the ELPA21 appointed EL Expert, were available to support the standard setting activities.
- The high level of EL specific discussion observed among panelists reflected thoughtful, knowledgeable, and well-qualified standard setting panelists.
- There was ample observational evidence that participants were actively using the concept of the “Target Student” to make their judgments. Understanding and using the concept of the “Target Student” is important to the validity of the method used. Panelists’ evaluations indicated they tended to have a consistent understanding of the Target Student and used that understanding when making their cut score recommendations.
- Participants affirmed their understanding of the tasks required of them through “readiness surveys.” This, at a minimum, supports participants’ self-efficacy with respect to their various tasks.
- Participants appeared to be highly engaged in the process and interacted appropriately with their colleagues.
- The auditor’s report, provided in Appendix H, supports the validity of the in-person standard setting workshop results.

7.3.3 Internal Consistency

Internal consistency evidence can be derived from an analysis of the data generated during the standard setting. Overall, data resulting from the ELPA21 standard setting appear reasonable and consistent with what would be expected from a consensus-building process of this type. An analysis of the data, as observed in Table 18, indicate a generally decreasing range of bookmark placements from round to round, which indicates that participants’ judgments tended toward convergence over the course of the standard setting. While convergence is not necessary, it is a reflection of (a) an increasingly common understanding

of the Target Student and (b) participants' willingness to modify their perspectives based on group discussion and in consideration of their colleagues' perspectives. This reflects participants' understanding that they are not just setting cut scores for their own students, but for the entire consortium, and thus, their fellow participants' perspectives are informative and important. This evidence supports the internal validity of the standard setting results.

7.3.4 External Criteria

External validity is supported by triangulation of the recommended cut scores with data external to the standard setting activity, when it is available. In this case, a review of the recommended cut scores and impact data in light of Consortium Council member-state data from their previous ELD assessments indicated that, for the most part, ELPA21 results were not inconsistent with what was expected.

7.4 Auditor's Report

The auditor's report provides strong, objective evidence of the validity of the in-person Bookmark Standard Setting Workshop. As stated in the auditor's report in Appendix H, "With the assumption that the participants' evaluations do not suggest cause for concern, the available information and observations suggest that cut score recommendations produced at the workshop can be considered to be valid and reliable estimates of the cut scores for the ELPA21. Policy makers should have confidence that the recommendations from the standard setting activity were based on sound procedures, and produced trustworthy, valid, and defensible results."

7.5 Evaluations

The strong support for the validity of the in-person Bookmark standard setting workshop, as noted in Section 7.4 Auditor's Report was dependent on the support of the workshop's evaluation results. As reported in Table 12, Table 14, Table 19, Table 20,

Table 21, Table 22, and Table 24, all readiness survey and evaluation results support the validity of the results:

- Positive responses to the Opening Session Readiness Survey were endorsed by approximately 98% of panelists.
- Positive responses to the Bookmark Training Readiness Survey were endorsed by 99.5% of panelists.
- Positive responses to the Reading Evaluation were endorsed by 98% of panelists.
- Positive responses to the Speaking Evaluation were endorsed by 96% of panelists.
- Positive responses to the Writing Evaluation were endorsed by 98% of panelists.
- Positive responses to the Listening Evaluation were endorsed by 99% of panelists.
- Positive responses to the Proficiency Determination Evaluation were endorsed by 99% of panelists.

7.6 Conclusions

The above cited evidence, taken together, strongly supports the validity of the adopted cut scores, Proficiency Determination rules, and Achievement Level Descriptors. Implementation of the TAC recommendation to review the cut scores following subsequent ELPA21 administrations, after the collection of additional validity evidence and with the support of longitudinal data, will further enhance the validity of the cut scores, Proficiency Determination rules, and ALDs, and their value with respect to supporting the growth of English Learners.

8. References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- CTB/McGraw-Hill (2013). *Washington English Language Proficiency Assessment (WELPA): Annual Assessment Technical Report 2011–2012 School Year*. CTB/McGraw-Hill, author.
- CTB/McGraw-Hill (2006). *California English Language Development Test (CELDT) Bookmark Standard Setting Technical Report*. CTB/McGraw-Hill, author.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Westport, CT: American Council on Education/Praeger.
- Hambleton, R. K., Pitoniak, M. J., & Copella, J. (2012). Setting performance standards. In G. J. Cizek (Ed.), *Setting performance standards: Theory and applications* (2nd ed., pp. 47–76). New York, NY: Routledge.
- Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53–88). Mahwah, NJ: Lawrence Erlbaum.
- Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998, April). *The Bookmark standard setting procedure: Methodology and recent implementations*. Paper presented at the annual meeting of the National Council for Measurement in Education, San Diego, CA.
- Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (pp. 225–254). New York, NY: Routledge.
- Mehrens, W. A. (1995). Methodological issues in standard setting for educational exams. In *Proceedings of joint conference on standard setting for large-scale assessments* (pp. 221–263). Washington, DC: National Assessment Governing Board and National Center for Educational Statistics.
- Mitzel, H. C., Lewis, D. M., 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 Erlbaum.
- Thissen, D., Liu, Y., Magnus, B., & Quinn, H. (2015). Extending the use of multidimensional IRT calibration as projection: Many-to-one linking and linear computation of projected scores. In *Quantitative Psychology Research* (pp. 1-16). Springer International Publishing.

9. Appendices

Appendices

Contents

Appendix A. Target Student Descriptors	2
Appendix B. Materials and Results from the Contrasting Groups Study	13
Appendix B1. Screenshots from the Online Survey	14
Appendix B2. Screenshots of the Excel Workbook	24
Appendix B3. Contrasting Groups Training Materials	26
Appendix B4. CGS: Profiles	35
Appendix B5. CGS Results: Comparison of CGS and State Data	41
Appendix C. EL Expert Advisory Panel Materials and Results	45
Appendix C1. Slides from May 11th EL Expert Meeting	46
Appendix C2. EL Expert Panel Survey	58
Appendix D. In-Person Bookmark Standard Setting Workshop Training	64
Appendix D1. Table Facilitator Training Slides	65
Appendix D2. Opening Session Slides	85
Appendix D3. Bookmark Training Slides	121
Appendix D4. Introduction to Round 3 and Impact Data Slides	132
Appendix D5. Proficiency Determination Training Slides	136
Appendix E. In-Person Bookmark Standard Setting Workshop Materials	154
Appendix E1. Table Facilitator Agenda	155
Appendix E2. Bookmark Rating Form	170
Appendix E3. Bookmarks	171
Appendix E4. Sample Item Map	173
Appendix E5. Materials Summary	174
Appendix F. In-Person Bookmark Standard Setting Workshop Results	176
Appendix F1. Round by Round Results	177
Appendix F2. Calculating a Meaningful Standard Error for the Bookmark Cut Score	249
Appendix G. Evaluations	251
Appendix G1. Opening Session Readiness Survey	252
Appendix G2. Bookmark Training Readiness Survey	254
Appendix G3. Reading Evaluation and Results	255
Appendix G4. Listening Evaluation and Results	269
Appendix G5. Speaking Evaluation and Results	284
Appendix G6. Writing Evaluation and Results	307
Appendix G7. Proficiency Determination Evaluation and Results	325
Appendix H. Auditor's Report	334

Appendix A.

Target StudentDescriptors

Listening Level 3 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(1) identify key words and phrases in read-alouds and oral presentations	(1) identify the main topics in oral presentations of literary and informational texts	(1) identify key words, phrases and the main topic or message in oral presentations and read-alouds	(1) determine the main idea or theme; identify key details from oral presentations	(1) identify the main topic in oral communication	(1) identify the main topic in oral communication and some details about the topic
(2) participate in short conversations and respond to simple yes/no questions and wh-questions about familiar topics	(2) participate in short conversations and discussions on familiar topics and texts and answer simple questions	(2) participate in short conversations and discussions and respond to simple yes/no and wh-questions about familiar topics; respond to the comments of others	(2) participate in short conversations and discussions about familiar topics and answer questions	(2) participate in short conversations on familiar topics and respond to simple questions and wh-questions	(2) participate in conversations on familiar topics and respond to questions and comments, and restate the main idea
(5) gather information from an oral source to answer a question with a few details	(5) gather and summarize information from oral sources	(5) gather and record some information from oral	(5) gather information from oral sources to answer a question and identify key information	(5) gather information from a few oral sources and record some data and information	(5) gather information from oral sources and summarize data and information
(6) Identify one main point the author is making	(6) identify a reason a speaker gives to support the main point	(6) identify one or two reasons a speaker gives to support a point	(6) identify how one or two reasons support or do not support the specific points a speaker makes	(6) identify the main argument a speaker makes and one reason a speaker gives to support the argument	(6) identify the main argument a speaker makes and several reasons a speaker gives to support the argument
(8) recognize the meaning of frequently occurring words and phrases in oral presentations about familiar topics, experiences, or events	(8) determine the meaning of some less frequently occurring words and phrases in simple oral presentations about familiar topics, experiences, or events	(8) determine the meaning of some frequently occurring words, phrases and expressions in simple oral presentations about familiar topics, experiences, or events	(8) determine the meaning of some general academic and content-specific words, phrases, and idiomatic expressions about familiar topics, experiences, or events in oral presentations	(8) determine the meaning of frequently occurring words, phrases, and expressions in oral presentations topics, experiences, or events	(8) determine the meaning of frequently occurring and some content-specific words, phrases, and expressions in oral presentations about topics, experiences, or events

Listening Level 4 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(1) identify the main topics and answer questions about key details in read-alouds and oral presentations	(1) identify the main topics and answer questions about key details in oral presentations of literary and informational texts	(1) determine the main topic, idea or message; identify or answer questions about some key details that support the main idea/message in oral presentations	(1) determine the main idea or theme from oral presentations presentation and some details that support it	(1) determine a central idea or theme in an oral presentation and some details that support it	(1) determine the central idea or theme in an oral presentation and some details that support it
(2) participate in longer conversations and discussions and respond to questions about familiar and unfamiliar topics	(2) participate in conversations and discussions about familiar and unfamiliar topics and texts, answer questions and respond to others' comments	(2) participate in conversations and discussions, and ask and answer questions to clarify understanding about familiar and unfamiliar topics and texts	(2) participate in conversations and discussions on familiar and unfamiliar of topics, and answer questions	(2) participate in discussions on familiar and unfamiliar topics and texts; answer questions	(2) participate in discussions on familiar and unfamiliar topics, texts, and issues; paraphrasing key ideas expressed, and answer questions
(5) use information from oral sources to answer a question using increasing details/events	(5) gather, summarize, and answer questions about information from oral sources	(5) gather and record information from oral presentations	(5) gather and record information from oral sources	(5) gather information from oral sources and paraphrase key information	(5) gather information from multiple oral sources and paraphrase key information
(6) Identify a reason a speaker gives to support a point	(6) identify reasons a speaker gives to support the main point	(6) identify how one or two reasons support the main point a speaker makes	(6) identify how reasons support the specific points a speaker makes or fails to make	(6) determine the speaker's argument and distinguish claims that are supported by reasons and evidence from those that are not	(6) determine the speaker's argument and distinguish claims that are supported by reasons and evidence from those that are not
(8) answer and sometimes ask questions to help determine the meaning of words and phrases in oral presentations & read-alouds about familiar, and some unfamiliar, topics, experiences, or event	(8) determine the meaning of some less frequently occurring words, phrases, and idiomatic expressions in oral presentations about familiar and unfamiliar of topics, experiences, or events	(8) determine the meaning of less-frequently occurring words and phrases, content-specific words, and some idiomatic expressions, and some general academic and content-specific vocabulary in oral presentations and read-alouds about familiar and unfamiliar topics, experiences, or events	(8) determine the meaning of general academic and content-specific words and phrases, and idiomatic expressions about familiar and unfamiliar topics, experiences, or events in oral presentations	(8) determine the meaning of general academic and content-specific words, phrases, and idiomatic expressions in oral presentations about familiar or unfamiliar topics, experiences, or events	(8) determine the meaning of general academic and content-specific words, phrases, and idiomatic expressions in oral presentations about familiar and unfamiliar topics, experiences, or events

Reading Level 3 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(1) identify some key words and phrases from read-alouds of literary and informational text	(1) identify main topics, answer simple questions about key details from read-alouds and simple texts	(1) identify some key words and phrases and the main topic or message in simple literary and informational texts	(1) determine the main idea or theme and identify a few key details in simple written texts	(1) identify the main topic and a few key details in simple written texts	(1) identify the main topic and key details in written text, provide basic summary
(2) respond to simple yes/no and wh- questions about familiar topics	(2) answer simple questions about familiar topics	(2) respond to simple yes/no and wh- questions about familiar topics	(2) respond to others' comments and answer questions about familiar topics and texts and some unfamiliar simple non-fiction and fiction text	(2) respond to simple questions and wh-questions on some familiar and some unfamiliar topics and texts	(2) respond to a variety of questions on familiar and some unfamiliar topics and texts
(5) gather information from a provided source to answer a question			(5) gather some information from provided sources and identify key information	(5) gather information from provided sources and record some data and information	(5) gather information from provided sources and summarize data and information
	(6) identify a reason an author gives to support the main point	(6) identify one or two reasons an author gives to support the main point	(6) identify one or two reasons that support the specific points an author makes or fails to make	(6) identify the main argument an author makes and identify one reason an author gives to support the argument	(6) identify the main argument an author makes and list the reasons the author gives to support the argument
(8) recognize the meaning of some frequently occurring words and phrases in read-alouds about familiar experiences, or events	(8) answer simple questions to help determine the meaning of some less frequently occurring words and phrases in simple texts and read-alongs about familiar topics, experiences, or events	(8) determine the meaning of frequently occurring words, phrases, and expressions in simple literary and informational texts about familiar topics, experiences, or events	(8) determine the meaning of some frequently occurring words, phrases, and expressions in written text about familiar topics, experiences, or events	(8) determine meaning of frequently occurring words, phrases, and expressions in written text about familiar and some unfamiliar topics, experiences, or events	(8) determine meaning of frequently occurring words, some content-specific words, phrases, and expressions in written text about familiar and some unfamiliar topics, experiences, or events

Reading Level 4 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(1) identify main topics and answer questions about key details from read-alouds of literary and informational text; retell some events from read-alouds of literary and informational text and picture books	(1) identify main topics and ask and answer questions about an increasing number of key details in written texts and read-alouds; retell familiar stories or episodes of stories in written texts and read-alouds	(1) identify the main topic or message and answer questions about key details in literary and informational texts	(1) determine main idea or theme, explain how key details support the main idea or theme in written texts, and summarize part of a text	(1) determine the central idea or theme and identify how the theme is supported by specific details in written texts; summarize part of a text	(1) determine the central idea or theme and identify how the theme is developed by specific details in a written text; summarize a text
(2) answer questions about a variety of topics	(2) answer questions about a variety of topics and texts	(2) answer questions about a variety topics and texts	(2) answer questions about a variety of topics and texts	(2) respond to others' comments and answer questions on a variety of topics and texts	(2) respond to others' comments and answer questions on topics, texts, and issues; add information and evidence and restate key ideas
(5) use information from provided sources and some basic gathered sources to answer a question			(5) gather information from multiple sources and identify key information	(5) gather information from multiple sources and summarize or paraphrase	(5) gather information from multiple sources and begin to determine reliability of each source
	(6) identify reasons an author gives to support the main point	(6) identify and tell how one or two reasons support the main point an author makes	(6) identify how reasons support the specific points an author makes or fails to make	(6) explain the argument an author makes; distinguish between claims that are supported by reasons and evidence from those that are not	(6) explain the argument an author gives to support a claim and identify textual evidence to support the explanation; distinguish between claims that are supported by reasons and evidence from those that are not
(8) determine the meaning of frequently occurring words and phrases in read-alouds about familiar, and some unfamiliar, topics, experiences, or events	(8) determine the meaning of less frequently occurring words, phrases, and some idiomatic expressions in written texts about familiar and unfamiliar topics, experiences, or events	(8) determine the meaning of some less-frequently occurring words and phrases, content-specific words, and some idiomatic expressions in written texts about familiar and unfamiliar topics, experiences, or events	(8) determine the meaning of general academic and content specific words, phrases, and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events	(8) determine the meaning of general academic and content-specific words and phrases and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events	(8) determine the meaning of general academic and content-specific words and phrases, and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events

SPEAKING Level 3 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(2) participate in short conversations; respond to simple yes/no and wh- questions about familiar topics	(2) participate in short conversations; ask and answer simple questions about familiar topics	(2) participate in short conversations and discussions; respond to simple yes/no and wh- questions-about familiar topics; take turns; ask some questions	(2) participate in short conversations and discussions, ask and answer questions, respond to the comments of others, and contribute his or her own comments about familiar topics	(2) participate in short conversations about familiar and some unfamiliar topics and texts, and respond to simple questions and some wh-questions	(2) participate in conversations about familiar and some unfamiliar topics and texts, present information and ideas, respond to questions and comments, and restate main idea
(3) communicate information or feelings about familiar topics, experiences, events, or objects in the environment	(3) deliver short simple oral presentations about familiar topics, experiences, or events	(3) deliver short oral presentations about familiar topics, experiences, or events	(3) deliver short oral presentations about familiar texts, topics, or events, including a few details	(3) deliver short oral presentations about familiar and some unfamiliar texts, topics, experiences, or events	(3) deliver oral presentations about familiar and some unfamiliar texts, topics, experiences, or events including one or two examples
(4) express an opinion and preference about a familiar topic	(4) express an opinion about a familiar topic, experiences, or events and give a reason for the opinion	(4) express an opinion about a familiar topic or story; give one reason for the opinion	(4) construct a claim about a familiar topic; introduce the topic and provide several supporting reasons or facts in a logical order	(4) construct a claim about a familiar or unfamiliar topic and give a reason to support the claim	(4) construct a claim about a familiar or unfamiliar topic; introduce the topic and give several reasons to support the claim; provide a concluding statement
(5) recall information from experience or use information from a provided source to answer a question	(5) summarize information and answer simple questions from provided sources		(5) recount a short sequence of events with events in order; introduce an informal topic and present one or two facts about the topic; use a range of temporal and other linking words, and provide a concluding statement		
				(6) identify the main argument an author or speaker makes and identify one reason an author or speaker gives to support the argument	(6) identify the main argument an author or speaker makes and several reasons given to support the argument
(9) retell several events from experience, attempting to use one or more frequently occurring linking words	(9) retell a simple sequence of events and present simple information; use frequently occurring linking words	(9) present some information about a topic and recount some events in a simple sequence of		(9) recount a brief sequence of events in order; introduce an informational topic, presenting one or two facts about the topic;	(9) recount a sequence of events; introduce an informational topic and provide several facts about the topic; use common linking

SPEAKING Level 3 Target Students

		events; use frequently occurring linking words		use some commonly occurring linking words, and provide a concluding statement	words to connect events and ideas and provide a basic conclusion
(10) respond to questions using one or more longer sentences; use frequently occurring nouns and use verbs in present tense	(10) produce simple sentences, use some singular and plural nouns, frequently occurring present and past tense verbs, prepositions and conjunctions	(10) produce simple sentences; use some frequently occurring collective nouns, verbs, adjectives, adverbs, and conjunctions	(10) produce and expand simple and compound sentences and use some relative pronouns and adverbs, and prepositional phrases	(10) produce and expand simple and compound sentences and use nouns, pronouns, verbs, prepositions, adjectives, adverbs, conjunctions, and prepositional phrases	(10) produce and expand simple and compound sentences use frequently occurring verbs, nouns, adjectives, adverbs, prepositions and conjunctions

SPEAKING Level 4 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(2) participate in short conversations and discussions; respond to questions about familiar and unfamiliar topics	(2) participate in discussions and conversations; ask and answer questions; respond to comments of others; make comments of his/her own	(2) participate in discussions and conversations; ask and answer questions about a variety of topics, building on the comments and ideas of other and contributing his/her own comments and ideas.	(2) participate in conversations and discussions about a variety of topics, building on the ideas of others and expressing his or her own ideas; asking and answering relevant questions, and adding relevant information and evidence	(2) participate in conversations and discussions on a variety of topics and texts; build on the ideas of others and contribute his or her own ideas; ask and answer questions to clarify ideas; and summarize key ideas	(2) participate in conversations and discussions on a variety of topics, texts, and issues, building on the ideas of others and contributing his or her own ideas; support points with specific evidence; ask and answer questions to clarify ideas; and summarize key ideas
(3) communicate coherent messages about familiar topics, experiences, or events	(3) deliver short oral presentations about a variety of texts, topics, experiences, or events	(3) deliver short oral presentations about a variety of topics, experiences, or events	(3) deliver short oral presentations about a variety of texts, topics, and experiences with some details	(3) deliver short oral presentations about a variety of texts, topics, and experiences, with some specific details	(3) deliver oral presentations about a variety of texts, topics, or events, with some details, concepts, examples, and information
(4) express an opinion and preference about familiar and some unfamiliar topics or stories	(4) express opinions about a variety of texts and topics, experiences, or events; give reasons for the opinions	(4) for a variety of topics, introduce a topic, express an opinion about a topic or story opinions, and give one or more reasons for the opinions	(4) construct a claim about a familiar or unfamiliar topics, introduce the topic, provide several reasons or facts to support the claim, and provide a concluding statement	(4) construct a claim about a familiar or unfamiliar topic; introduce the topic, provide several supporting reasons or facts to support the claim, and provide a concluding statement	(4) construct a claim about a familiar or unfamiliar topic; introduce the topic, provide logically ordered reasons or facts to support the claim, and provide a concluding statement
(5) recall information from experiences or use information from provided sources to answer a question and give an explanation for their answer	(5) summarize information and answer questions from provided sources		(5) recount a detailed sequence of events with a beginning, middle and end, and introduce an informational topic with facts and details; provide a conclusion; use transitional words and phrases to connect events, ideas, and opinions		
				(6) explain the argument and specific claims an author or speaker makes, and distinguish between claims that are supported by reasons and evidence from those that are not	(6) analyze the reasoning and use of rhetoric in persuasive texts and cite textual evidence to support the analysis

SPEAKING Level 4 Target Students

(9) retell several events from experience, or a familiar story in sequence, using several frequently occurring linking words	(9) recount two or three events in sequence and present information about a topic; use some temporal words and frequently occurring linking words	(9) introduce facts or information about a topic and recount a sequence of events; use common temporal and linking words to connect ideas and/or events		(9) recount a short sequence of events with a beginning, middle, and end; introduce and develop an informational topic with a few facts and details; use transitional words and phrases to connect events and ideas; provide a conclusion	(9) recount a detailed sequence of events or steps with a clear structure; introduce and develop an informational topic with facts, details, and evidence; use complex transitions to clarify
(10) use and respond to question words; produce simple and some compound sentences; use frequently occurring regular plural nouns, regular verbs in past and present tense, prepositions, question words, and make an attempt at irregular verbs and irregular plural nouns	(10) produce and expand simple and some compound sentences; use singular and plural nouns, present and past verb tense with appropriate subject-verb agreement, and frequently occurring prepositions and conjunctions	(10) produce and expand simple and some compound sentences and a few complex sentences; use some collective nouns, adjectives, adverbs, and conjunctions, and the past tense of some frequently occurring irregular verbs, and frequently occurring adjectives, adverbs, and conjunctions	(10) produce and expand simple, compound and a few complex sentences and use relative pronouns and adverbs, subordinating conjunctions, and prepositional phrases	(10) produce and expand simple, compound, and a few complex sentences and use relative pronouns, conjunctions, and prepositional phrases	(10) produce and expand simple, compound and complex sentences and use relative pronouns and adverbs, subordinating conjunctions, and prepositional phrases

Writing Level 3 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(2) respond to simple yes/no questions and wh-questions about familiar topics	(2) participate in short written exchanges; answer simple questions about familiar topics	(2) participate in short written exchanges and respond to simple yes/no questions and wh-questions about familiar topics	(2) participate in short written exchanges about familiar topics and texts; respond to others' comments and add some comments of his or her own, and ask and answer questions about the topics or texts, adding relevant information and evidence	(2) participate in short written exchanges on familiar and unfamiliar topics and texts, presenting information and ideas; respond to simple questions and wh-questions	(2) participate in written exchanges on familiar topics and texts, presenting information and ideas, respond to general questions, and restate main idea
(3) communicate information or feelings about familiar topics, experiences or events	(3) compose short written texts about familiar topics, stories, experiences, or events using simple sentences	(3) compose short written texts about familiar topics, experiences, or events	(3) compose narratives or informational texts about familiar topics, texts, or experiences, and develop the topic with a few details	(3) compose narratives or informational texts about familiar and unfamiliar texts, topics and experiences	(3) compose written narratives or informational texts about familiar texts, topics, experiences or events
(4) express an opinion or preference about a familiar topic	(4) express an opinion about a familiar topic or story and give a reason for the opinion	(4) express an opinion about a topic or story	(4) construct a simple claim about familiar topics, introducing the topic and providing a few reasons or facts to support the claim	(4) construct a claim about familiar and unfamiliar topics and give a reason to support the claim	(4) construct a claim about familiar topics, introduce the topic, give reasons to support the claim, and provide a concluding statement
	(9) retell a simple sequence of events and present simple information using frequently occurring linking words	(9) communicate simple information about a topic and recount two events in sequence; use some frequently occurring linking words	(9) recount a sequence of events in order; introduce an informational topic including one or two facts about the topic; use common transitional words and phrases to connect events, ideas and opinions	(9) recount a brief sequence of events in order, introduce an informational topic, using some commonly occurring linking words and providing a concluding statement	(9) recount a sequence of events in order, introduce an informational topic, provide several facts about the topic and use common linking words to connect events and ideas
(10) produce a few simple sentences and use frequently occurring nouns, verbs, and short phrases	(10) produce and expand simple sentences; use some singular and plural nouns, verbs in the present tense, and frequently occurring prepositions and conjunctions	(10) produce simple sentences; use some frequently occurring collective nouns, verbs, adjectives, adverbs, and conjunctions	(10) produce and expand simple and compound sentences and use some relative pronouns and adverbs, and prepositional phrases	(10) produce simple and compound sentences and use nouns, pronouns, verbs, prepositions, adjectives, adverbs, conjunctions and prepositional phrases	(10) use frequently occurring verbs, nouns, adjectives, prepositions, and conjunctions; produce simple and compound sentences

Writing Level 4 Target Students

Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7	High School
(2) respond to yes/no questions and wh- questions about familiar and some unfamiliar topics;	(2) participate in written exchanges on familiar and unfamiliar texts and topics; ask and answer questions and respond to comments of others, adding some comments of his/her own	(2) participate in written exchanges on familiar and unfamiliar texts and topics, and texts; respond to; build on the ideas of others and contribute his or her own comments and ideas	(2) participate in written exchanges on familiar and unfamiliar topics and texts, building on the ideas of others, expressing his or her own ideas, asking and answering relevant questions, and adding information and evidence	(2) participate in written exchanges on familiar and unfamiliar topics and texts, building on the ideas of others, expressing his or her own ideas, answering questions, and adding information and evidence	(2) participate in written exchanges on familiar and unfamiliar topics, texts, and issues, building on the ideas of others, expressing his or her own ideas, answering questions, adding information, and restating most of the key ideas expressed
(3) write grade-appropriate text about familiar and some unfamiliar topics, experiences or events	(3) compose written texts about a variety of texts, topics, experiences, or events using simple or compound sentences	(3) compose written narratives or informational texts about a variety of topics, experiences, or events	(3) compose narratives or informational texts about a variety of texts, topics, and experiences, including some details	(3) compose narrative or informational texts with some details about a variety of texts, topics and experiences, and develop the topic with some details	(3) compose narrative or informational texts about variety of texts, topics, or events, and develop the topic with some details and examples
(4) express an opinion or preference about familiar and unfamiliar topics or stories	(4) express opinions about familiar and unfamiliar texts and topics and give a reason for the opinion	(4) introduce a topic, express an opinion about a familiar and unfamiliar topic or story, and give several reasons for the opinion	(4) construct a claim about a familiar or unfamiliar topic, introducing the topic, providing supporting reasons or facts to support the claim, and providing a concluding statement	(4) construct a claim about a familiar and unfamiliar topic, introduce the topic and provide supporting reasons and facts in logical order; provide a concluding statement	(4) construct a claim about a familiar and unfamiliar topic, introduce the topic and provide sufficient reasons or facts to support the claim; provide a concluding statement
	(9) recount two or three events in sequence and present information about a topic using some temporal words and frequently occurring linking words	(9) introduce and present a few pieces of facts about an informational topic and recount a short sequence of events; use common temporal and linking words to connect ideas and/or events	(9) introduce and present a sequence of events, with a beginning, middle and end; and introduce an informational topic including facts and details; provide a conclusion; using a variety of transitional words and phrases to connect events, ideas, and opinions	(9) introduce and present a sequence of events with a beginning, middle, and end; introduce and develop an informational topic with facts and details, using a variety of transitional words and phrases to connect events and ideas; provide a conclusion based on evidence	(9) introduce and present a sequence of events, with a beginning, middle, and end; introduce and develop an informational topic with facts and details and use a variety of transitional words and phrases to connect events, ideas, and opinions; provide a conclusion based on evidence.
(10) produce simple sentences and some compound sentences, and use frequently occurring regular plural nouns, verbs, and prepositions	(10) produce and expand simple and some compound sentences; use an increasing number of singular and plural nouns, present, past and future verb tenses with appropriate subject-verb agreement, and frequently occurring prepositions and conjunctions	(10) produce and expand simple and some compound sentences and a few complex sentences; use some collective nouns, some the past tense of frequently occurring past tense irregular verbs, and frequently occurring adjectives, adverbs, and conjunctions	(10) produce and expand simple, compound and a few complex sentences and use relative pronouns and adverbs, subordinating conjunctions, and prepositional phrases		(10) use phrases and clauses to produce and expand simple, compound, and complex sentences

Appendix B. Materials and Results from the Contrasting Groups Study

Appendix B1. Screenshots from the Online Survey

Contrasting Groups – ELPA21 Study

Thank you for choosing to participate in the Contrasting Groups Study for the ELPA21 assessments. The Contrasting Groups Study is a standard setting process that allows input from large groups of ELL educators from multiple states. By participating in this study, you are helping the Consortium decide what students should know and be able to do in order to be placed into ELPA21 achievement levels and to succeed academically without EL program support. Your input will also help us to understand how well the ELPA21 test itself is measuring students' ELL proficiency.

In this study, we ask you to please classify your students into ELL achievement levels based on the policy definitions shown below. It is important that you base your classifications solely on the policy definitions and your own understanding of the students' ELL skill. Your understanding may be based on classroom interactions, homework, and classroom assessments. You should not base your classifications on how well you think each student will perform, or did perform on the ELPA21 test. ELPA21 recognizes that you have deep and valuable knowledge of your ELL students' English language skills and of what your ELL students need to know and be able to do in order to succeed academically without ELL support. We wish to bring this expertise to bear on the standard setting process independently of the ELPA21 test.

You do not need to complete the survey in a single sitting. You may exit and return to the survey at a later time in order to complete it. Please be sure to save your progress before exiting.

Materials

In order to participate in this study, you will want to have ready access to the following materials:

1. Your school ID.
2. Your students' first and last names, with their grade level and **state student ID**.
 - If you teach in Iowa, Ohio, Oregon, or Washington, you may find your state student ID in the TIDE system.
 - If you teach in Arkansas, Kansas, Nebraska, or West Virginia, you may find your state student ID in the Nextera system.
3. The Policy Descriptors and Assessment Targets. These may be found on a later screen. We encourage you to print out the Policy descriptors for reference by choosing the **File** menu at the top of the screen and then selecting **Print** from the drop-down menu.

Please tell us a bit about yourself. The information in this survey is confidential. Your personally identifiable information will not be shared outside of this research project.

What is your first name? *

What is your last name? *

What is the name of your school district? *

What is your school's name? *

What is your school ID? *

Please print and study the *Policy Descriptors* and *Assessment Targets*. The **Policy Descriptors** provide you with high-level guidance on the expectations for students in each performance level. The *Assessment Targets* describe what it means to be proficient in each domain as well as to have overall proficiency in English.

To print, choose the File menu at the top of the screen and then select Print from the drop-down menu.

Policy Descriptors	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Begins to develop a few grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs. Students scoring at this level have demonstrated proficiency sufficient to succeed academically without EL program support.
Level 5: Advanced	Demonstrates superior English language skills, as measured by ELPA21.

Assessment Targets	
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Overall	ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks without on-going EL Program support. Once proficient on ELPA21, ELs can be considered for reclassification.

[pipe:17], you can choose two different ways to fill out the information in this survey: 1) You can enter your students' information through this survey, or 2) You can fill out an Excel template, which was provided to you via email. Once you have entered your students' information in the Excel file, you will be directed to upload the file on this site.

[pipe:17], how would you prefer to take this survey? *

☐ Online survey ☐ Upload Excel form

You received the Excel survey in your email with the survey link.

Please upload the Excel file with your students' information.

No file chosen

Please identify the students you will be rating for the Contrasting Groups Study in the roster below. In order to participate in this study, you will want to have ready access to the following materials:

1. Your students' first and last names with their grade level and **state student ID**.
 - If you teach in Iowa, Ohio, Oregon, or Washington, you may find your state student ID in the TIDE system.
 - If you teach in Arkansas, Kansas, Nebraska, or West Virginia, you may find your state student ID in the Nextera system.
2. The Policy descriptors.
3. The Assessment Targets.

Please enter information for the students enrolled in your EL Program. Remember to save your progress before proceeding to the next page.

	First Name	Last Name	Student ID	Grade
1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

For your students, consider the Policy descriptors and provide ratings for their Overall performance as well as their performance in each domain: Reading, Writing, Listening, and Speaking.

- For a domain, rate each student's performance within only that area.
- For Overall performance, rate each student's global performance in English.
- Please be sure to provide ratings for overall performance as well as for all four domains.

It is expected that your ratings of student performance may vary across the domains. For example, you may rate a student at Level 4 overall, Level 2 on writing, Level 5 on speaking, Level 3 on Listening, and Level 2 on Reading. It is not expected that you assign the same level to all areas. However, it is expected that if you rate the student at Level 4 or higher overall, you believe that the student's profile of proficiency on the individual domains is *sufficient to allow them to succeed academically without English language program support*.

Reading Assessment Target: An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities

Reading					
	Level 1	Level 2	Level 3	Level 4	Level 5
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Speaking Assessment Target: An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

[illegible]

Listening Assessment Target: An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

Listening					
	Level 1	Level 2	Level 3	Level 4	Level 5
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Writing Assessment Target: An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

Writing					
	Level 1	Level 2	Level 3	Level 4	Level 5
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Overall Proficiency: ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level, content-related academic tasks without on-going EL program support. Once proficient on ELPA21, ELs can be considered for reclassification.

Overall					
	Level 1	Level 2	Level 3	Level 4	Level 5
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[pipe:17], you said that 5 is not the right number of levels. How many levels would you recommend? *

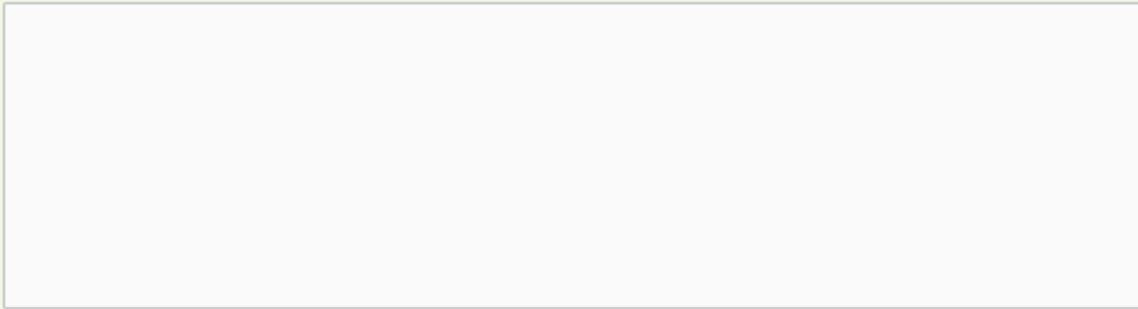
- ☐ 2 levels
 ☐ 3 levels
 ☐ 4 levels
 ☐ 6 levels

Please tell us if you think five levels are sufficient and necessary for categorizing the range of proficiency in your ELs.

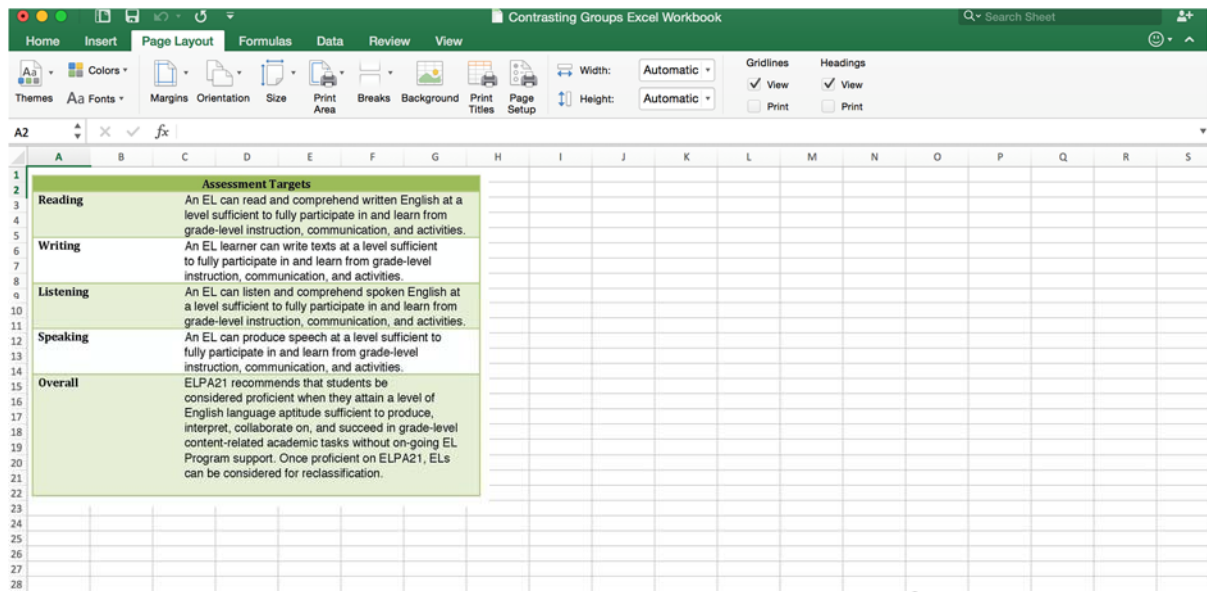
I believe that five is the right number of levels. *

☐ Yes ☐ No

(Optional) Do you have any comments or suggestions to share with ELPA21?

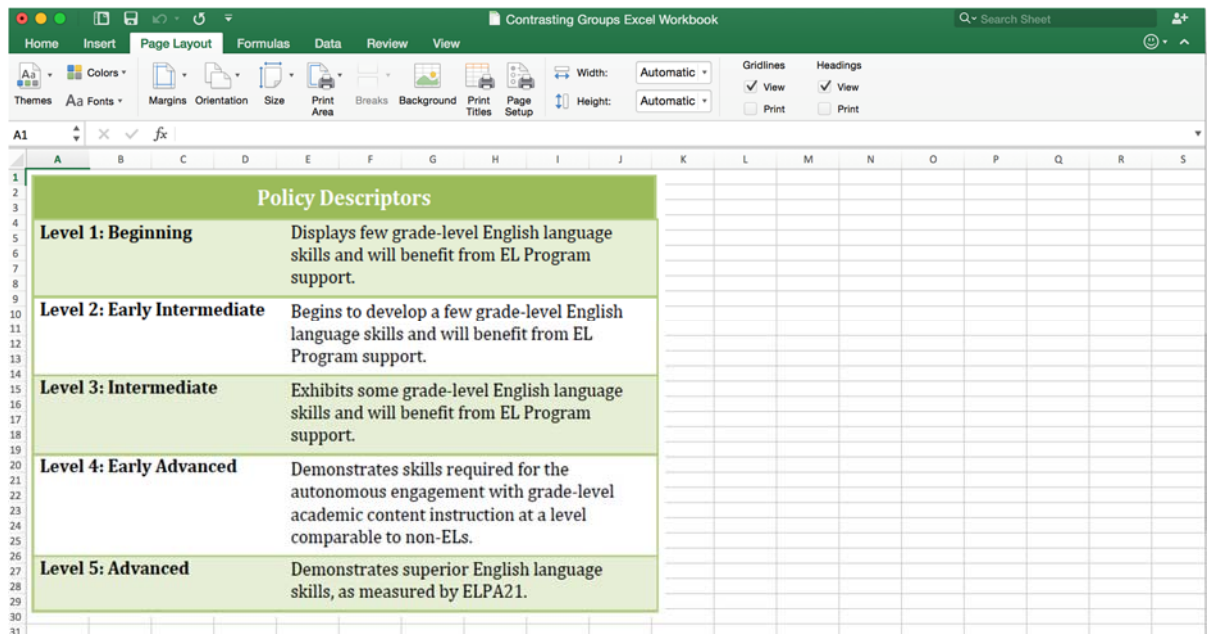


Appendix B2. Screenshots of the Excel Workbook



The screenshot shows the 'Assessment Targets' table in the 'Contrasting Groups Excel Workbook'. The table is located in the range A2:G15. It has two columns: the first column lists the skill areas (Reading, Writing, Listening, Speaking, Overall) and the second column provides detailed descriptions of the targets for each skill area.

Assessment Targets	
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Overall	ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks without on-going EL Program support. Once proficient on ELPA21, ELs can be considered for reclassification.



The screenshot shows the 'Policy Descriptors' table in the 'Contrasting Groups Excel Workbook'. The table is located in the range A1:G30. It has two columns: the first column lists the levels (Level 1: Beginning, Level 2: Early Intermediate, Level 3: Intermediate, Level 4: Early Advanced, Level 5: Advanced) and the second column provides detailed descriptions of the descriptors for each level.

Policy Descriptors	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Begins to develop a few grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Demonstrates superior English language skills, as measured by ELPA21.

Contrasting Groups Excel Workbook

Home Insert Page Layout Formulas Data Review View

Themes Colors Margins Orientation Size Print Area Breaks Background Print Titles Page Setup Width: Automatic Height: Automatic Gridlines View View Print Print

B6 fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	About You																
2	First Name																
3	Last Name																
4	School District																
5	School																
6	School ID																
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	

Contrasting Groups Excel Workbook

Home Insert Page Layout Formulas Data Review View

Themes Colors Margins Orientation Size Print Area Breaks Background Print Titles Page Setup Width: Automatic Height: Automatic Gridlines View View Print Print

AD8 fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Student Information				Reading					Writing					Listening					Speaking					Overall						
2	First Name	Last Name	State ID	Grade	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5		
3																															
4																															
5																															
6																															
7																															
8																															
9																															
10																															
11																															
12																															
13																															
14																															
15																															
16																															
17																															
18																															
19																															

Appendix B3. Contrasting Groups Training Materials

Contrasting Groups Training



Overview



Standard Setting



Materials



Process





What is Standard Setting?

- Process allows educators to make decisions about the knowledge and skills that students demonstrate when acquiring English-language proficiency.



Why Standard Setting?

- Process provides educators a means to discuss knowledge and skills
- Outcome aids test interpretation for stakeholders
- Create connection between the assessment and content standards





How do we do this?

You provide:

- Knowledge of student performance
- Knowledge of English-language proficiency

We provide:

- Policy level descriptors
- Assessment targets
- Process for setting cut scores



Policy Descriptors

- High-level definition of type of performance expected at each proficiency level
- Five levels of proficiency





Policy Descriptors

Policy Descriptors	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Begins to develop a few grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs. Students scoring at this level have demonstrated proficiency sufficient to succeed academically without EL program support.
Level 5: Advanced	Demonstrates superior English language skills, as measured by ELPA21.



ELPA21
English Language Proficiency
Assessment for the 21st Century



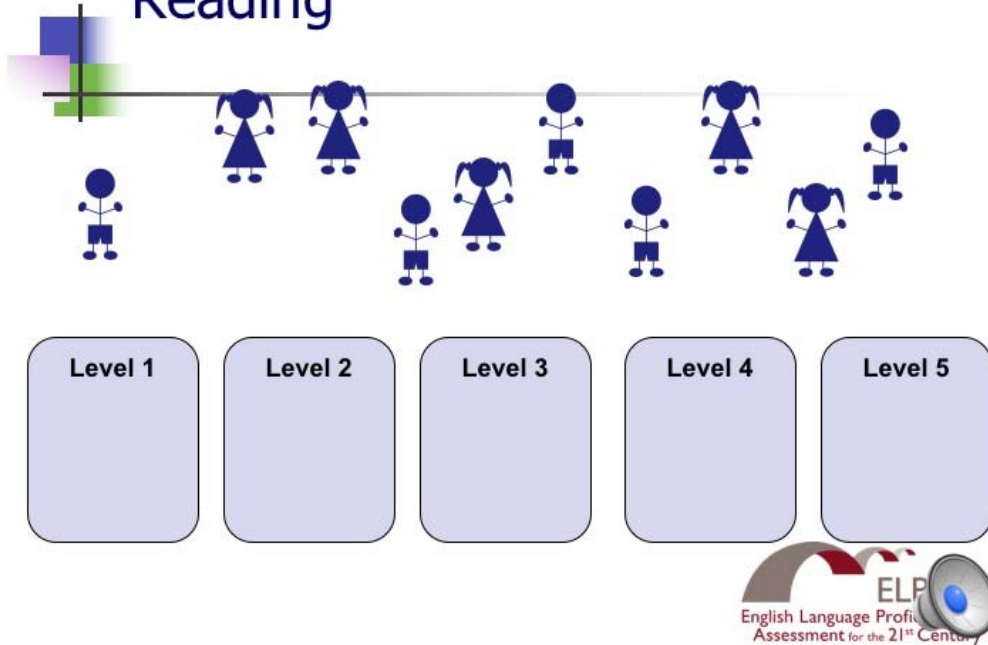
Assessment Targets

Assessment Targets	
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Overall	ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks without on-going EL Program support. Once proficient on ELPA21, ELs can be considered for reclassification.



ELPA21
English Language Proficiency
Assessment for the 21st Century

Reading



Contrasting Groups Ratings

- Reading
- Writing
- Listening
- Speaking
- Overall

Assessment Targets	
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Overall	ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks without on-going EL Program support. Once proficient on ELPA21, ELs can be considered for reclassification.



Proficiency Levels & Cut Scores

- Experts make ratings
- Ratings translated to cut scores
- Cut score: test score necessary to achieve each level of proficiency

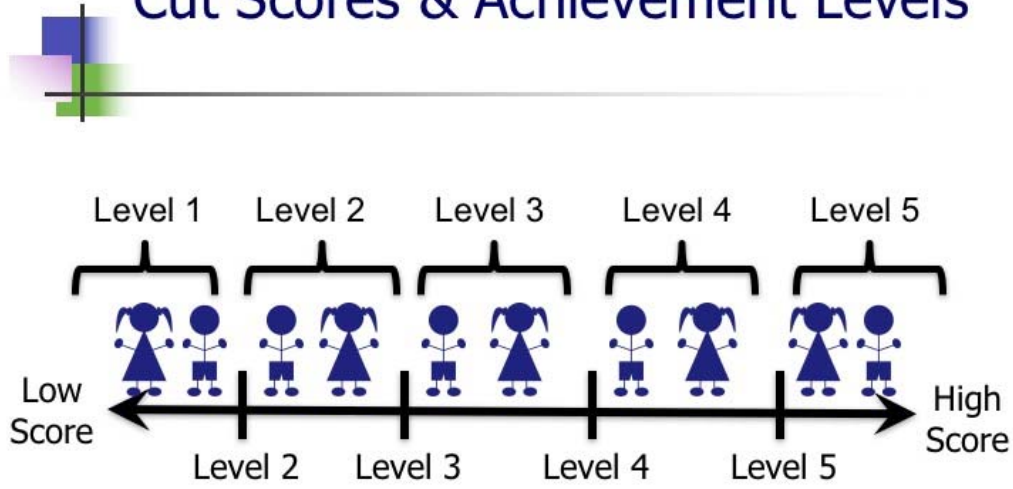


Reading

- Assign students to achievement levels
 - Policy descriptors
 - Assessment targets

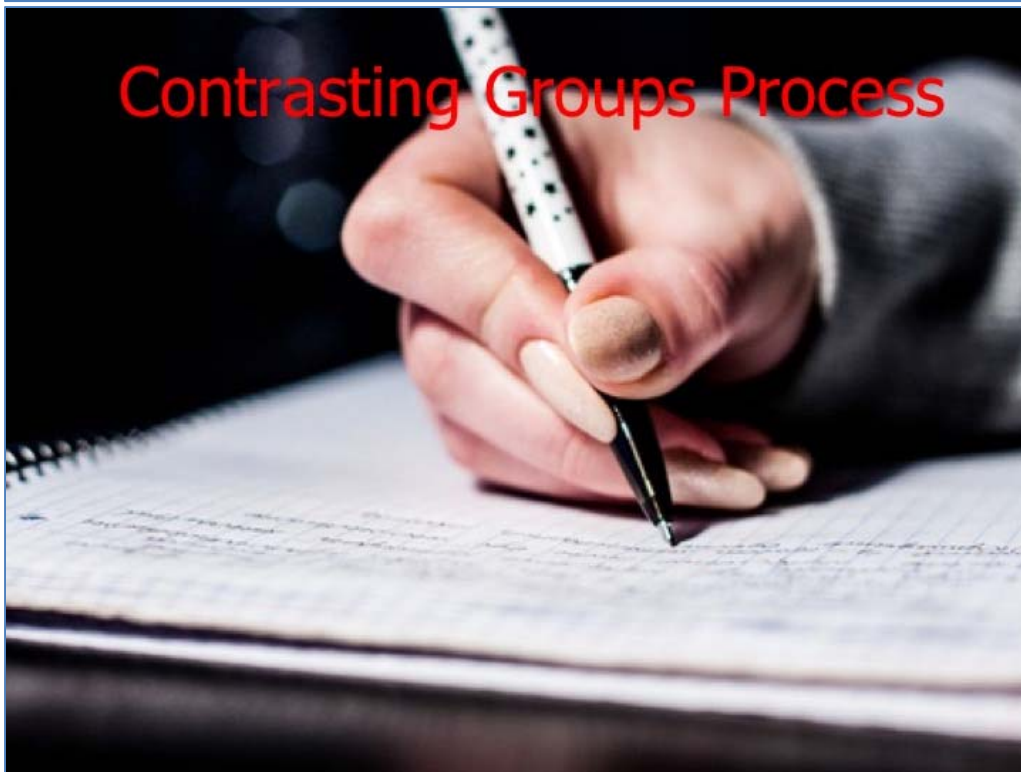


Cut Scores & Achievement Levels



 **ELPA21**
English Language Proficiency
Assessment for the 21st Century

Contrasting Groups Process





Before beginning

1. Class roster with state student ID
 - WA, OR, OH, IA: TIDE system
 - NE, AR, KS, WV: Nextera
2. School identification number



When you need help

sphillips@pacificmetrics.com





Tips and tricks

- All modern browsers are supported
- Most easily completed on a desktop or laptop
 - Mobile devices may be used, but the tool is not optimized for their use



Tips and tricks

- PC: Ctrl-C, Ctrl-V
- Mac: Command-C, Command-V
- Do not need to complete survey in one sitting



Appendix B4. Preliminary CGS Results: Profiles

Contrasting Groups Profiles for Level 4

Profile	N All Grades	N Grade K	N Grade 1	N Grade 2	N Grade 3	N Grade 4	N Grade 5	N Grade 6	N Grade 7	N Grade 8	N HS (9-12)
44444	352	24	43	59	62	39	32	22	22	28	21
44455	185	10	10	7	16	22	12	17	34	15	42
44344	165	8	33	19	30	24	14	11	2	5	19
43344	145	8	23	25	14	23	9	11	8	12	12
44454	75	6	7	8	10	6	6	2	4	2	24
44445	62	8	4	4	7	5	5	2	5	12	10
43444	50	6	12	9	1	4	2	8	3	3	2
44355	50	4	2	3	7	11	2	6	6	4	5
43354	42	1	2	5	2	8	4	1	5	3	11
43355	37	3	3	2	4	4	3	1	6	5	6
45455	33	1	5	2	6	2	1	4	3	0	9
43455	30	0	0	0	2	0	2	6	8	1	11
44434	27	5	4	3	2	4	1	2	1	1	4
44354	24	0	1	2	0	6	6	1	4	0	4
43343	23	0	2	6	4	4	1	2	2	0	2
43454	23	1	1	1	1	1	1	3	4	2	8
45444	20	4	2	4	4	4	0	0	1	1	0
44345	18	0	2	4	1	1	2	2	2	2	2
43334	14	1	2	2	0	0	2	1	2	3	1
43345	14	3	1	2	2	1	0	1	1	0	3
44443	14	1	1	5	3	0	3	0	1	0	0
44334	13	0	3	3	2	1	1	1	1	1	0
44555	12	2	0	1	0	0	3	1	2	0	3
44544	11	1	0	2	5	1	0	0	0	0	2
42344	9	3	1	1	2	1	0	0	0	0	1
43244	9	1	1	3	1	1	1	1	0	0	0
44343	8	0	1	2	2	3	0	0	0	0	0
44433	7	1	3	0	0	0	0	0	1	0	2
43445	6	0	0	0	0	3	0	0	0	0	3
45355	6	0	0	1	0	0	2	1	1	1	0
45445	6	0	1	2	1	0	0	0	0	0	2
45454	6	0	1	0	2	1	0	1	0	0	1
43443	5	0	1	1	1	0	0	2	0	0	0
44333	5	0	2	1	2	0	0	0	0	0	0
45544	5	2	0	2	0	0	0	0	0	1	0
42243	4	0	1	2	0	1	0	0	0	0	0
42244	4	2	0	1	1	0	0	0	0	0	0
43353	4	0	0	0	1	1	1	0	0	0	1
43434	4	2	1	0	0	0	1	0	0	0	0

44435	4	0	0	0	0	2	0	0	1	0	1
44453	4	1	0	0	0	0	1	0	0	1	1
44545	4	0	0	0	0	2	0	0	1	0	1
44554	4	0	0	0	0	1	2	0	0	0	1
43234	3	0	0	0	1	1	0	0	0	0	1
43243	3	1	1	1	0	0	0	0	0	0	0
43254	3	0	0	0	0	1	1	0	1	0	0
45345	3	0	1	1	0	0	0	0	1	0	0
45554	3	0	0	0	0	0	1	0	0	1	1
41144	2	1	0	0	0	0	0	0	0	0	1
42255	2	1	0	0	0	0	0	0	1	0	0
43245	2	1	0	0	0	0	0	0	0	0	1
43255	2	0	0	1	0	0	0	0	0	1	0
43342	2	0	0	0	0	0	0	0	1	0	1
44245	2	0	0	1	0	0	0	0	0	0	1
44335	2	0	0	0	0	1	0	0	0	0	1
44424	2	1	1	0	0	0	0	0	0	0	0
45353	2	0	0	0	0	1	0	0	1	0	0
45453	2	0	0	0	0	0	0	0	1	1	0
41114	1	0	0	0	0	0	1	0	0	0	0
41124	1	0	0	0	0	0	0	0	0	0	1
41135	1	0	0	0	0	0	0	0	0	0	1
41354	1	0	0	0	0	0	0	0	0	1	0
41355	1	0	1	0	0	0	0	0	0	0	0
42144	1	1	0	0	0	0	0	0	0	0	0
42254	1	0	0	0	0	0	0	0	0	1	0
42343	1	0	0	1	0	0	0	0	0	0	0
42444	1	0	0	0	0	1	0	0	0	0	0
42455	1	0	0	0	0	0	0	1	0	0	0
42543	1	0	1	0	0	0	0	0	0	0	0
43155	1	0	0	0	0	0	0	1	0	0	0
43335	1	0	1	0	0	0	0	0	0	0	0
43351	1	0	0	0	1	0	0	0	0	0	0
43422	1	0	1	0	0	0	0	0	0	0	0
43423	1	0	0	1	0	0	0	0	0	0	0
43453	1	0	0	0	0	0	0	0	1	0	0
43544	1	0	1	0	0	0	0	0	0	0	0
43554	1	0	0	1	0	0	0	0	0	0	0
44144	1	1	0	0	0	0	0	0	0	0	0
44243	1	0	0	0	0	0	0	0	1	0	0
44244	1	0	0	1	0	0	0	0	0	0	0
44254	1	0	0	0	0	0	1	0	0	0	0
44324	1	0	0	0	0	0	1	0	0	0	0
44353	1	0	0	0	0	0	0	0	0	1	0

44534	1	0	0	0	0	1	0	0	0	0	0
45335	1	0	0	1	0	0	0	0	0	0	0
45343	1	0	0	1	0	0	0	0	0	0	0
45433	1	0	0	1	0	0	0	0	0	0	0
45434	1	1	0	0	0	0	0	0	0	0	0
45545	1	0	0	0	1	0	0	0	0	0	0
45553	1	1	0	0	0	0	0	0	0	0	0

Contrasting Groups Study: 25 Most Common Patterns by Grade

Grade K Total N = 534				Grade 1 Total N = 548				Grade 2 Total N = 503			
Profile	N	%	Cum %	Profile	N	%	Cum %	Profile	N	%	Cum %
22222	91	11%	11%	22222	74	9%	9%	33333	63	8%	8%
11111	66	8%	19%	33333	71	9%	18%	44444	59	8%	16%
33333	53	7%	26%	44444	43	5%	24%	22222	30	4%	20%
33233	35	4%	30%	44344	33	4%	28%	32233	30	4%	24%
44444	24	3%	33%	33233	26	3%	31%	33344	29	4%	28%
21122	23	3%	36%	33344	23	3%	34%	55555	26	3%	31%
32233	22	3%	39%	55555	23	3%	37%	22233	26	3%	34%
33334	18	2%	41%	43344	23	3%	40%	43344	25	3%	38%
22233	17	2%	43%	11111	22	3%	43%	11111	23	3%	41%
11112	17	2%	45%	22233	22	3%	45%	44344	19	2%	43%
22223	16	2%	47%	33334	18	2%	48%	22223	19	2%	46%
22122	15	2%	49%	32233	17	2%	50%	33233	17	2%	48%
33344	14	2%	51%	32333	17	2%	52%	22232	17	2%	50%
32333	14	2%	53%	21122	15	2%	54%	33343	16	2%	52%
11122	13	2%	54%	22223	13	2%	56%	33244	12	2%	54%
22232	12	1%	56%	33343	13	2%	57%	32244	10	1%	55%
21132	11	1%	57%	22232	12	2%	59%	55455	9	1%	56%
21123	11	1%	58%	43444	12	2%	60%	43444	9	1%	58%
44455	10	1%	60%	11112	11	1%	62%	34333	9	1%	59%
44344	8	1%	61%	44455	10	1%	63%	33334	8	1%	60%
43344	8	1%	62%	22122	10	1%	64%	54455	8	1%	61%
44445	8	1%	63%	54455	9	1%	65%	21122	8	1%	62%
55555	7	1%	63%	32243	9	1%	66%	44454	8	1%	63%
33343	7	1%	64%	55455	8	1%	67%	32333	8	1%	64%
32244	7	1%	65%	44454	7	1%	68%	32232	8	1%	65%

Contrasting Groups Study: 25 Most Common Patterns by Grade (continued)

Grade 3 Total N = 525				Grade 4 Total N = 472				Grade 5 Total N = 273			
Profile	N	%	Cum %	Profile	N	%	Cum %	Profile	N	%	Cum %
33333	75	10%	10%	33333	50	7%	7%	44444	32	8%	8%
44444	62	8%	19%	33344	46	7%	14%	33333	25	6%	14%
33344	40	5%	24%	44444	39	6%	20%	55555	22	5%	20%
11111	32	4%	28%	55555	32	5%	24%	22222	19	5%	24%
44344	30	4%	32%	44344	24	3%	28%	33344	19	5%	29%
22222	27	4%	36%	11111	23	3%	31%	44344	14	3%	33%
55555	27	4%	40%	43344	23	3%	34%	11111	13	3%	36%
22233	24	3%	43%	55455	23	3%	38%	44455	12	3%	39%
33233	23	3%	46%	22222	22	3%	41%	22233	12	3%	42%
32233	22	3%	49%	44455	22	3%	44%	55455	10	2%	44%
44455	16	2%	51%	33343	20	3%	47%	43344	9	2%	47%
33343	15	2%	53%	33334	17	2%	49%	33244	9	2%	49%
43344	14	2%	55%	33244	14	2%	51%	33233	8	2%	51%
33334	14	2%	57%	33233	12	2%	53%	32233	8	2%	53%
22122	13	2%	59%	32233	12	2%	55%	22232	6	1%	54%
22223	12	2%	60%	22232	12	2%	57%	44454	6	1%	56%
22232	12	2%	62%	22233	11	2%	58%	44354	6	1%	57%
55455	10	1%	63%	22223	11	2%	60%	22223	5	1%	58%
44454	10	1%	65%	44355	11	2%	61%	33343	5	1%	60%
21122	7	1%	66%	54455	10	1%	63%	44445	5	1%	61%
44445	7	1%	67%	43354	8	1%	64%	11122	5	1%	62%
44355	7	1%	68%	33243	7	1%	65%	32244	5	1%	63%
32243	7	1%	69%	44454	6	1%	66%	32243	5	1%	65%
34343	7	1%	70%	32333	6	1%	67%	33243	5	1%	66%
33244	6	1%	70%	44354	6	1%	68%	33334	4	1%	67%

Contrasting Groups Study: 25 Most Common Patterns by Grade (continued)

Grade 6 Total N = 295				Grade 7 Total N = 340				Grade 8 Total N = 347			
Profile	N	%	Cum %	Profile	N	%	Cum %	Profile	N	%	Cum %
11111	38	9%	9%	11111	38	7%	7%	11111	55	11%	11%
33344	23	5%	14%	44455	34	7%	14%	44444	28	6%	17%
44444	22	5%	19%	33333	32	6%	20%	33333	27	6%	23%
22222	19	4%	24%	55555	32	6%	27%	55555	26	5%	28%
44455	17	4%	28%	33344	27	5%	32%	33344	17	4%	32%
33333	13	3%	31%	44444	22	4%	36%	44455	15	3%	35%
55555	12	3%	34%	22222	17	3%	40%	22233	15	3%	38%
44344	11	3%	36%	22223	15	3%	43%	22222	14	3%	41%
32233	11	3%	39%	22233	12	2%	45%	22223	14	3%	44%
43344	11	3%	42%	33343	10	2%	47%	54455	13	3%	47%
33343	11	3%	44%	32233	9	2%	49%	43344	12	2%	49%
22233	10	2%	46%	43344	8	2%	50%	44445	12	2%	52%
33334	10	2%	49%	43455	8	2%	52%	33334	11	2%	54%
55455	10	2%	51%	33233	7	1%	53%	11122	10	2%	56%
11112	8	2%	53%	33334	7	1%	55%	11112	9	2%	58%
43444	8	2%	55%	55455	7	1%	56%	54555	9	2%	60%
22223	7	2%	57%	54455	7	1%	57%	55455	8	2%	61%
22232	7	2%	58%	44355	6	1%	58%	32333	8	2%	63%
54455	7	2%	60%	33354	6	1%	60%	33343	7	1%	64%
33244	6	1%	61%	43355	6	1%	61%	32233	6	1%	66%
32244	6	1%	63%	21122	5	1%	62%	21122	6	1%	67%
44355	6	1%	64%	33244	5	1%	63%	44344	5	1%	68%
43455	6	1%	65%	22122	5	1%	64%	33233	5	1%	69%
11121	6	1%	67%	44445	5	1%	65%	34344	5	1%	70%
32333	5	1%	68%	11122	5	1%	66%	43355	5	1%	71%

Contrasting Groups Study: 25 Most Common Patterns by Grade (continued)

Grade 9-12 Total N = 638			
Profile	N	%	Cum %
11111	95	10%	10%
22222	55	6%	16%
33333	51	5%	22%
55555	42	5%	26%
44455	42	5%	31%
33344	37	4%	35%
55455	24	3%	37%
44454	24	3%	40%
11112	23	2%	42%
44444	21	2%	44%
44344	19	2%	47%
22223	19	2%	49%
33343	19	2%	51%
54455	19	2%	53%
22233	18	2%	55%
33233	17	2%	56%
33244	13	1%	58%
32233	12	1%	59%
43344	12	1%	60%
33354	12	1%	62%
22122	11	1%	63%
43354	11	1%	64%
34344	11	1%	65%
43455	11	1%	66%
22232	10	1%	67%

Appendix B5. CGS Results: Comparison of CGS and State Data

Grade	Data Source	N		Level 1	Level 2	Level 3	Level 4	Level 5	Level 4 + Level 5
K	CGS	808		15%	36%	32%	15%	2%	17%
	State Median			16%	38%	30%	12%	3%	14%
	Difference: CGS - State			-1%	-2%	2%	4%	-1%	3%
	Ohio OTELA	6240		23%	41%	27%	8%	1%	9%
	Oregon ELPA			20%	36%	25%	15%	4%	19%
	Arkansas ELDA			11%	29%	40%	16%	5%	21%
	WELPA	17,082		9%	49%	32%	8%	0%	8%
1	CGS	792		6%	27%	37%	23%	7%	30%
	State Median			9%	32%	30%	20%	6%	26%
	Difference: CGS - State			-3%	-5%	7%	3%	1%	4%
	Ohio OTELA			9%	28%	30%	25%	7%	32%
	Oregon ELPA			20%	36%	25%	15%	4%	19%
	Arkansas ELDA			8%	27%	29%	28%	8%	36%
	WELPA	16,947		2%	37%	47%	13%	0%	13%
2	CGS	763		5%	20%	39%	27%	8%	35%
	State Median			7%	28%	33%	20%	9%	29%
	Difference: CGS - State			-2%	-8%	6%	7%	-1%	6%
	Ohio OTELA			9%	28%	30%	25%	7%	32%
	Oregon ELPA			13%	27%	35%	15%	10%	26%
	Arkansas ELDA			5%	19%	27%	36%	13%	49%

Grade	Data Source	N		Level 1	Level 2	Level 3	Level 4	Level 5	Level 4 + Level 5
	WELPA	15,092		2%	32%	50%	15%	0%	15%
3	CGS	738		6%	19%	40%	27%	7%	34%
	State Median			11%	25%	34%	18%	5%	24%
	Difference: CGS - State			-5%	-6%	6%	9%	2%	10%
	Ohio OTELA			8%	23%	33%	27%	8%	35%
	Oregon ELPA			13%	27%	35%	15%	10%	26%
	Arkansas ELDA			16%	30%	33%	21%	1%	22%
	WELPA	12,799		2%	18%	65%	14%	0%	14%
4	CGS	691		5%	15%	40%	28%	12%	40%
	State Median			7%	17%	32%	31%	6%	40%
	Difference: CGS - State			-2%	-2%	8%	-3%	6%	0%
	Ohio OTELA			8%	23%	33%	27%	8%	35%
	Oregon ELPA			6%	11%	24%	36%	23%	59%
	Arkansas ELDA			8%	19%	30%	41%	3%	44%
	WELPA	10,790		2%	15%	65%	17%	0%	17%
5	CGS	402		5%	17%	35%	31%	11%	42%
	State Median			6%	12%	29%	31%	8%	47%
	Difference: CGS - State			-1%	5%	6%	0%	3%	-5%
	Ohio OTELA			8%	23%	33%	27%	8%	35%
	Oregon ELPA			6%	11%	24%	36%	23%	59%
	Arkansas ELDA			5%	12%	25%	51%	7%	58%

Grade	Data Source	N		Level 1	Level 2	Level 3	Level 4	Level 5	Level 4 + Level 5
	WELPA	9,348		2%	12%	58%	26%	0%	26%
6	CGS	426		15%	19%	31%	26%	8%	34%
	State Median			7%	13%	29%	28%	6%	46%
	Difference: CGS - State			8%	6%	2%	-2%	2%	-12%
	Ohio OTELA			14%	19%	30%	28%	8%	36%
	Oregon ELPA			6%	10%	29%	27%	28%	55%
	Arkansas ELDA			7%	12%	25%	52%	4%	56%
	WELPA	6,562		3%	14%	56%	25%	0%	25%
7	CGS	510		12%	18%	31%	27%	12%	39%
	State Median			7%	11%	29%	28%	8%	46%
	Difference: CGS - State			5%	7%	2%	-1%	4%	-7%
	Ohio OTELA			14%	19%	30%	28%	8%	36%
	Oregon ELPA			6%	10%	29%	27%	28%	55%
	Arkansas ELDA			8%	9%	18%	57%	8%	65%
	WELPA	5,785		3%	13%	60%	22%	0%	22%
8	CGS	481		18%	18%	28%	23%	14%	37%
	State Median			7%	11%	29%	31%	10%	46%
	Difference: CGS - State			11%	7%	-1%	-8%	4%	-9%
	Ohio OTELA			14%	19%	30%	28%	8%	36%
	Oregon ELPA			6%	10%	29%	27%	28%	55%
	Arkansas ELDA			7%	8%	16%	58%	11%	69%

Grade	Data Source	N		Level 1	Level 2	Level 3	Level 4	Level 5	Level 4 + Level 5
	WELPA	5,154		3%	11%	51%	33%	0%	33%
HS	CGS	931		15%	20%	30%	24%	11%	35%
	State Median			11%	16%	26%	33%	13%	46%
	Difference: CGS - State			5%	4%	4%	-9%	-2%	-11%
	Ohio OTELA			16%	20%	30%	29%	5%	34%
	Oregon ELPA			11%	5%	22%	40%	22%	62%
	Arkansas ELDA			10%	12%	19%	38%	21%	59%
	WELPA	13,244		5%	22%	55%	18%	0%	18%

Appendix C. EL Expert Advisory Panel Materials and Results

Appendix C1. Slides from May 11th EL Expert Meeting

6/11/2016

Meeting Agenda for ELPA21 EL Experts Overview




Introductions,
Standard Setting
Review, and
Q and A

Introduction to
Proficiency
Determination Profiles
and Discussion

Preview of Next EL
Expert Activities:
Introduction to
Achievement Level
Descriptors

Daniel Lewis
Pacific Metrics

Karla Egan
EdMetric LLC




Introductions

Name:

Affiliation and role:

Grade band experience (EL):

- Elementary ,
- Middle, and/or
- High School



ELPA21 Standard Setting

Review and Q & A



Background

Unique assessments were developed for the following six grade bands:

- K
- 1
- 2-3
- 4-5
- 6-8
- 9-12 (High School)

Assessments in each grade band measure the following four domains:

- Reading
- Writing
- Listening
- Speaking



Background

Five achievement levels describe performance on each domain.

Policy Definitions for Domain Performance	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Begins to develop a few grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Demonstrates superior English language skills, as measured by ELPA21.

Overall Proficiency Determination

ELPA21 policy guides the Overall Proficiency Determination, as follows:

- *ELPA21 recommends that students be considered proficient when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once proficient on ELPA21, ELs can be considered for reclassification.*



Overall Proficiency Determination

- Need to determine profiles associated with Overall Proficiency Determinations
 - Not Proficient
 - Nearly Proficient
 - Proficient



Process

- Systematic process to determine profiles of proficiency for Overall Proficiency Determination
 - Contrasting Groups Study
 - EL Expert recommendations
 - In person standard setting workshop in July
 - ELPA21 makes final determination based on above input



EL Expert Engagement for Overall Proficiency Determination

1. Today's web meeting (introduction)
2. Each EL Expert completes a survey (to be sent early next week; completed by end of next week)
3. Pacific Metrics aggregates survey results
4. EL Experts will be reconvened in web meeting to review results, discuss differences, and reconsider recommendations
5. Pacific Metrics summarizes EL Expert recommendations



Contrasting Groups Study Results

- Teachers in ELPA21 states read the policy descriptors and completed surveys
 - They classified their students into one of five levels for Reading, Writing, Listening, Speaking, and Overall based on the policy descriptors and their knowledge of the students' classroom achievement.



Sample of Raw CGS Results Pattern = ORWLS*

Pattern	Overall	Reading	Writing	Listening	Speaking	All Grades	%	Cum %
33333	3	3	3	3	3	449	6.9%	6.9%
11111	1	1	1	1	1	395	6.1%	13.0%
22222	2	2	2	2	2	357	5.5%	18.5%
44444	4	4	4	4	4	343	5.3%	23.7%
33344	3	3	3	4	4	267	4.1%	27.8%
55555	5	5	5	5	5	244	3.8%	31.6%
44455	4	4	4	5	5	179	2.8%	34.4%
22233	2	2	2	3	3	165	2.5%	36.9%
44344	4	4	3	4	4	163	2.5%	39.4%
33233	3	3	2	3	3	150	2.3%	41.7%
32233	3	2	2	3	3	145	2.2%	43.9%
43344	4	3	3	4	4	142	2.2%	46.1%
22223	2	2	2	2	3	130	2.0%	48.1%
33343	3	3	3	4	3	120	1.8%	50.0%
33334	3	3	3	3	4	115	1.8%	51.7%
55455	5	5	4	5	5	113	1.7%	53.5%
22232	2	2	3	3	2	92	1.4%	54.9%
11112	1	1	1	1	2	80	1.2%	56.1%
54455	5	4	4	5	5	80	1.2%	57.3%

* ORWLS = Overall Reading Writing Listening Speaking

 **ELPA21**
English Language Proficiency
Assessment for the 21st Century

Most Common CGS Profiles

Profile	K-12	%	Cum %	
44444	343	21.7%	21.7%	All 4s
44455	179	11.3%	33.0%	All 4s and 5s
44344	163	10.3%	43.3%	4s with a 3 in Writing
43344	142	9.0%	52.3%	4s with 3s in Reading and Writing
44454	73	4.6%	56.9%	All 4s and 5s
44445	59	3.7%	60.6%	All 4s and 5s
44355	49	3.1%	63.7%	4s and 5s with a 3 in Writing
43444	49	3.1%	66.8%	4s with a 3 in Reading
43354	42	2.7%	69.5%	4s and 5s with 3s in Reading and Writing
43355	36	2.3%	71.7%	5s with a 3 in Reading and Writing
45455	32	2.0%	73.8%	5s with a 4 in Writing

ORWLS = Overall Reading Writing Listening Speaking

 **ELPA21**
English Language Proficiency
Assessment for the 21st Century

Profiles by Grade

Profile	K-HS %	K %	G1 %	G2 %	G3 %	G4 %	G5 %	G6 %	G7 %	G8 %	HS%
44444	21.7%	20.0%	23.4%	28.9%	30.5%	20.5%	25.0%	19.8%	14.8%	24.3%	24.3%
44455	11.3%	7.8%	4.9%	3.5%	7.1%	11.4%	9.7%	15.3%	25.2%	13.5%	13.5%
44344	10.3%	7.0%	17.9%	9.5%	14.7%	13.0%	11.3%	9.9%	1.5%	4.5%	4.5%
43344	9.0%	7.0%	12.5%	11.9%	7.1%	11.9%	7.3%	9.0%	5.9%	10.8%	10.8%
44454	4.6%	5.2%	3.8%	3.5%	5.1%	3.2%	4.8%	1.8%	2.2%	1.8%	1.8%
44445	3.7%	6.1%	2.2%	2.0%	3.6%	2.7%	3.2%	1.8%	3.7%	10.8%	10.8%
44355	3.1%	3.5%	1.1%	1.5%	3.6%	5.4%	1.6%	5.4%	4.4%	3.6%	3.6%
43444	3.1%	5.2%	6.5%	4.5%	0.5%	2.2%	1.6%	7.2%	2.2%	1.8%	1.8%

ORWLS = Overall Reading Writing Listening Speaking



Summary of Main Results

- Profiles associated with an Overall Proficiency Determination of Proficient may be defined by 3 or 4 rules.
 - Overall Proficiency can be indicated if W = 3 and other domains are 4 or 5
 - Overall Proficiency can be indicated if W and R = 3 and S and L = 4 or 5
 - Any combination of all 4s and 5s indicates overall proficiency



Discussion

- Overall Proficiency Determination of Proficient is generally associated with Level 4 on all domains
 - There may be special cases, other profiles representing different skill combinations, that may be associated with proficiency
 - We saw such cases in the CGS study
 - Reading and/or Writing at Level 3 and Level 4 or 5 on the remaining domains
 - Is this defensible? Why or why not?
 - Are there other special cases you would consider for the Overall Proficiency Determination of Proficient?



Discussion: Profiles

- Should the profiles associated with Proficient, Nearly Proficient, and Not Proficient be the same for all grade bands?
 - Should they differ by grade band?
 - Why or why not?



Discussion: Nearly Proficient & Not Proficient

- We have a clear definition of Proficient
 - ELPA21 recommends that students be considered proficient when they reach a level of English language aptitude sufficient to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English.
- Policy definitions for Nearly Proficient and Not Proficient are still being developed
 - Recommendations?
 - Students are considered Nearly Proficient if they have not yet reached the proficient level, but demonstrate sufficient progress toward proficiency that with continued program support they are expected to demonstrate proficiency within ? years.
 - Other definitions that may be useful?



Discussion: Nearly Proficient & Not Proficient

- We need a definition of Nearly Proficient to answer the following questions
 - Should Nearly Proficient be associated with an Overall Proficiency Determination of 3?
 - So Overall Proficiency Determination of 1 or 2 is Not Proficient?
 - Could we apply the same rules to Overall Proficiency Determination for 3 as we do for 4?
 - e.g., 3s and perhaps 2s on Reading or Writing?
 - Is a student with 2222 Nearly Proficient?
- A student will be classified as Not Proficient if his or her profile is not associated with Nearly Proficient or Proficient



Questions



Next Steps:

Draft Target Achievement Level Descriptors

- Achievement level *descriptors* define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of achievement.

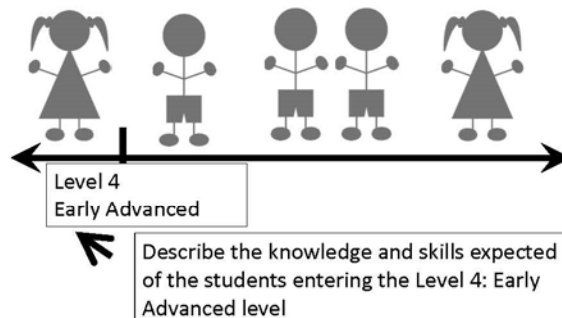


What are Achievement Level Descriptors (ALDs)?

- Achievement levels
 - Aggregate students into groups based on test achievement.
 - Under NCLB: *Below Basic, Basic, Proficient, or Advanced*
 - Target ALDs describe what students *should be able to do* to enter a category
 - to the content standards



Target ALDs for Early Advanced



Future Meeting

- Create Preliminary Target ALDs for use at ELPA21 standard setting
 - ELPA21 proficiency standards and policy descriptors will frame this work



Appendix C2. EL Expert Panel Survey

EL Expert Survey: Introduction

In each grade band, student achievement in each domain is reported in terms of five achievement levels. The five levels and associated policy definitions follow:

Policy Definitions for Domain Performance	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Begins to develop a few grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Exhibits some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Demonstrates superior English language skills, as measured by ELPA21.

Level 1 is the lowest level and Level 5 is the highest level. Level 4 is the assessment target where students demonstrate the skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.

The teachers in the Contrasting Group Study (CGS) classified their students' achievement into 5 levels on their Overall performance and on each domain based on the policy definitions and observation.

We report the teachers' classifications as ORWLS (Overall, Reading, Writing, Listening, and Speaking).

Some common teacher classifications follow:

44444 (Overall Level 4, Reading Level 4, Writing Level 4, Listening Level 4, Speaking Level 4)

33333 (Overall Level 3, Reading Level 3, Writing Level 3, Listening Level 3, Speaking Level 3)

44454 (Overall Level 4, Reading Level 4, Writing Level 4, Listening Level 5, Speaking Level 4)

33343 (Overall Level 3, Reading Level 3, Writing Level 3, Listening Level 4, Speaking Level 3)

Survey Part I

The Overall Proficiency Determination—a determination based on the pattern of achievement on all 4 domains—will be reported in terms of three levels-- Not Proficient, Nearly Proficient, and Proficient.

Draft definitions of these three levels are provided below, and are subject to modification.

Not Proficient	ELPA21 recommends that students be considered Not Proficient when they have not yet attained a level of English language necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. Students scoring Not Proficient on ELPA21 are eligible for ongoing program support.
Nearly Proficient	ELPA21 recommends that students be considered Nearly Proficient when they are approaching a level of English language necessary to produce, interpret, and collaborate, with support, on grade-level content-related academic tasks in English. Students scoring Nearly Proficient on ELPA21 are recommended to receive ongoing program support.
Proficient	ELPA21 recommends that students be considered Proficient when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. Once Proficient on ELPA21, ELs can be considered for reclassification.

We need to translate the five-level Overall classification system used by the CGS teachers into the three levels that will be used by ELPA21 to report students' Overall Proficiency Determination.

In each row, please associate each of the five Overall levels with one of the three Overall Proficiency Determination levels based on your expert opinion (place an X to the left of your choice).

Five-Level Overall Classification		Three-Level Overall Proficiency Determination		
Level 1 (Beginning)	=	____ Not Proficient	____ Nearly Proficient	____ Proficient
Level 2 (Early Intermediate)	=	____ Not Proficient	____ Nearly Proficient	____ Proficient
Level 3 (Intermediate)	=	____ Not Proficient	____ Nearly Proficient	____ Proficient
Level 4 (Early Advanced)	=	____ Not Proficient	____ Nearly Proficient	____ Proficient
Level 5 (Advanced)	=	____ Not Proficient	____ Nearly Proficient	____ Proficient

Comments (optional):

Survey Part II

We would like your expert opinion on specific profiles that should be associated with an Overall Level 4 based on the five-level Overall classification scheme used by CGS teachers.

We examined the CGS teachers' profiles classified as Overall Level 4. Note that we are not examining Overall Level 5s here, just Level 4 for this exercise. We organized all profiles classified to Overall Level 4 into the following categories:

Category (Percent of Profiles Classified as Overall 4)	Exemplar Patterns Assigned as Overall Level 4 RWLS Reading, Writing, Listening, Speaking (Percent of Profiles Classified as Overall 4)	Description of Category
A (48.4%)	4444 (21.9%), 4455 (11.5%), 4454 (4.7%), etc.	All domains are classified as 4s or 5s.
B (38.0%)	4344 (10.2%), 3344 (9.0%), 3444 (3.1%) , etc.	Reading and/or Writing are classified as 3 but Listening and Speaking are classified as 4.
C (7.5%)	4434 (1.7%), 3343 (1.4%), 3334 (.9%), etc.	There are one or more 3s in a domain other than Reading and Writing and other domains are 4s.
D (1.9%)	3353 (.2%) , 4435 (.2%) , 4453 (.2%), etc.	There are at least one 3 and one 5 in a domain and the other domains are 4s or 5s. This is a compensatory profile where the teachers may have assumed that one or more 5s in one domain compensated for one or more 3s in another domain.
E (4.3%)	2344 (.6%), 3244 (.6%), 2243 (.2%), 1144 (.1%), etc.	One or more domains are classified as a 1 or 2.

Use the following form to agree or disagree with the assignment of each category to Level 4 by placing an X to the left of one of the following three choices for each row in the table:

Agree. You agree that profiles matching the category description should be Overall Level 4

Disagree. You disagree; you believe that profiles matching the category description should be Level 3 or lower, or

Other. You neither agree nor disagree and believe that neither choice captures your opinion. If you select "other," please state your opinion in the comment box.

You may also make a comment to clarify your agreement or disagreement. If you are unable to decide, please note what you would need in order to make a decision.

Category	Exemplar Patterns Assigned as Overall Level 4 RWLS Reading, Writing, Listening, Speaking	Decision Criteria	Agree, Disagree, or Other	Comment
A	4444, 4544, 4554, etc.	All domains classified as 4s or 5s <i>should be Overall Level 4</i>	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
B	3444, 4344, 3344, etc.	Reading and/or Writing classified as 3 but Listening and Speaking classified as 4 <i>should be Overall Level 4</i>	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
C	4434, 4443, 4433, 4434, etc.	One or more 3s in a domain other than Reading and Writing and the other domains are 4s <i>should be classified as Overall Level 4</i>	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
D	4435, 3445, 3355, etc.	At least one 3 and one 5 in a domain and the other domains are 4s <i>should be classified as Overall Level 4. (compensatory profile)</i>	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
E	3243, 1135, 2244, etc.	One or more domains classified as a 1 or 2 <i>can be classified as Overall Level 4.</i>	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	

Survey Part III

We would like your expert opinion on specific profiles that should be associated with an Overall Level 3 based on the five-level Overall classification scheme used by CGS teachers.

We examined the CGS teachers' profiles classified as Overall Level 3. Note that we are not examining Overall Levels other than 3s here, just Level 3 for this exercise. We organized all profiles classified to Overall Level 3 into the following categories:

Category (Percent of Profiles Classified as Overall 3)	Exemplar Patterns Assigned as Overall Level 3 RWLS Reading, Writing, Listening, Speaking (Percent of Profiles Classified as Overall 3)	Description of Category
A (55.8%)	3333 (20.2%), 3344 (12.1%), 3343 (5.4%), etc.	All domains are classified as at least 3.
B (33.4%)	3233 (6.7%), 2233 (6.5%), 3244 (3.2%), etc.	Reading and/or Writing are classified as 2 but Listening and Speaking are classified as at least 3.
C (7.5%)	2232 (1.1%), 3323 (1.1%), 2223 (1.1%), etc.	There are one or more 2s in a domain other than Reading and Writing.
D (1.0%)	3342 (.4%), 3324 (.2%), 3423 (.1%), etc.	There is at least one 2 and at least as many 4s or 5s as there are 2s. There are no 1s. This is a compensatory profile where the teachers may have assumed that the 4s or 5s compensated for the 2s.
E (2.3%)	2133 (.4%), 1144 (.1%), 2123 (.1%), etc.	One or more domains are classified as a 1.

Note that some profiles classified as B could have also been classified as D (but were not so that the Category percentages sum to 100%). That is, some profiles with 2s in Reading and/or Writing may have been compensated by as many 4s and 5s in Listening and/or Speaking. We can address this as a special case in your recommendations below.

Use the following form to agree or disagree with the assignment of each category to Level 3 by placing an X to the left of one of the following three choices for each row in the table:

Agree. You agree that profiles matching the category description should be Overall Level 3

Disagree. You disagree; profiles matching the category description should be Level 2 or lower, or

Other. You neither agree nor disagree and believe that neither choice captures your opinion. If you select "other," please state your opinion in the comment box.

You may also make a comment to clarify your agreement or disagreement. If you are unable to decide, please note in the Comment box what you would need in order to make a decision.

Category	Exemplar Patterns Assigned Overall Level 3 RWLS Reading, Writing, Listening, Speaking	Decision Criteria	Agree, Disagree, or Other	Comment
A	3333, 3344, 3343, etc.	All domains classified as at least 3 should be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
B	3233, 2233, 3244, etc.	Reading and/or Writing classified as 2 and Listening and Speaking classified as at least 3 should be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
C	2232, 3323, 2223, etc.	One or more 2s in a domain other than Reading and Writing can still be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
D	3342, 3324, 3423, etc.	There is at least one 2 and at least as many 4s or 5s as there are 2s (with no 1s) should be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
E	2133, 1144, 2123, etc.	One or more domains classified as a 1 can still be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	
Special Case	3244, 2244, 3243, etc.	If there are 2s only in Reading and/or Writing but there are at least as many 4s or 5s as 2s, the Overall Profile should be classified as Overall Level 3.	<input type="checkbox"/> Agree <input type="checkbox"/> Disagree <input type="checkbox"/> Other	

Appendix D. In-Person Bookmark Standard Setting Workshop Training

Appendix D1. Table Facilitator Training Slides

ELPA 21 Standard Setting

Process Overview & Training

Karla Egan

Mary Seburn, ELPA21/Quantiful



Overview of Table Leader Training



Standard
Setting



Bookmark
Procedure



Your Role



Q & A



Understanding the Standard Setting Process



What is Standard Setting?

Outcomes aid
interpretations

Process that
allows experts to
make decisions

Connects
assessments
and standards

Standard
Setting



How do we do this?

You provide:

- Knowledge of student performance
- Knowledge of English-language proficiency

We provide:

- English Language Proficiency Standards
- Policy level descriptors
- Target students
- Process for setting cut scores



English Language Proficiency Standards

- ELP Standards highlight skills needed by ELLs as they develop competencies in academic areas
 - Language functions: what students do with language to accomplish content-specific tasks
 - Language forms: vocabulary, grammar, and discourse specific to a content area
- Divided into five levels
 - What might an ELL's language use look like at each ELP level as she progresses toward independent participation in grade-appropriate activities?



Writing—ELPA21 Preliminary Achievement Level Indicators—Kindergarten

Beginner	Early Intermediate	Intermediate	Early Advanced	Advanced
<p>The learner can:</p> <ul style="list-style-type: none"> • respond to simple yes/no questions and some wh-questions about familiar topics • communicate simple information about familiar topics or experiences • express a feeling or opinion about a familiar topic • use a small number of frequently used nouns and verbs 	<p>The learner can:</p> <ul style="list-style-type: none"> • respond to simple yes/no questions and wh-questions about familiar topics • communicate simple information about familiar topics, experiences or events • express an opinion or preference about a familiar topic • use frequently occurring nouns, verbs, and short phrases; produce a few simple sentences 	<p>The learner can:</p> <ul style="list-style-type: none"> • respond to simple yes/no questions and wh-questions about familiar topics; • communicate simple information about familiar topics, experiences or events • express an opinion or preference about a familiar topic or story • use frequently occurring regular plural nouns, verbs and prepositions; produce simple sentences 	<p>The learner can:</p> <ul style="list-style-type: none"> • ask and answer simple questions and wh-questions • dictate simple messages about a variety of topics or experiences • express an opinion or preference about a variety of topics or stories • use frequently occurring regular plural nouns, verbs and prepositions; produce and expand simple sentences 	<p>The learner can:</p> <ul style="list-style-type: none"> • ask and answer simple questions and wh-questions • compose short written texts about a variety of topics, experiences, or events • express an opinion or preference about a variety of topics or stories • use frequently occurring regular plural nouns, verbs, and prepositions; produce and expand simple sentences



Policy Level Descriptors

- High-level definition of type of performance expected at each proficiency level

A STUDENT AT THIS LEVEL...	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of developing grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits superior English language skills, as measured by ELPA21.



Assessment Targets

Assessment Targets	
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Overall	ELPA21 recommends that students be considered proficient when they attain a level of English language aptitude sufficient to produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks without on-going EL Program support. Once proficient on ELPA21, ELs can be considered for reclassification.



ELPA21 Cut Scores

- Four cut scores
 - Level 2, Level 3, Level 4, Level 5
- Five achievement levels
 - Level 1, Level 2, Level 3, Level 4, Level 5
- You will recommend: Levels 3 & 4



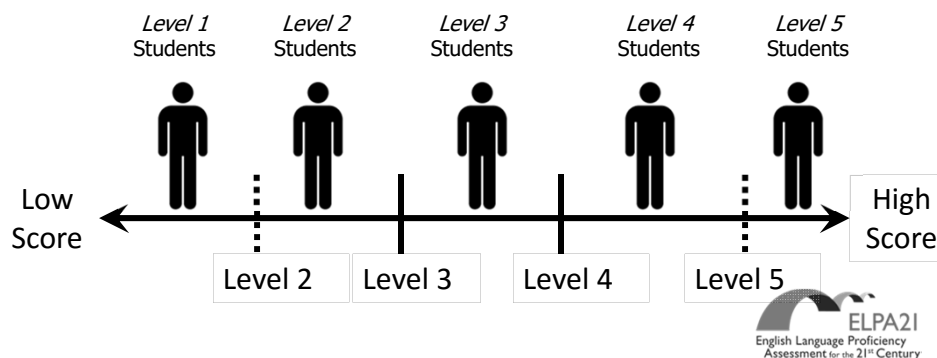
How do we set cut scores?

- Percentages
 - Arbitrary
 - Test specific
 - Content is not considered
- Content
 - Use pre-established ELP proficiency standards
 - Considers educational objectives
- Bookmark Standard Setting Procedure



Purpose of the Standard Setting

- Allows cut scores to be set on the test scale
 - Test scale represents range of student scores
 - Cut scores will separate students into achievement levels



Purpose of the Standard Setting

Establish Four Cut Scores

Recommended by content experts	On the test scale	Decisions based on content
-----------------------------------	-------------------	-------------------------------

Students placed in Achievement Levels

Students who meet or exceed the cut score have enough knowledge, skills, and abilities to be classified as Level 4



Developing Expertise

- You know students and content
- Activities to enhance knowledge:
 1. Study ELP standards and ALIs
 2. Study test items
 - Take practice test
 - Study ordered item booklets using item maps
 3. Study target student descriptors



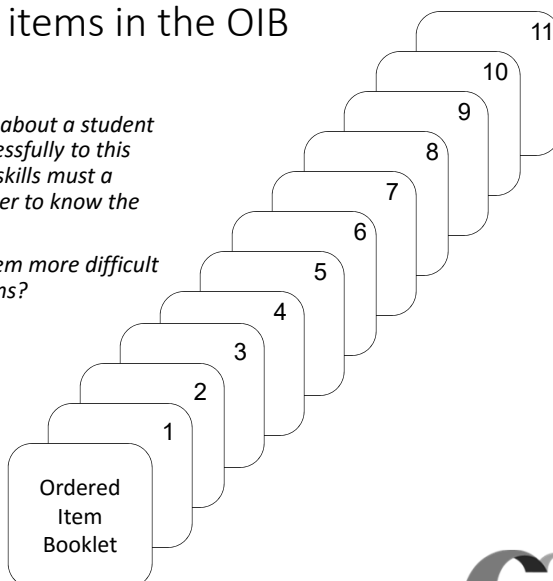


Bookmark Standard Setting

- Key Materials
 - Ordered Item Booklets
 - Item Map

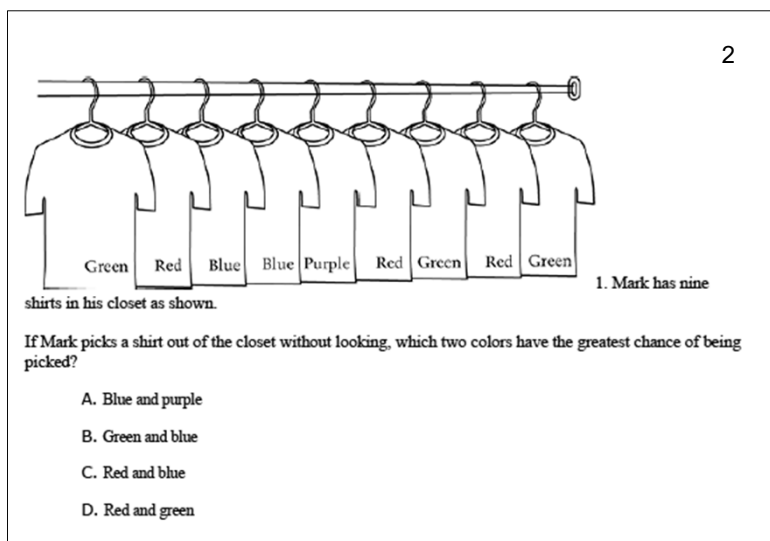
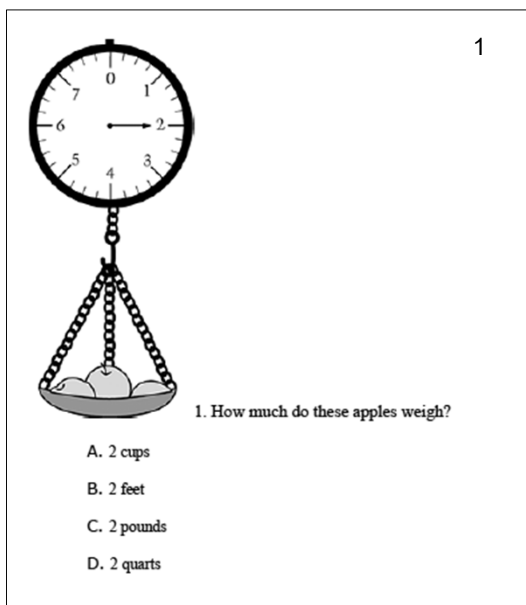
Studying the items in the OIB

1. *What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?*
2. *What makes this item more difficult than preceding items?*



Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID	Content Classification	Item Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	160	1	2013-4M7 #1 M154	Measurement	MC		
2	236	1	2013-4M7 #2 M170	Data analysis, Statistics, & Probab	MC		
3	251	1 of 2	2013-4M6 #14 M13	Data analysis, Statistics, & Probab	CR		
4	257	1	2013-4M7 #6 M158	Geometry	MC		
5	260	1	2013-4M7 #5 M16	Algebra	MC		
6	282	2 of 2	2013-4M6 #14 M13	Data analysis, Statistics, & Probab	CR		
7	295	1	2013-4M7 #12 M16	Number properties and operations	MC		

1. *What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?*
2. *What makes this item more difficult than preceding items?*



3

1 of 2 score points

1. Al, Bev, and Carmen are going on a ride at the park. Only 2 people can go on the ride at a time. They can pair up 3 different ways, as shown below.

Al and Bev

Al and Carmen

Bev and Carmen

Derek decides to join the group. How many different ways can the 4 students pair up?

Answer: _____

Show your work or explain how you got your answer.

You will have rubrics for constructed response items



4



shapes shown?

1. Which statement is true about all four

- A. Each shape is a rectangle.
- B. Each shape is a quadrilateral.
- C. Each shape has two pairs of parallel sides.
- D. Each shape has one or more right angles.



7
2 of 2 score points

1. Al, Bev, and Carmen are going on a ride at the park. Only 2 people can go on the ride at a time. They can pair up 3 different ways, as shown below.

Al and Bev

Al and Carmen

Bev and Carmen

Derek decides to join the group. How many different ways can the 4 students pair up?

Answer: _____

Show your work or explain how you got your answer.



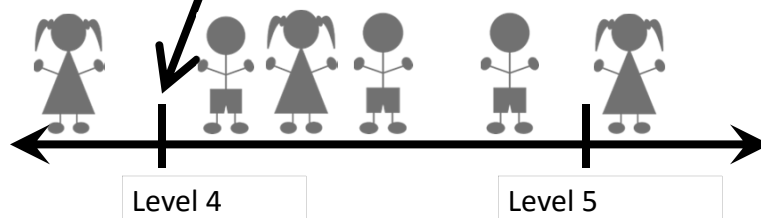
The Bookmark Process

- Study content in a book
- Tell a story about the content students in each level should be able to do
- Set a Bookmark



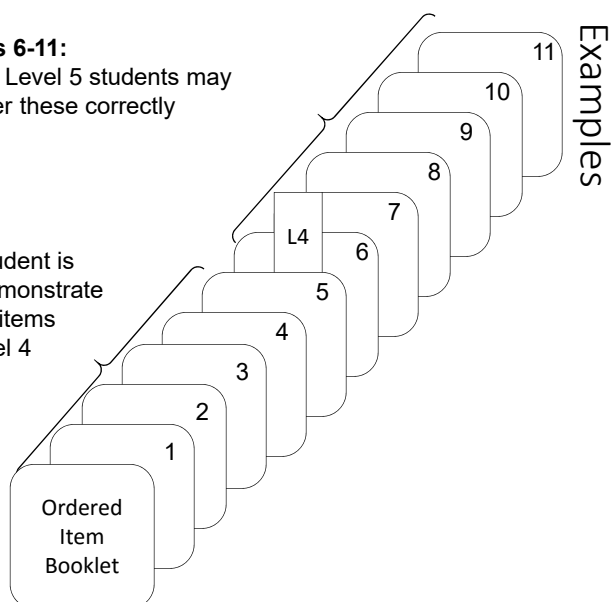
Target Student for Level 4

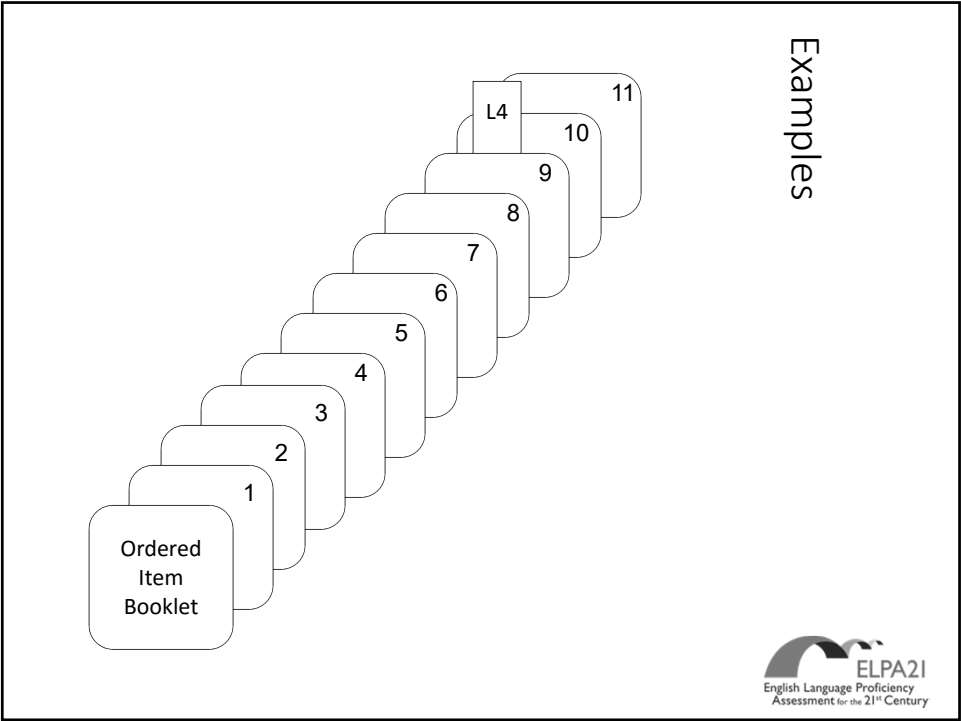
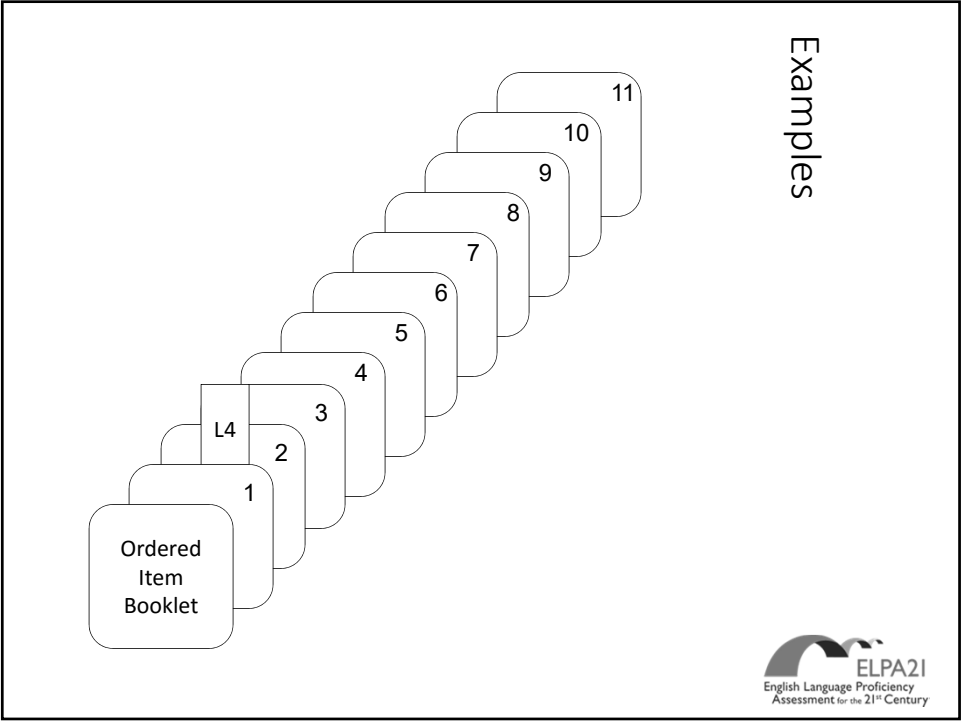
Describes the knowledge and skills of the student entering the Level 4 Achievement Level

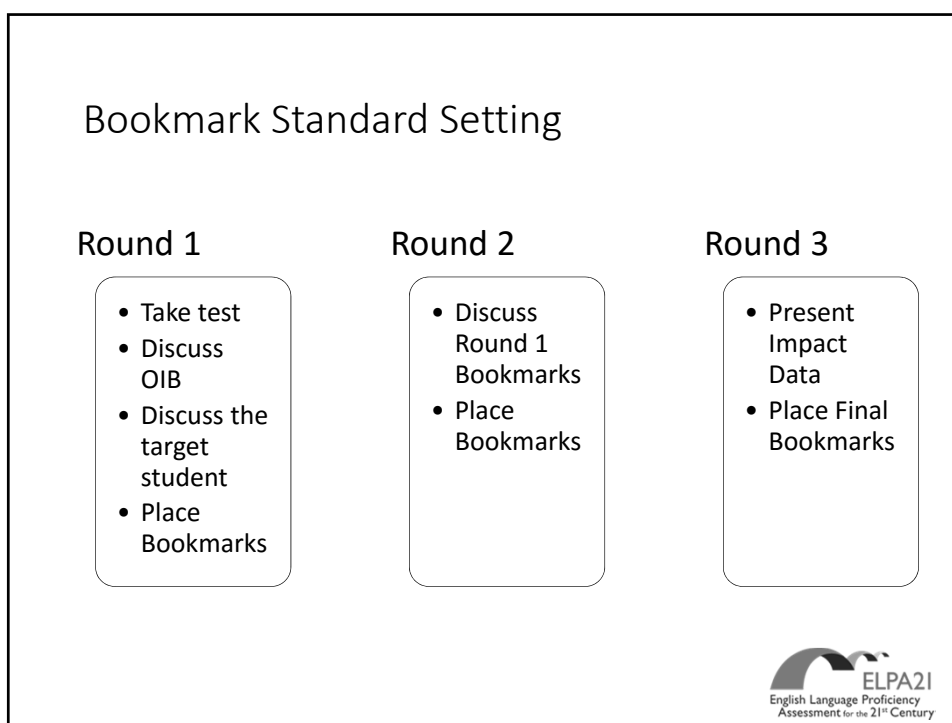
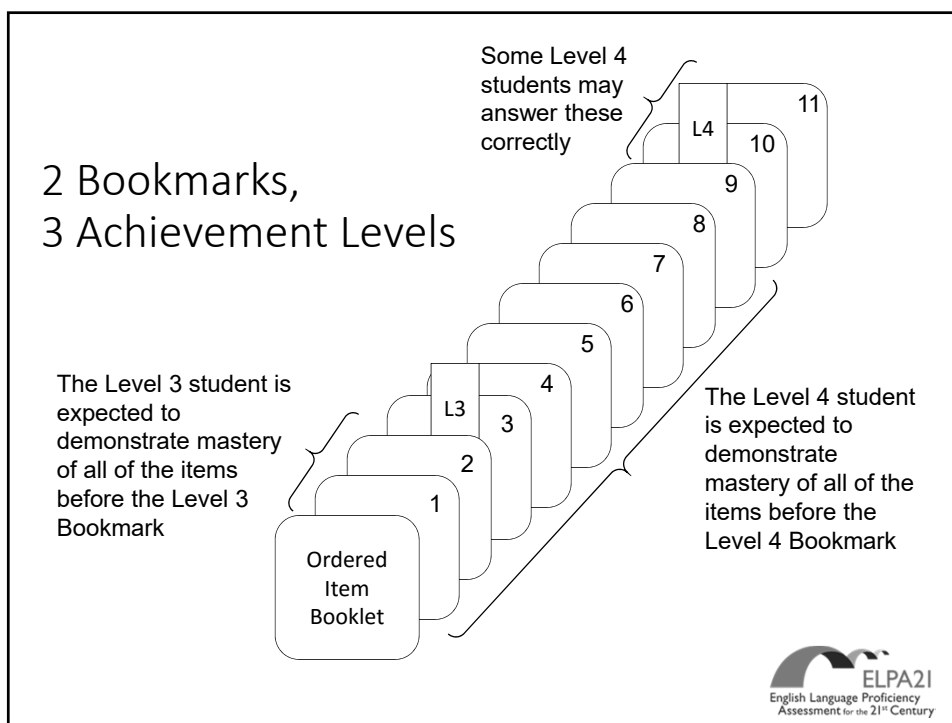


Pages 6-11:
Some Level 5 students may
answer these correctly

Pages 1-5:
The Level 4 student is
expected to demonstrate
mastery of the items
before the Level 4
bookmark.







Understanding Your Role as Table Leader



Table Leader Responsibilities

- Facilitate discussion at the table level
 - Studying the OIB
 - Leading table level discussions (Rounds 2 and 3)
- Bring important questions to the attention of the room facilitators
- Check understanding at table
- Keep an eye on time
- Monitoring and collection of secure materials at your table



Table Leader Responsibilities

- Facilitate discussion at the table level
 - Studying the OIB
 - Facilitate conversation around the two questions
 - 1. What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?
 - 2. What makes this item more difficult than preceding items?
 - Remind panelists to put thoughts into comments
 - Monitor time for this task
 - Discourage sidebar conversation
 - Try to get all panelists to participate in discussion



Table Leader Responsibilities

- Facilitate discussion at the table level
 - Following Round 1, announce the bookmark locations for each panelist at your table
 - Levels 3 & 4
 - Discuss the reasons for panelist differences one level at a time, beginning with Level 3.
 - Begin with the first bookmark, then the highest, to spur conversation.
 - When conversation is complete for Level 3, you should understand the rationale for the differences, and move on to the next level's bookmarks



Table Leader Responsibilities

- Monitor and collect secure materials at your table
 - At the end of each day we will collect materials
 - You will have a list of secure materials. Have panelists stack materials according to the list.
 - This allows us to have an efficient and orderly collection of materials to support security.



Agenda: Day 1

- Morning
 - Opening Session
 - Round 1 for Reading
- Afternoon
 - Bookmark training
 - Round 2 for Reading
 - Round 3 for Reading
 - Evaluation for Reading



Agenda: Day 2 & 3

- Repeat process for Listening, Writing, and Speaking



Agenda: Day 4

- Cross-Grade Work
 - Vertical Articulation
 - ALD Development
 - Overall Proficiency



ELPA21 Standard Setting Table Facilitator Training

Daniel Lewis, Pacific Metrics
Mary Seburn, ELPA21/Quantiful



Background

Contrasting Groups Study



Contrasting Groups Study

- EL teachers recruited by ELPA21 states complete a survey
 - Classify students into achievement levels on each domain and Overall using
 - policy descriptors
 - knowledge of their students' proficiency on the domains



Sample Results

	B	C	D	E	F	G	H	I
1	First	Last	Grade	Reading	Writing	Listening	Speaking	Overall
2			k	3	3	4	4	3
3			k	1	1	1	1	1
4			k	3	3	3	4	3
12			k	3	2	3	3	3
13			k	1	1	3	3	2
14			k	2	2	2	2	2
16			k	2	2	3	3	3
17			k	3	2	5	4	3
18			k	3	3	5	4	4
24			k	4	3	5	5	4
25			k	4	4	4	4	4



Use of CGS Results

- The CGS study will yield
 - an external frame of reference for the in-person panels to consider as feedback during standard setting
 - Match survey data to student test results
 - Conduct analysis to see range of scores associated with, say, Level 3 and Level 4
 - “See where Level 3 ends on scale and the other begins” to estimate CGS cut scores

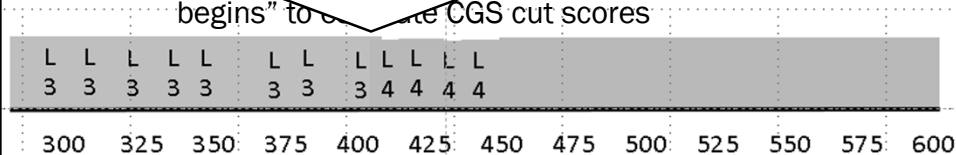




Table Facilitator Training

Review of Annotated Agenda

Daniel Lewis
Pacific Metrics





Standard Setting: High-Level Agenda



Note: Times are approximate and will be adjusted as needed.
Appropriate breaks will be provided throughout.

Tuesday, July 19: Day 1

Tuesday Morning	<p>7:30 AM: Table Facilitator Refresher in Cagglej Ballroom</p> <p>8:00 AM: Opening Session: Welcome and Training</p> <ul style="list-style-type: none"> ELPA21 welcomes panelists, provides background information associated with test, and highlights relevant policy information Dan Lewis provides overview of standard setting and describes relevance of Contrasting Groups Study (CGS) and EL Expertwork to date Karla Egan provides overview of Bookmark Standard Setting and prepares panelists for their first activities <p>9:30 AM: Readiness survey administered</p> <p>10:00 AM: Review the ELP Standards and Achievement Level Indicators (ALIs) for Reading</p> <p>10:15 AM: Table review of Reading online operational form</p> <p>10:45 AM: Study Reading ordered item booklet (OIB)</p> <ul style="list-style-type: none"> Review items one-by-one, together as a table, in order of difficulty <p>Noon: Lunch</p>
Tuesday Afternoon	<p>1:00 PM: Reading Round 1 (R1)</p> <ul style="list-style-type: none"> Discuss the Reading Target Student Descriptors, which describe the knowledge, skills, and processes (KS Ps) associated with students at the threshold of an achievement level Bookmark Training: How to express recommendations for the KS Ps needed for item hold Level 3 and Level 4 students Bookmarking Readiness Survey Panelists independently place R1 bookmarks <p>2:30 PM: Reading Round 2 (R2)</p> <ul style="list-style-type: none"> Review R1 results and CGS bookmarks at each table Table-level discussion: Differences in panelists' R1 ratings Panelists independently place R2 bookmarks <p>3:30 PM: Reading Round 3 (R3)</p> <ul style="list-style-type: none"> Presentation of Table 1, Table 2, and Overall R2 median ratings and bookmarks associated with CGS cut scores Presentation of impact data based on Overall R2 medians

1

Standard Setting: Table Facilitator Agenda

Note: Times are approximate and will be adjusted as needed. Breaks will be provided throughout.

Tuesday Morning, July 19th

Morning agenda and goals:



- Introductions
- Distribute and track secure materials
- Identify a volunteer scribe to take notes on the computer for your table
- EL Expert leads Q & A of the ELP standards and ALIs
- Support the review of one form of your grade band assessment online
- Study the ordered item booklet (OIB) at your table. This is intended to help panelists attain a deeper understanding of what the test measures

What you need:

- Panelist materials: assignment and tracking form
- Secure materials: packets (printed copies of ALIs, OIBs, item maps, and stimulus booklet)
- Laptop (provided by Pacific Metrics)

Schedule and Description of Activities and Roles

- 7:30 AM:** Table Facilitators meet with Workshop Facilitators in Cagglej Ballroom
- 8:00 AM:** Opening Session
- 9:30 AM:** Introductions and secure materials sign-out
 - A packet of secure materials will be provided for each panelist. Each secure piece has a Panelist ID number in the top right hand corner. The Table Facilitator for a grade band should assign secure packets so that they are consecutively numbered at your tables (e.g., Table 1 assigns sets 1-4 and Table 2 assigns sets 5-8). The Table Facilitators work together to complete the tracking form for your grade band. Have panelists sign and print their names on the tracking form by the ID number of the materials they receive. They should also put their names on each piece of secure material.
 - Introduce yourself and support brief introductions around the table. There will be many opportunities to get to know your fellow panelists so keep initial introductions brief (modeling this with your own brief introduction). One minute per panelist is a good target.
 - Ask for a volunteer to act as scribe at the table (this may be a rotating role). The scribe will take notes on the computer about items when studying the OIB. The scribe should sit next to you so that you can share the laptop controls (there is a large monitor at each table for panelists to view). You may choose to be the scribe but you have enough responsibility and it is probably better for someone else to take that on. There will be a modest amount of opening different files (all contained in a single folder on the desktop) so the scribe should have strong keyboarding skills and be able to work with you to move to the various files.

Facilitator Annotated Agenda



Standard Setting Table Facilitator Agenda



Note: Times are approximate and will be adjusted as needed. Breaks will be provided throughout.

Tuesday Morning, July 19th

Morning agenda and goals:

- Introductions
- Distribute and track secure materials
- Identify a volunteer scribe to take notes on the computer for your table
- EL Expert leads Q & A of the ELP standards and ALIs
- Support the review of one form of your grade band assessment online
- Study the ordered item booklet (OIB) at your table. This is intended to help panelists' attain a deeper understanding of what the test measures

What you need:

- Panelist materials assignment and tracking form
- Secure materials packets (printed copies of ALIs, OIBs, item maps, and stimulus booklets)
- Laptop (provided by Pacific Metrics)



Facilitator Annotated Agenda

Schedule and Description of Activities and Roles

- **7:30 AM:** Table Facilitators meet with Workshop Facilitators in Cupples Ballroom
- **8:00 AM:** Opening Session
- **9:30 AM:** Introductions and secure materials sign-out
 - A packet of secure materials will be provided for each panelist. Each secure piece has a Panelist ID number in the top right hand corner. The Table Facilitators for a grade band should assign secure packets so that they are consecutively numbered at your tables (e.g., Table 1 assigns sets 1-4 and Table 2 assigns sets 5-8). The Table Facilitators work together to complete the tracking form for your grade band. Have panelists sign and print their names on the tracking form by the ID number of the materials they receive. They should also put their names on each piece of secure material.
 - Introduce yourself and support brief introductions around the table. There will be many opportunities to get to know your fellow panelists so keep initial introductions brief (modeling this with your own brief introduction). One minute per panelist is a good target.
 - Ask for a volunteer to act as scribe at the table (this may be a rotating role). The scribe will take notes on the computer about items when studying the OIB. The scribe should sit next to you so that you can share the laptop controls (there is a large monitor at each table for panelists to view). You may choose to be the scribe but you have enough responsibility and it is probably better for someone else to take that on. There will be a modest amount of opening different files (all contained in a single folder on the desktop) so the scribe should have strong keyboarding skills and be able to work with you to move to the various files.

Sign
out



Facilitator Annotated Agenda

- **9:45 AM:** Q & A for the ELP Standards and ALIs for Reading 1. The Workshop Facilitators and EL Expert will direct panelists to the appropriate materials and open the floor up for questions. ELP standards and ALIs should be reviewed prior to the workshop and this Q & A is intended to answer any questions panelists may have.
- **10:00 AM:** Table-level review of Reading online operational form. The Workshop Facilitators will demonstrate how to access the online operational forms on your computer. This activity is intended to provide an opportunity to see a full form of the Reading assessment in its online format as a student sees it.
 - Follow the Workshop Facilitators prompts and let them know if you have challenges logging in. Move through each item as a student would; perhaps a bit quicker. You will have many opportunities to discuss these items so use this one to experience the assessment as a student and not as a means to prompt discussion.
 - Take a 10-minute break after completion.



Facilitator Annotated Agenda

- **10:30 AM:** OIB Review
 - Upon return from break, begin studying the OIB.
 - Panelists should have their OIBs, item maps, and stimulus booklets out.
 - Ask the scribe to open the item map so it appears on the monitor and take notes with input from the table.
 - Panelists turn to the first item in their OIBs and locate it on their item maps. Review the item:
 - Prompt with the first question: *What does this item measure? That is, what do you know about a student who responds successfully to this item?*
 - The scribe should be taking notes of summarizing the responses, which will be used during achievement level descriptor (ALD) writing. Notes that are informative, succinct responses will be most useful for creating valid ALDs to help teachers and parents understand students' KSPs based on their test results.
 - Table Facilitator asks the second question: *Why is this item more difficult than the preceding items?*
 - Scribe takes notes.
 - When discussion ceases to provide new information, move to the next question.
 - Repeat for all items in the OIB.

Item
Map

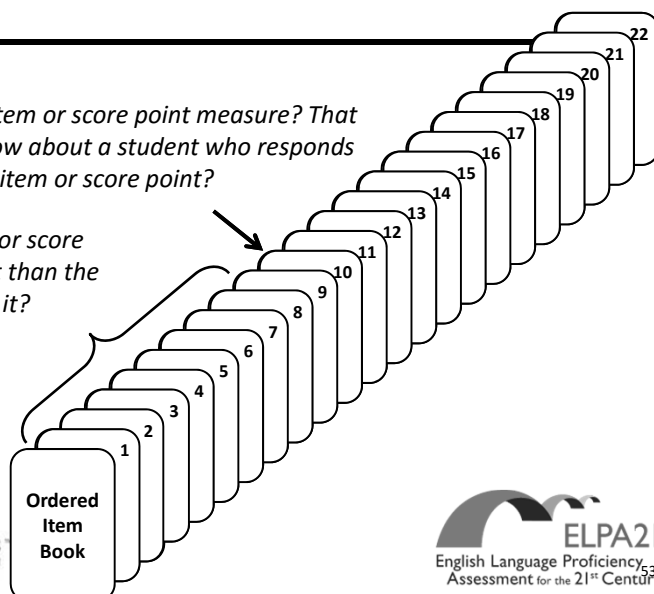


Facilitator Annotated Agenda

OIB Review

1. What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?

2. Why is this item or score point more difficult than the items that precede it?



PACIFIC METRICS™
CORPORATION



Facilitator Annotated Agenda

OIB Review

□ Considerations:

- Remember, you will see multiple-choice items only once in the OIB. You will see 1-point constructed response items only once in the OIB. You will see items with a maximum score higher than 1 multiple times, once for each non-zero score point. So, for example, an item scored from 0-2 appears two times.
 - The first time you see the item it is for score point 1 (e.g., out of 2). Ask and answer the question: What do you know about a student who scores a 1 on this 2-point item?
 - The second time you see the item it is for score point 2 (out of 2). Ask and answer the question: What do you know about a student who scores a 2 (a perfect score) on this 2-point item?

Excel File



PACIFIC METRICS™
CORPORATION



Facilitator Annotated Agenda

OIB Review

- Considerations:
 - Stimulus booklet. Longer stimuli that are common to multiple items are presented in a separate booklet to maintain the flow of the OIB. This booklet is ordered by Set Leader, which can be found on each associated page in the OIB. To view an item's stimulus, identify the Set Leader in the OIB and find that number in the table of contents of the stimulus booklet. Then find the page on which the stimulus appears. [Stimulus Book](#)
 - Partial credit item rubric: A paper copy of the simple rubrics used for the automated scoring of Technology Enabled items will be provided. The scoring rubric for these items is frequently intuitive. When it is not, please refer to the paper copy of the rubrics for these items. [Rubric](#)



Facilitator Annotated Agenda

OIB Review

- Considerations:
 - Monitor time. Encourage thoughtful discussion, and do not rush through items. Remember, you will be discussing the items in this activity and after Rounds 1 and 2 before your final round of ratings. You *do not* need to exhaustively dissect every item. It is important to the validity of the process that panelists have enough time and don't feel rushed. It is also important to complete the process in our 4-day workshop. These activities must be completed for all four domains in that time period.
 - Support equitable and diverse discussion by encouraging all panelists to contribute to the discussion.



Facilitator Annotated Agenda

OIB Review

- Caveats
 - Do not spend time critiquing items. While this is natural, this is not an item review workshop. If panelists feel the need to provide item input, have them write their concerns on an index card provided at the table for ELPA21 staff review.
- Noon: Lunch
 - All secure materials are color coded. Be sure all color-coded materials are left in the room. Panelists should never remove these materials from the room.



Facilitator Annotated Agenda

Afternoon agenda and Goals:

- Review Target Student Descriptors
- Bookmark Training: Receive training to make Bookmark recommendations
- Round 1: Panelists make their first cut score recommendations
- Round 2: Panelists discuss differences in cut score recommendations within their tables
- Round 3: Panelists review the reasonableness of impact data associated with the grade band median recommendation and discuss differences between the grade band Table 1 and Table 2 cut score recommendations
- View final results
- Distribute, complete, and collect Domain evaluations
- Support systematic collection of secure materials

What you need:

- OIBs, item maps, stimulus booklets
- Rating forms
- Excel file
- Readiness surveys
- Evaluations
- Secure materials collection checklist



Facilitator Annotated Agenda

Schedule and Description of Activities and Roles

- **1:00 PM:** Workshop Facilitator and EL Expert review Target Student Descriptors
- **1:15 PM:** Workshop Facilitator provides Bookmark training
- **1:40 PM:** Readiness surveys. You will pass out the Bookmark readiness surveys following training. If someone at your table marks “yes” to the last two questions, then raise your hand that additional training is needed for your table.
- **1:45 PM:** Round 1 ratings
 - ☐ Panelists make their ratings on their paper rating forms independently and without discussion. If there are questions, ask a Workshop Facilitator.
 - ☐ Collect panelists’ rating forms.
 - Review each panelist’s rating form to see that the Level 4 Bookmark is higher than the Level 3 Bookmark.
 - When all are collected open the Excel “Enter R1 Ratings” tab and enter all data in the orange highlighted cells for Level 3 and the green cells for Level 4. See Figure 1 for a screenshot of the “Enter R1 Ratings” tab.
 - Ask another panelist to review your entries for accuracy. When you have entered them and the entries have gone through QA, save the excel file.
 - ☐ Have panelists take a short break after providing you their completed rating forms.

[Readiness
Survey](#)

[Place
Bookmarks](#)

[Rating
Form](#)

[Excel File](#)



Facilitator Annotated Agenda

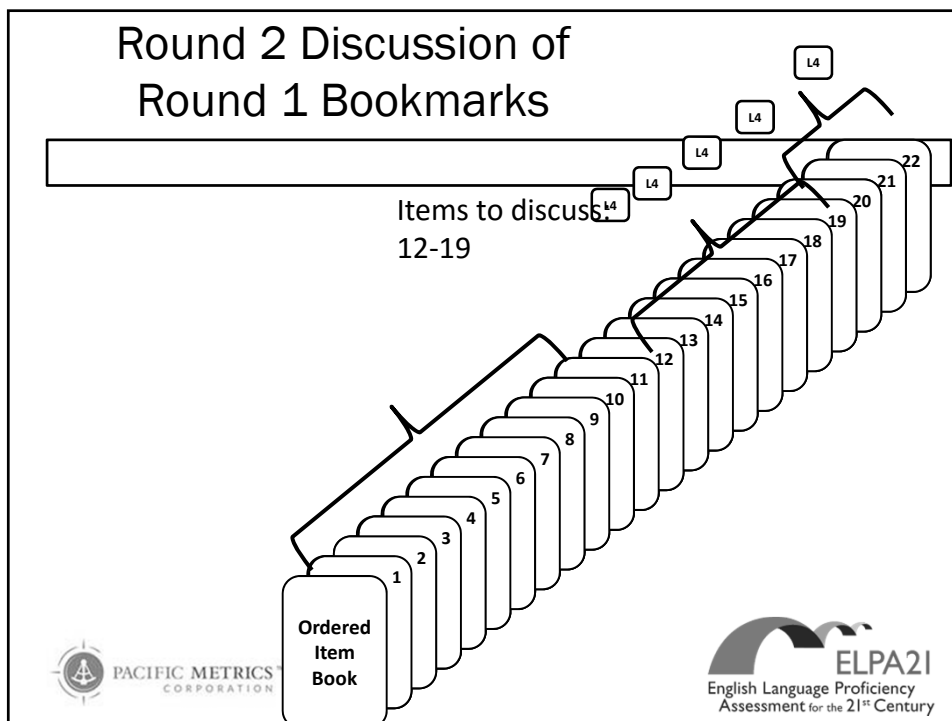
Round 2

- **2:30 PM:** Discussion of Round 1 results and Round 2 ratings
 - ☐ Open the Excel file to “R2 Display” tab and follow steps 1-9. See Figure 2 for a screenshot of the “R2 Display” tab.
 1. Each panelist places a green post-representing your table’s panelists Level 4 bookmarks. We start with Level 4, not Level 3, because Level 4 is the “anchor level.”

[Excel File](#)



Round 2 Discussion of Round 1 Bookmarks

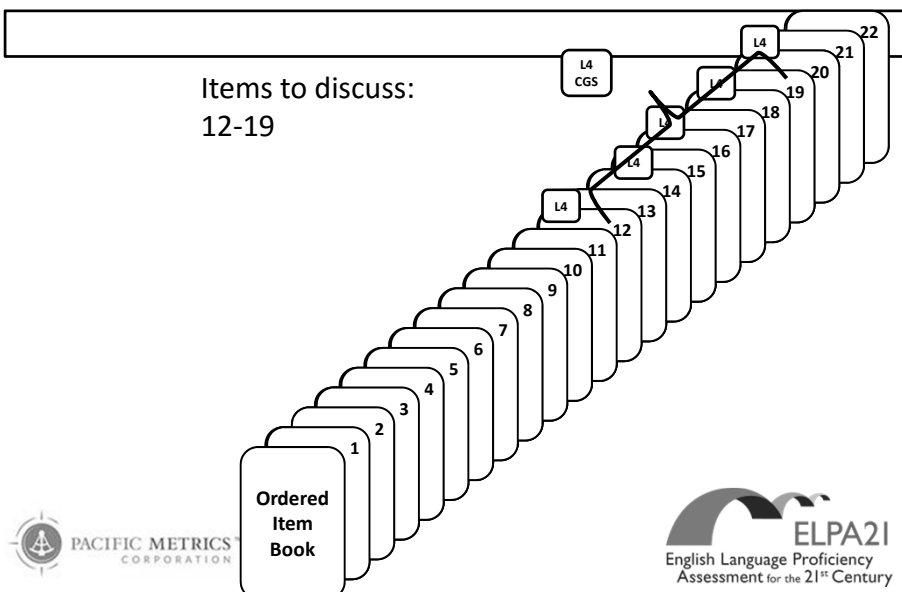


Facilitator Annotated Agenda Round 2

2. Panelists place a blue post-it bookmark representing the Level 4 Contrasting Groups Study (CGS) results.
 - Remind panelists of the CGS training from the opening session. The CGS results are another reference point, described here:
 - Contrasting Groups Study. The CGS is a study conducted by Pacific Metrics for ELPA21. EL teachers in ELPA21 states classified their EL students into 5 levels of achievement for each domain and overall based on the ELPA21 Achievement Level Policy Descriptors. Pacific Metrics analyzed this data and by associating teachers' classifications with the operational ELPA21 test results for these students we are able to estimate cut scores and bookmarks associated with the teachers' judgments.
 - The results of the CGS will be provided to panelists by translating the CGS cut score estimates into bookmarks in the OIBs so that panelists can see the CGS results expressed in the same metric used to express their own ratings.

Round 2 Discussion of Round 1 Bookmarks

Items to discuss:
12-19



PACIFIC METRICS
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Facilitator Annotated Agenda Round 2

3. Discussion of Round 1 Level 4 bookmarks. Ask panelists to discuss the reasons for their various bookmark placements. One way to do this is to begin by having the panelists with the lowest and highest bookmarks discuss their rationales for their bookmark placement.
 - Remind panelists that there are no wrong bookmark placements, just differences of opinion. Round 2 allows panelists to reflect on their own and others' bookmark placements.
4. Panelists record Round 2 ratings for Level 4 based on their updated perspective. [Rating Form](#)
5. Remove Level 4 bookmarks. Each panelist now places an orange post-it for each of the panelist's Round 1 Level 3 bookmarks.
6. Place a blue post-it representing the Level 3 CGS results.
7. Discuss Round 1 Level 3 differences.
8. Record Round 2 ratings for Level 3.
9. Collect the rating forms, enter them into the Excel "R2 Ratings" tab, and QA [Excel File](#) them.



PACIFIC METRICS
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Facilitator Annotated Agenda Round 3

3:30 PM (Time tentative, depending on R2 duration. May carry over to the next day if necessary.): Discussion of Round 2 results and Round 3 ratings

- ☐ The two tables join together as one group representing the grade band. You no longer work as separate, independent tables.
- ☐ The Workshop Facilitator will introduce Round 3 activities

	A	B	C	D	E	F	G
1				Data based on R2 Bookmarks		Level 3 Bookmark	Level 4 Bookmark
2				Table 1 Median		4	16
3				Table 2 Median		4	14
4				Grade Band (T1 & T2) Median		4	16
5				CGS		?	13
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							

	A	B	C	D	E	F	G
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							

Facilitator Annotated Agenda Round 3

- Open the Excel "R3 Display" tab when instructed by the Workshop Facilitator. Figure 3 provides a screenshot of the Excel "R3 Display" tab. [Excel File](#)
- Table Facilitator Visual Check. We have linked the Table 1 and Table 2 Excel files. Both grade band Table Facilitators should check to see that:
 - ☐ the same data is displayed in each of their Excel files,
 - ☐ that the bookmarks for their respective tables are correctly represented (you can look to the "Enter R3 Ratings" tab to check your own table's bookmarks
 - ☐ If the Table Facilitator Visual Check indicates correct data, then proceed. If there are any discrepancies call a Workshop Facilitator to update your spreadsheet.
- Observe the impact data. If you are very surprised by the data, call a Workshop Facilitator to your table to help frame the discussion.
- As a single group (Tables 1 and 2) discuss the differences between the two tables' median bookmarks and the CGS bookmarks.
 - ☐ Begin with Level 4
 - Discuss differences between each table's Level 4 Bookmarks.
 - Panelists at both tables record their Round 3 Level 4 ratings on the rating form.
 - ☐ Repeat for Level 3
- Collect the rating forms and enter them in the Excel "R3 Ratings" tab



Facilitator Annotated Agenda

Round 3

- **4:45 PM:** Session close. Reading Evaluations and collection of secure materials. Sessions may run longer if we are near completion of an important task.
 - ☐ Workshop Facilitators will initiate close of the daily session. [Reading](#)
 - ☐ Panelists complete Reading evaluations [Evals](#)
 - ☐ Table Facilitators collect evaluations and submit to Workshop Facilitators
 - ☐ Table Facilitators follow systematic secure materials collection as described by Workshop Facilitators [Secure](#) [Materials](#)
- **5:15 PM:** Daily Debrief
 - ☐ The Table Facilitators are asked to attend the first part of the daily debrief in order to provide feedback:
 - Challenges that should be shared
 - What went well and what could be improved
 - Share any useful information associated with the quality of the workshop, table dynamics, etc.



Facilitator Annotated Agenda

Note that we have provided detailed directions for one domain. We will begin with Reading and when have completed all three rounds for Reading we will proceed and repeat this process for Listening, Writing, and Speaking. The only difference between domains will be the nature of the content and that Writing and Speaking have rubrics and examples of student work to help understand the scoring of some items (written work for writing and mp3 sound files for Speaking). We will provide training on the review of rubrics and examples of student work for Writing and Speaking at the workshop.



Follow up

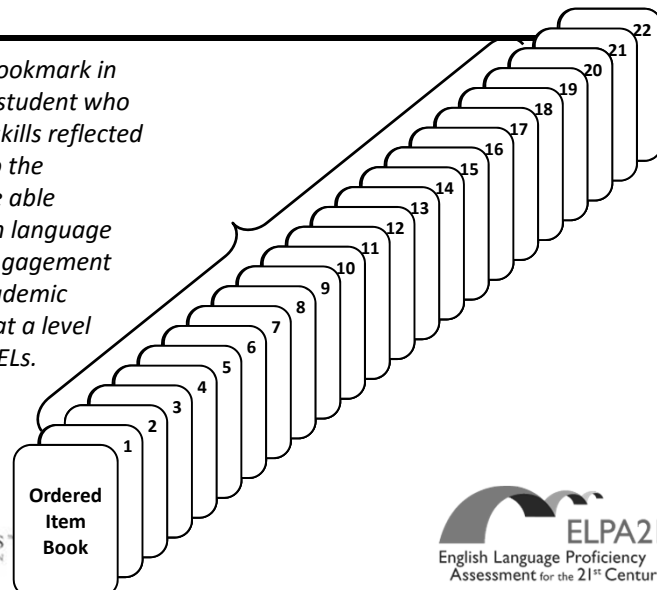
- I will send a copy of the high-level agenda and the annotated Facilitators Agenda (that we just reviewed) by tomorrow AM
- Would you like a Q & A session on Friday?
- Concerns?
- Questions now?



Panelists make cut score recommendations

L4

Place your Level 4 bookmark in the OIB such that a student who has mastery of the skills reflected by the items prior to the bookmark should be able demonstrate English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.



Appendix D2. Opening Session Slides

ELPA 21 Standard Setting

Standard Setting Overview



Policy Level Descriptors

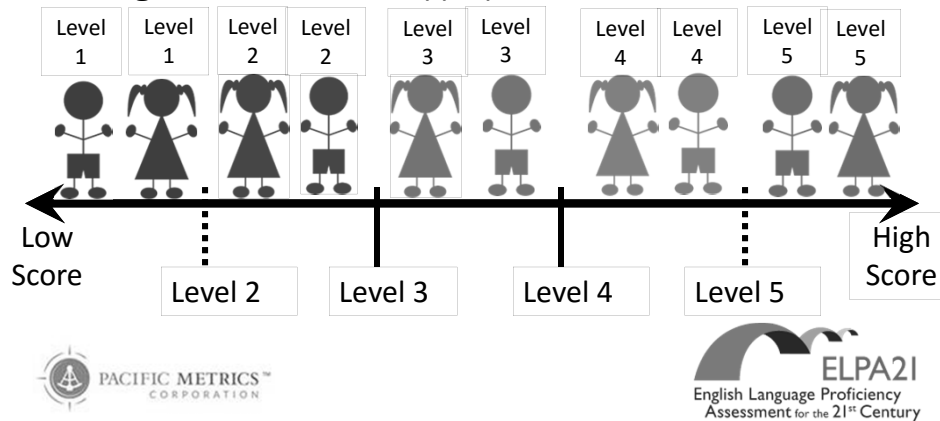
- High-level definition of type of performance expected at each proficiency level

A STUDENT AT THIS LEVEL...	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of developing grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits superior English language skills, as measured by ELPA21.



Purpose of the Standard Setting

Standard setting is a process that lets experts make judgments that result in scores on the test scale used to categorize students into appropriate levels of achievement.



What is Bookmark Standard Setting?

- Bookmark Standard Setting sets up a partnership between SMEs (you) and testing experts (us)
 - SMEs indicate the KSPs students should have to be in each achievement level
 - We (testing experts) use standardized methods, in this case, the Bookmark method, to identify cut scores on the test scale such that students that meet or exceed those cut scores tend to demonstrate those KSPs and students below those cut scores do not.

English Language Proficiency Standards

English Language Proficiency Standards

with Correspondences to K-12 English Language Arts (ELA), Mathematics, and Science Practices, K-12 ELA Standards, and 6-12 Literacy Standards

English Language Proficiency (ELP) Standards

with Correspondences to K-12 English Language Arts (ELA), Mathematics, and Science Practices, K-12 ELA Standards, and 6-12 Literacy Standards

English Language Proficiency (ELP) Standards

with Correspondences to K-12 English Language Arts (ELA), Mathematics, and Science Practices, K-12 ELA Standards, and 6-12 Literacy Standards

- [

d by ELLs
idemic

with
c tasks
and

like at each
ependent
ities?



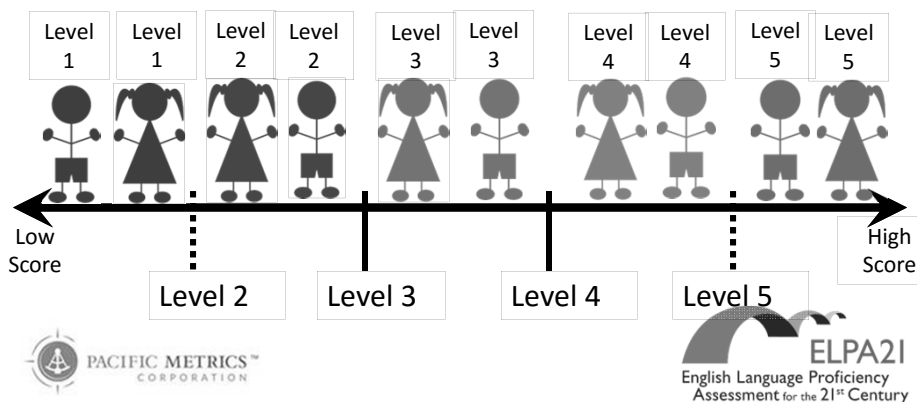
Achievement Level Indicators (ALIs)

Reading—ELPA21 Preliminary Achievement Level Indicators—Grade 1				
1	2	3	4	5
<p>The learner can:</p> <ul style="list-style-type: none"> • (1) identify a few key words in read-alongs • (2) respond to simple yes/no and some wh- questions about familiar topics • (8) recognize the meaning of a few frequently occurring words and phrases read-alongs about familiar topics, experiences, or events 	<p>The learner can:</p> <ul style="list-style-type: none"> • (1) identify key words and phrases in read-alongs and simple texts • (2) respond to simple yes/no and wh- questions about familiar topics • (6) identify a reason an author gives to support the main point • (8) answer simple questions to help determine the meaning of frequently occurring words and phrases in read-alongs about familiar topics, experiences, or events 	<p>The learner can:</p> <ul style="list-style-type: none"> • (1) identify main topics, answer questions about key details, and retell some key details or events from read-alongs and simple texts • (2) answer simple questions about familiar topics • (6) identify one or two reasons an author gives to support the main point • (8) answer questions to help determine the meaning of some less frequently occurring words and phrases in simple texts and read-alongs about familiar topics, experiences, or events 	<p>The learner can:</p> <ul style="list-style-type: none"> • (1) identify main topics and ask and answer questions about an increasing number of key details in written texts and read-alongs; retell familiar stories or episodes of stories in written texts and read-alongs • (2) answer questions about a variety of topics and texts • (6) identify reasons an author gives to support the main point • (8) answer questions to help determine the meaning of less common words, phrases, and simple idiomatic expressions in written texts about a variety of topics, experiences, or events 	<p>The learner can:</p> <ul style="list-style-type: none"> • (1) identify main topics and ask and answer questions about key details in written texts and read-alongs; retell stories in written text and read-alongs, including key details • (2) answer questions about a variety of topics and texts • (6) identify reasons an author gives to support the main point • (8) answer questions to help determine or clarify the meaning of words, phrases, and idiomatic expressions in written texts about a variety of topics, experiences, or events



ELPA21 Levels and Cut Scores

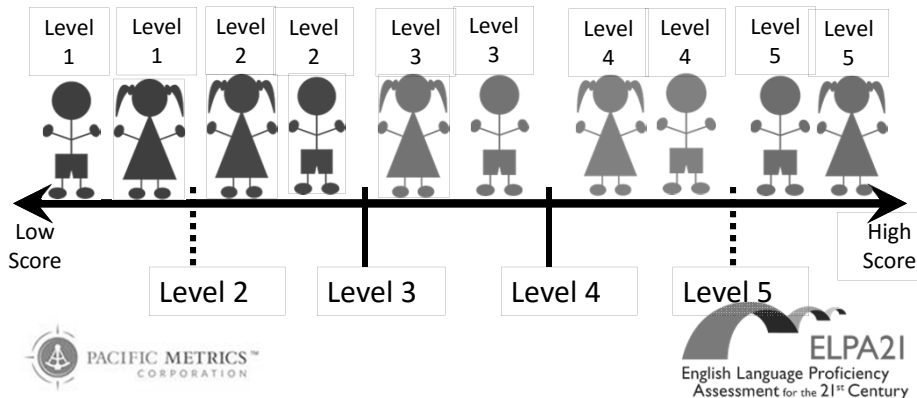
- Four cut scores: Level 2, Level 3, Level 4, Level 5
- Five achievement levels: Level 1, Level 2, Level 3, Level 4, Level 5



ELPA21 Levels and Cut Scores

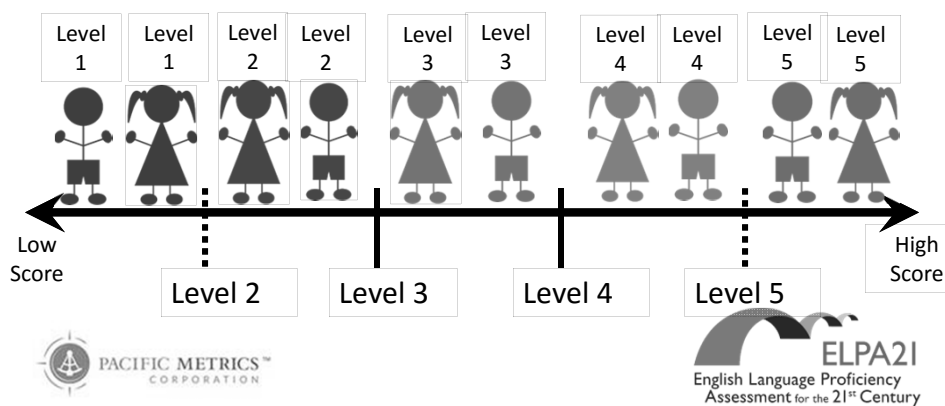
- You will recommend: Levels 3 and 4 directly

Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the engagement with grade-level academic content; instruction at a level comparable to non-ELs.



ELPA21 Levels and Cut Scores

- Levels 2 and 5 will be set, initially, using Contrasting Groups Study data (explained next)
 - You will have an opportunity to review, and modify these cut scores



Contrasting Groups Study

- EL teachers recruited by ELPA21 states completed a survey
 - They classified students into achievement levels for each domain and overall based on
 - policy descriptors
 - knowledge of their students' proficiency on the domains
 - Broad input from a large group of EL teachers
- EL teachers recruited by ELPA21 states
 - Over 200 teachers participated
 - Over 6,500 students classified



Contrasting Groups Study

Student Information				Assessment Results																								
First Name	Last Name	State ID	Grade	Reading					Writing					Listening					Speaking					Overall				
				level 1	level 2	level 3	level 4	level 5	level 1	level 2	level 3	level 4	level 5	level 1	level 2	level 3	level 4	level 5	level 1	level 2	level 3	level 4	level 5	level 1	level 2	level 3	level 4	level 5
																					<							



PACIFIC METRICS™
CORPORATION



Sample Results

	B	C	D	E	F	G	H	I
1	First	Last	Grade	Reading	Writing	Listening	Speaking	Overall
2			k	3	3	4	4	3
3			k	1	1	1	1	1
4			k	3	3	3	4	3
12			k	3	2	3	3	3
13			k	1	1	3	3	2
14			k	2	2	2	2	2
16			k	2	2	3	3	3
17			k	3	2	5	4	3
18			k	3	3	5	4	4
24			k	4	3	5	5	4
25			k	4	4	4	4	4



PACIFIC METRICS™
CORPORATION

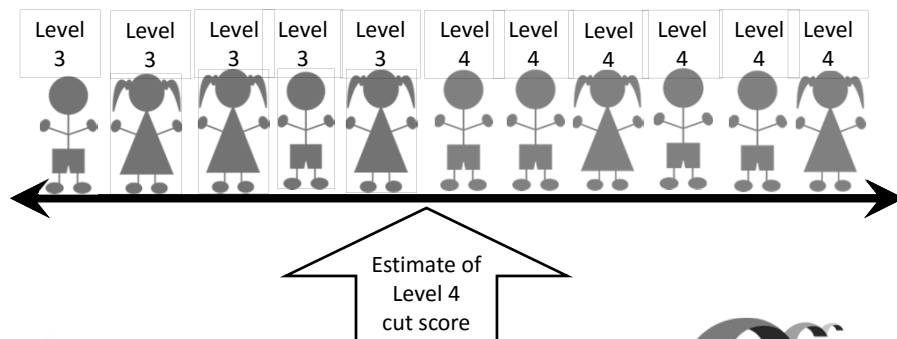


Use of CGS Results

- The CGS study yielded
 - an estimate of cut scores from a second standard setting method
 - Input from fellow teachers in the field
 - Analysis
 - Match survey data to student test results
 - Conduct analysis to see range of scores associated with, say Level 3 and Level 4
 - “See where one level ends on scale and the other begins” to estimate CGS cut scores



CGS “Estimate of Cut Score”

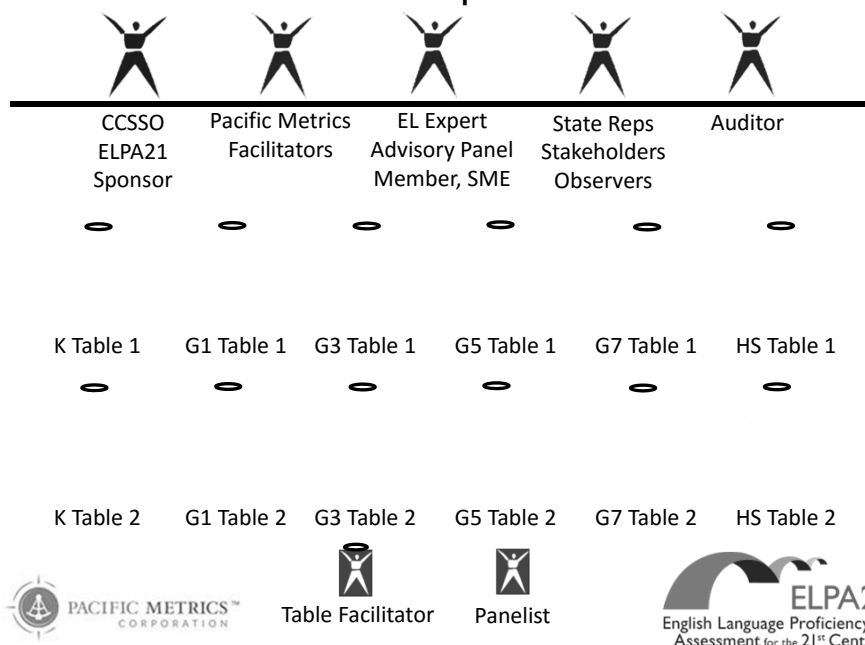


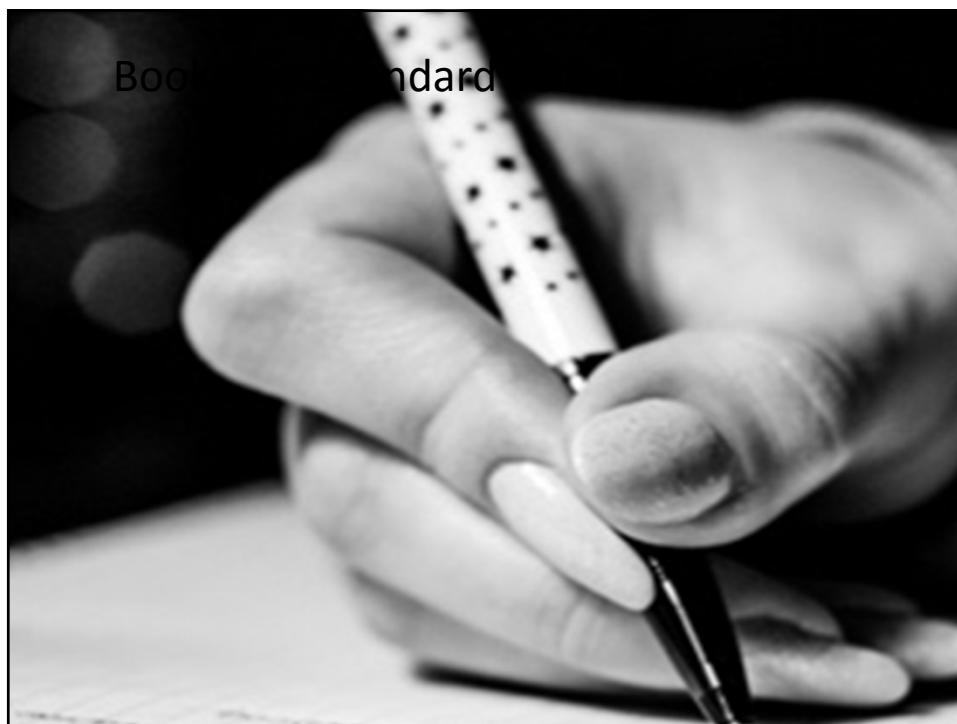
Developing Expertise

- You have expertise:
 - of the ALIs and ELP standards
 - working with ELLs
- You have knowledge the ELPA21 assessments
- Your work requires you all to have *advanced* knowledge of the ELPA21 assessments
 - You will participate in the following activities to support that requirement
 1. Study ELP standards and ALIs
 2. Study test items
 - Take an online form of the test as students do
 - Study and discuss items in an ordered item booklets
 3. Study target student descriptors



Workshop Roles

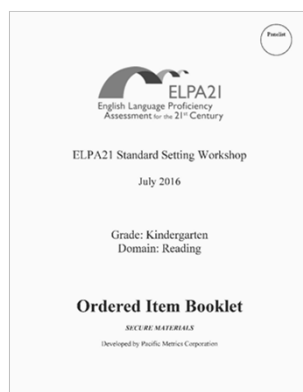




Bookmark Standard Setting Materials

Ordered Item Booklets

Item Maps



Item ID	Score	Domain	Task Type	Item Description
1	100	Reading	Word	Identify the word that is not a real word.
2	100	Reading	Word	Identify the word that is not a real word.
3	100	Reading	Word	Identify the word that is not a real word.
4	100	Reading	Word	Identify the word that is not a real word.
5	100	Reading	Word	Identify the word that is not a real word.
6	100	Reading	Word	Identify the word that is not a real word.
7	100	Reading	Word	Identify the word that is not a real word.
8	100	Reading	Word	Identify the word that is not a real word.
9	100	Reading	Word	Identify the word that is not a real word.
10	100	Reading	Word	Identify the word that is not a real word.
11	100	Reading	Word	Identify the word that is not a real word.
12	100	Reading	Word	Identify the word that is not a real word.
13	100	Reading	Word	Identify the word that is not a real word.
14	100	Reading	Word	Identify the word that is not a real word.
15	100	Reading	Word	Identify the word that is not a real word.
16	100	Reading	Word	Identify the word that is not a real word.
17	100	Reading	Word	Identify the word that is not a real word.
18	100	Reading	Word	Identify the word that is not a real word.

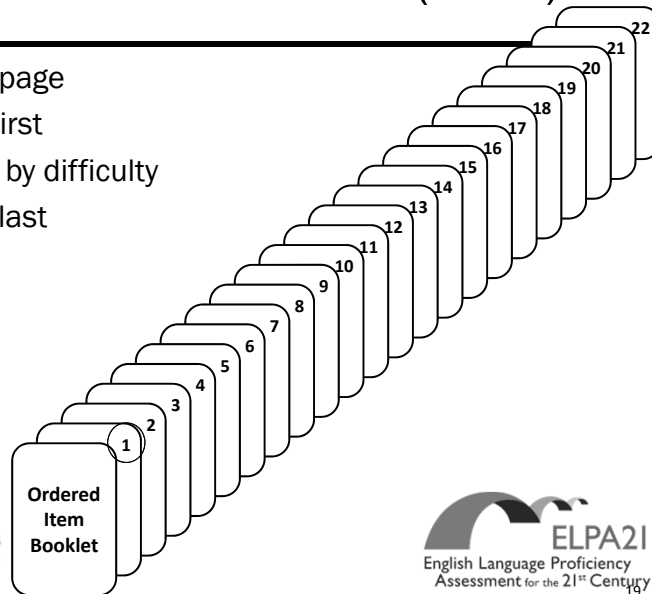


PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Ordered Item Booklets (OIBs)

- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last




PACIFIC METRICS™
CORPORATION



Item Map: Key Questions


- (1) What does this item measure?
- (2) Why is this item more difficult than the items that precede it?



FLORIDA
Department of Education
fldoe.org

ELPA21 Standard Setting July 2016

Grade 3 Reading Item Map



Order of Items by Subtest Number	Location Section	Item Number Item ID	Item Location Section Item ID	Task Type	Task Sub Type	What does this item or score point measure? * Task ID: What do you know about a student who responds correctly to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	301	1	VI019075_1274	Read-Using Sentence			
2	405	1	VI019206_1285	Short Comprehension			
3	240	1	VI019046_1218	Read and Match	Word		
4	330	1	VI022087_1291	Short Comprehension			
5	344	1	VI019975_1220	Read for Details			
6	347	2	VI019975_1220	Read for Details			
7	351	3	VI019975_1220	Read for Details			
8	352	1	VI019963_1220	Short Comprehension			
9	356	4	VI019975_1220	Read for Details			
10	373	1	VI019975_1220	Read for Details			

Developed by Florida Metric Corporation

SECURE MATERIALS

Page 1 of 2



PACIFIC METRICS™
CORPORATION

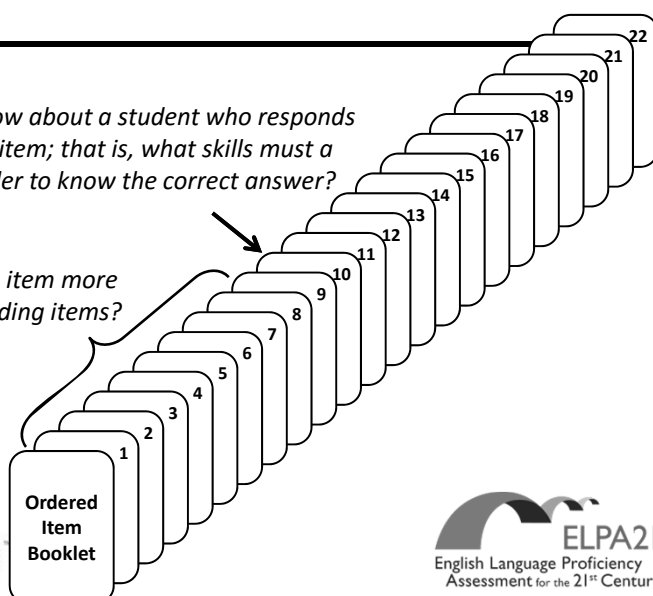


Study the Items in the OIB

As you study each item in the OIB, discuss two questions with your fellow panelists:

1. What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?

2. What makes this item more difficult than preceding items?



PACIFIC METRICS
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century



ELPA21 Standard Setting July 2016
Grade 3 Reading Item Map

Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID_AIR ITS Item ID	Set Leader_AIR Set Leader ID	Task Type	Task Sub Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	381	1	VH195927_1254		Read-Along Sentence			
2	440	1 of 2	VH203001_1174	VH201024_112	Procedural Text			
3	453	1	VH180306_1182		Read and Match	Sentence		
4	481	2 of 2	VH203001_1174	VH201024_112	Procedural Text			
5	485	1	VH182026_1288	VH182011_116	Short Correspondence			
6	510	1	VH180546_1218		Read and Match	Word		
7	539	1	VH200887_1292	VH182011_116	Short Correspondence			
8	552	1	VH199563_1290	VH182011_116	Short Correspondence			

Developed by Pacific Metrics Corporation

SECURE MATERIALS

Page 1 of 2



PACIFIC METRICS
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

1. What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?
2. What makes this item more difficult than preceding items?



PACIFIC METRICS™
CORPORATION


1A
1

Item Information			
Grade level/band:	2-3	Domain:	Reading
Item ID:	VH105027	Stimulus ID:	
Stimulus Name:			
ELP Standard(s):	1. Construct meaning from oral presentations and literary and informational text through grade-appropriate listening, reading, and viewing. 8. Determine the meaning of words and phrases in oral presentations and literary and informational text.		
Sub-Claim(s):	1R: The English language learner constructs meaning from literary and informational text. 8R: The English language learner determines the meaning of words and phrases in literary and informational text.		
CCSS/NGSS Practice(s):	L.16: Use English structures to communicate context-specific messages.		
Targeted Proficiency Level:		Text Complexity:	
Item Type:	TE	Linguistic Complexity:	
		Correct Answer:	3

Listen to the teacher and read along.

N: Listen to the teacher and read along.

1B



(Woman):: Put the yellow tulip in the vase.
Narrator: Choose the word "yellow" in the sentence.

Put the yellow tulip in the vase.

ELPA21
English Language Proficiency
Assessment for the 21st Century

Score Point 1 of 2 **2**

Item Information			
Grade level/band:	2-3	Domain:	Reading
Item ID:	VH203001	Stimulus ID:	VH201024
Stimulus Name:			
ELP Standard(s):	1. Construct meaning from oral presentations and literary and informational text through grade-appropriate listening, reading, and viewing. 8. Determine the meaning of words and phrases in oral presentations and literary and informational text.		
Sub-Claim(s):	1R: The English language learner constructs meaning from literary and informational text. 8R: The English language learner determines the meaning of words and phrases in literary and informational text.		
CCSS/NGSS Practice(s):	EP4: Build and present knowledge through research by integrating, comparing, and synthesizing ideas from texts.		
Targeted Proficiency Level:		Text Complexity:	
Item Type:	TE	Linguistic Complexity:	
		Correct Answer:	1

What **three** things do you need to make a paper-bag puppet?



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Grade 3 Reading Rubric

Grade	Item ID	Rubric
3	VH199073_1220	1 point for each correct match
3	VH199343_1224	1 point for each correct match
3	VH203001_1174	1 point for 2 correct selections, 2 points for 3
3	VH203017_1176	1 point for all 3 correct

You will have rubrics for constructed response items



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

3

Item Information			
Grade level/band:	2-3	Domain:	Reading
Item ID:	VH180305	Stimulus ID:	
Stimulus Name:			
ELP Standard(s):	1. Construct meaning from oral presentations and literary and informational text through grade-appropriate listening, reading, and viewing. 8. Determine the meaning of words and phrases in oral presentations and literary and informational text.		
Sub-Claim(s):	1R: The English language learner constructs meaning from literary and informational text. 8R: The English language learner determines the meaning of words and phrases in literary and informational text.		
CCSS/NGSS Practice(s):	L.P. Use English structures to communicate context-specific messages.		
Targeted Proficiency Level:	Text	Linguistic Complexity:	
Item Type:	SR	Correct Answer:	8

Look at the picture. Choose the sentence that matches the picture.

N: Look at the picture. Choose the sentence that matches the picture.



- A. The children are reading books.
- B. The children are wearing backpacks.
- C. The children are putting books in their backpacks.




PACIFIC METRICS™
CORPORATION


ELPA21
English Language Proficiency
Assessment for the 21st Century


Score Point 2 of 2 **4**

Item Information			
Grade level/band:	2-3	Domain:	Reading
Item ID:	VH203001	Stimulus ID:	VH201024
ELP Standard(s):	1. Construct meaning from oral presentations and literary and informational text through grade-appropriate listening, reading, and viewing; 8. Determine the meaning of words and phrases in oral presentations and literary and informational text.		
Sub-Claim(s):	1R The English language learner constructs meaning from literary and informational text; 8R The English language learner determines the meaning of words and phrases in literary and informational text.		
CCSS/NGSS Practice(s):	EP4. Build and present knowledge through research by integrating, comparing, and synthesizing ideas from texts.		
Targeted Proficiency Level:		Text Complexity:	
Item Type:	TE	Linguistic Complexity:	
		Correct Answer:	1

What **three** things do you need to make a paper-bag puppet?




PACIFIC METRICS™
CORPORATION

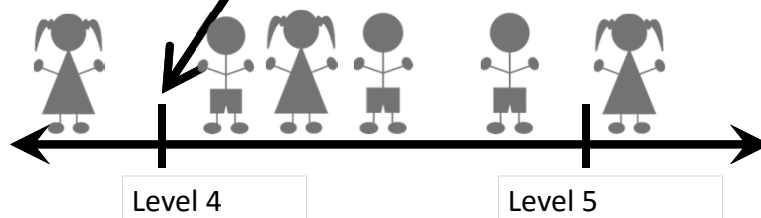

ELPA21
English Language Proficiency
Assessment for the 21st Century

The Bookmark Process

- Study content in a book
- Write a story about the content
- Set a Bookmark
 - Separate the content students should know to be in Level 4 from the content that is more than enough

Target Student for Level 4

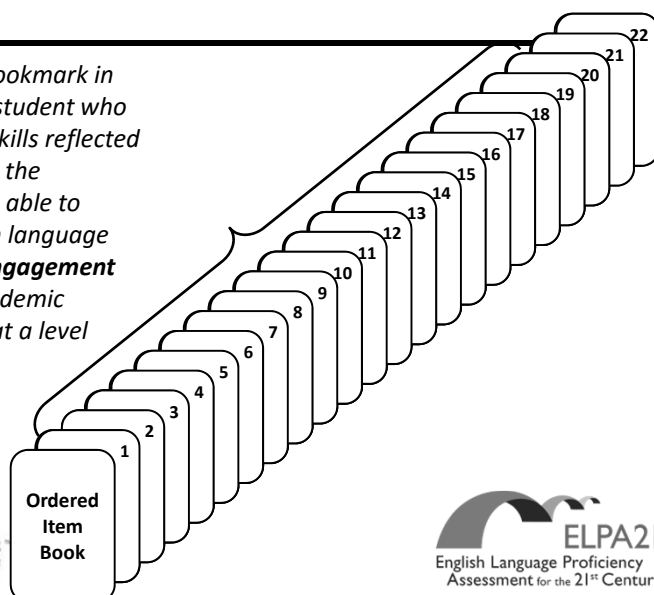
Describes the knowledge and skills of the student entering the Level 4 Achievement Level



Panelists make cut score recommendations

L4

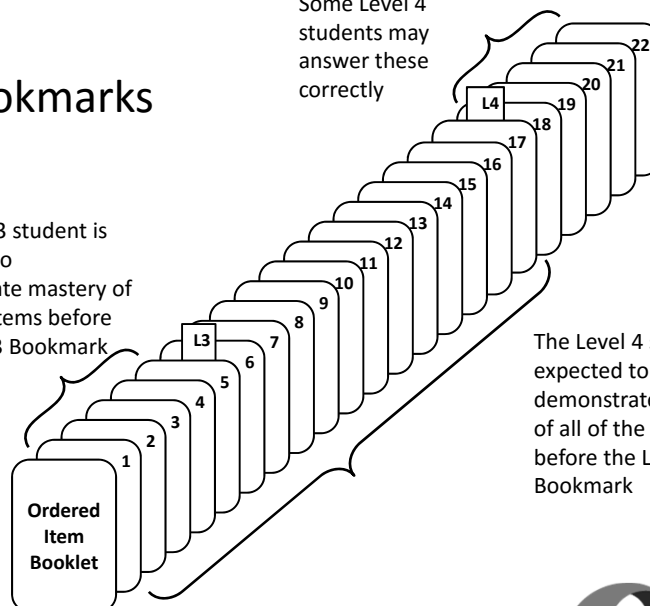
*Place your Level 4 bookmark in the OIB such that a student who has mastery of the skills reflected by the items prior to the bookmark should be able to demonstrate English language skills **required for engagement** with grade-level academic content instruction at a level comparable to non-ELs..*



2 Bookmarks

The Level 3 student is expected to demonstrate mastery of all of the items before the Level 3 Bookmark

Some Level 4 students may answer these correctly



The Level 4 student is expected to demonstrate mastery of all of the items before the Level 4 Bookmark



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Bookmark Standard Setting

Develop
expertise

- Take test
- Discuss OIB

Round 1

- Discuss the target student
- Place Bookmarks

Round 2

- Discuss Round 1 Bookmarks
- Place Bookmarks

Round 3

- Present Impact Data
- Place Final Bookmarks



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Agenda: Day 1

- Morning
 - Opening Session
 - Round 1 for Reading
- Afternoon
 - Bookmark training
 - Round 2 for Reading
 - Round 3 for Reading
 - Evaluation for Reading



PACIFIC METRICS™
CORPORATION



Agenda: Day 2 & 3

- Repeat process for Listening, Writing, and Speaking



PACIFIC METRICS™
CORPORATION



Agenda: Day 4

- Cross-Grade Work
 - Vertical Articulation
 - ALD Development
 - Proficiency Determination



Questions?



Readiness Survey



ELPA21 Standard Setting Post-Opening Session Readiness Survey

Please consider the statements below and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble only one of the four options for each statement.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
1. The orientation session provided a clear overview of the standard setting process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I understand the goals of the standard setting workshop.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I understand my role in the standard setting workshop.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. The orientation session provided a clear overview of the Contrasting Groups Study.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I understand how the Contrasting Groups Study data will be used in the standard setting process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. The orientation session provided a clear explanation of the development of ELPA21.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I understand how the results of the standard setting will be used to support the reporting of ELPA21 results.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I understand how to study the items in the ordered item booklet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you answered Disagree or Strongly Disagree to any of questions 1-8, then please answer the next two Yes/No questions.

	No (1)	Yes (2)
9. I would like additional training on studying the ordered item booklet.	<input type="radio"/>	<input type="radio"/>
10. I have additional questions on material presented during the opening session that I would like answered before I begin the next task.	<input type="radio"/>	<input type="radio"/>



Sign Out Materials



ELPA21 Standard Setting Secure Materials Sign-out/Collection List Kindergarten

Panelist	Table	Name	Reading	Listening	Writing	Speaking
1	1					
2	1					
3	1					
4	1					
5	1					
6	2					
7	2					
8	2					
9	2					
10	2					
11						
12						
13						

Table 1: Packets 1 – 5
Table 2: Packets 6 – 10



Sign Out Materials

Ordered Item Booklets



Item Maps

Karla Egan Profile

ELPA21 Standard Setting Workshop
Kindergarten Reading Item Map

Item ID	Item Name	Item Type	Item Description	Item Status	Item Score	Item Weight	Item Score Weight	Item Score Weight
1	001	1	001	001	001	001	001	001
2	002	1	002	002	002	002	002	002
3	003	1	003	003	003	003	003	003
4	004	1	004	004	004	004	004	004
5	005	1	005	005	005	005	005	005
6	006	1	006	006	006	006	006	006
7	007	1	007	007	007	007	007	007
8	008	1	008	008	008	008	008	008
9	009	1	009	009	009	009	009	009
10	010	1	010	010	010	010	010	010
11	011	1	011	011	011	011	011	011
12	012	1	012	012	012	012	012	012

Developed by Pacific Metrics Corporation



PACIFIC METRICS™
CORPORATION



Table Configuration

- Use headphones when using item viewer
- Feel free to rearrange equipment & materials
- Get comfortable!



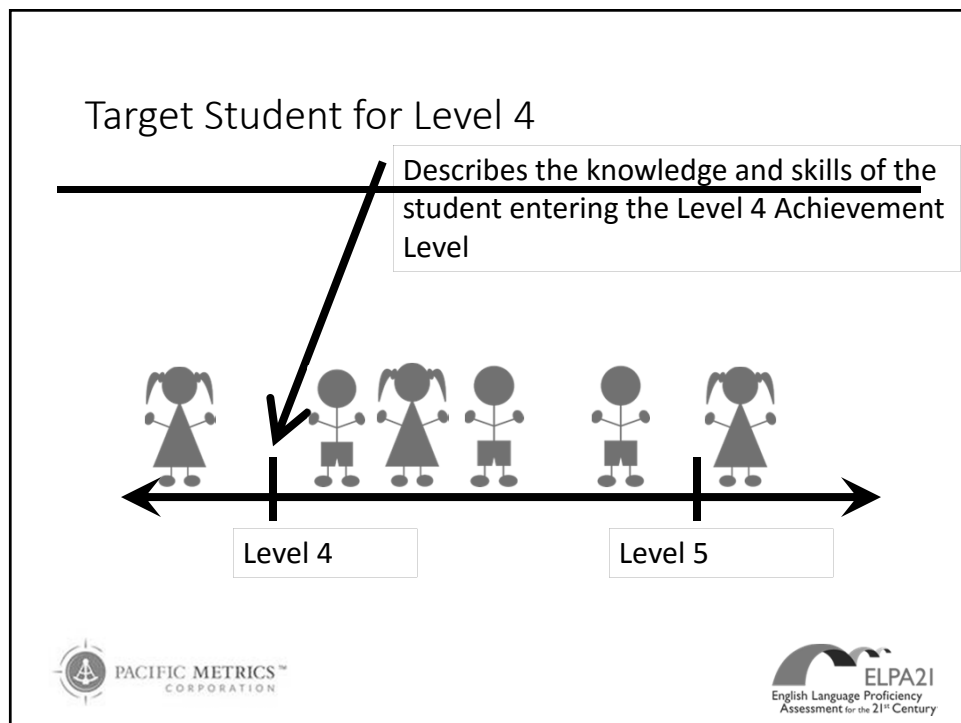
PACIFIC METRICS™
CORPORATION



Q & A FOR READING ELP STANDARDS & ALIS



Appendix D3. Bookmark Training Slides



Target Student Descriptors

Reading Level 4 Target Students

Kindergarten	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
(1) identify main topics and answer questions about key details from read-alouds of literary and informational text; retell some events from read-alouds of literary and informational text and picture books	(1) identify main topics and ask and answer questions about an increasing number of key details in written texts and read-alouds; retell familiar stories or episodes of stories in written texts and read-alouds	(1) identify the main topic or message and answer questions about key details in literary and informational texts	(1) determine main idea or theme; explain how key details support the main idea or theme in written texts; and summarize part of a text	(1) determine the central idea or theme and identify how the theme is supported by specific details in written texts; summarize part of a text	(1) determine the central idea or theme and identify how the theme is developed by specific details in a written text; summarize a text
(2) answer questions about a variety of topics	(2) answer questions about a variety of topics and texts	(2) answer questions about a variety of topics and texts	(2) answer questions about a variety of topics and texts	(2) respond to others' comments and answer questions on a variety of topics and texts	(2) respond to others' comments and answer questions on topics, texts, and issues; add information and evidence and restate key ideas
(3) use information from provided sources and some basic gathered sources to answer a question			(3) gather information from multiple sources and identify key information	(3) gather information from multiple sources and summarize or paraphrase	(3) gather information from multiple sources and begin to determine reliability of each source
	(4) identify reasons an author gives to support the main point	(4) identify and tell how one or two reasons support the main point an author makes	(4) identify how reasons support the specific points an author makes or fails to make	(4) explain the argument an author makes; distinguish between claims that are supported by reasons and evidence from those that are not	(4) explain the argument an author gives to support a claim and identify textual evidence to support the explanation; distinguish between claims that are supported by reasons and evidence from those that are not
(5) determine the meaning of frequently occurring words and phrases in read-alouds about familiar, and some unfamiliar, topics, experiences, or events	(5) determine the meaning of less frequently occurring words, phrases, and some idiomatic expressions in written texts about familiar and unfamiliar topics, experiences, or events	(5) determine the meaning of some less frequently occurring words and phrases, content-specific words, and some idiomatic expressions in written texts about familiar and unfamiliar topics, experiences, or events	(5) determine the meaning of general academic and content-specific words, phrases, and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events	(5) determine the meaning of general academic and content-specific words and phrases and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events	(5) determine the meaning of general academic and content-specific words and phrases, and some idiomatic expressions in written text about familiar and unfamiliar topics, experiences, or events

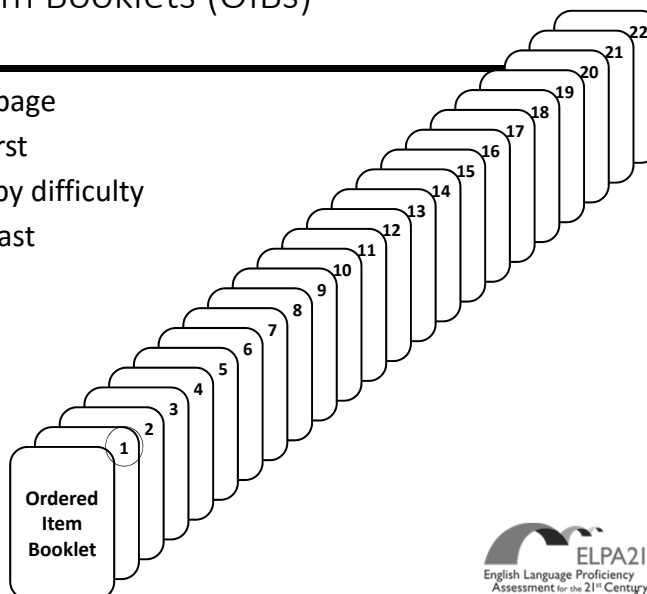


PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Ordered Item Booklets (OIBs)

- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last



PACIFIC METRICS™
CORPORATION

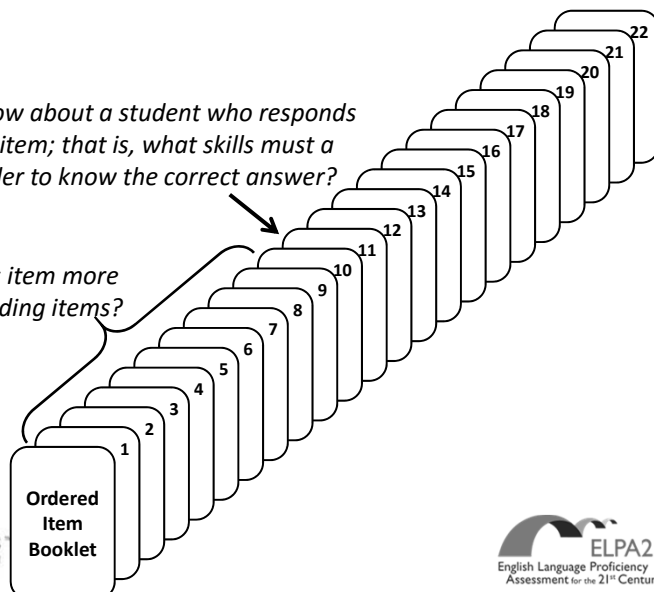
ELPA21
English Language Proficiency
Assessment for the 21st Century

Study the Items in the OIB

As you study each item in the OIB, discuss two questions with your fellow panelists:

1. What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?

2. What makes this item more difficult than preceding items?

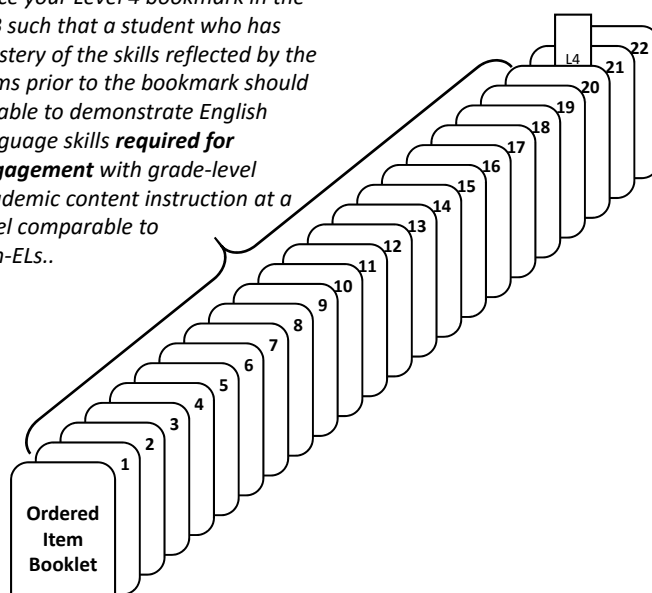


PACIFIC METRICS
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

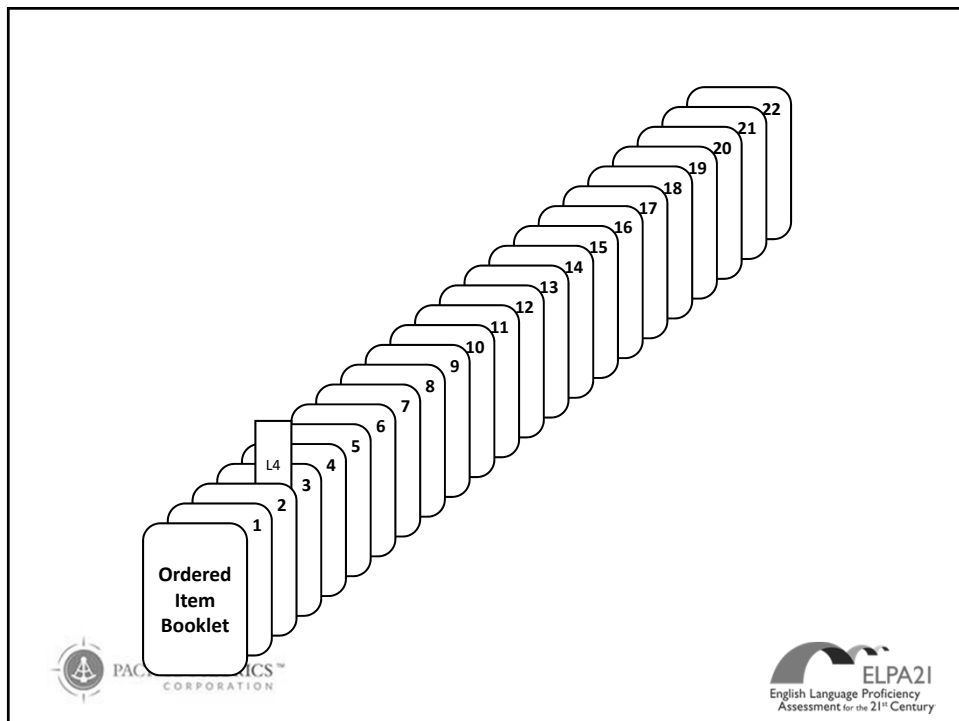
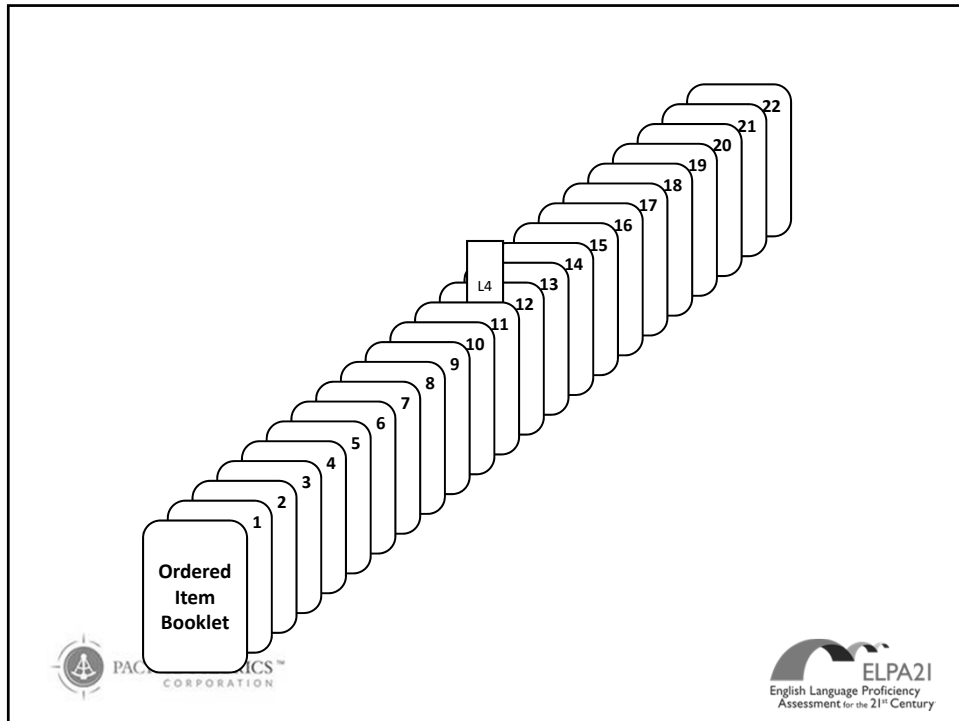
Place your Level 4 bookmark in the OIB such that a student who has mastery of the skills reflected by the items prior to the bookmark should be able to demonstrate English language skills **required for engagement** with grade-level academic content instruction at a level comparable to non-ELs..



PACIFIC METRICS
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century



ELPA21 Cut Scores

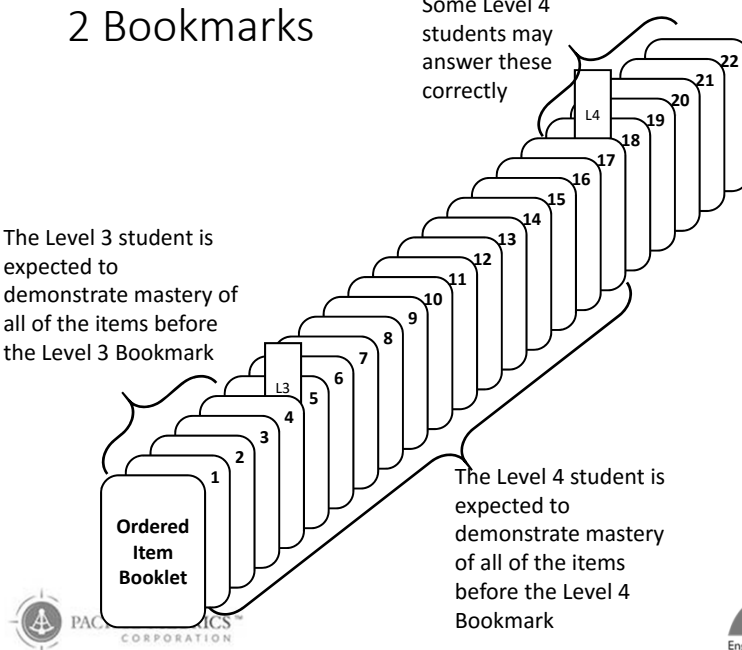
- Four cut scores
 - Level 2, Level 3, Level 4, Level 5
- Five achievement levels
 - Level 1, Level 2, Level 3, Level 4, Level 5
- You will recommend: Levels 3 & 4



2 Bookmarks

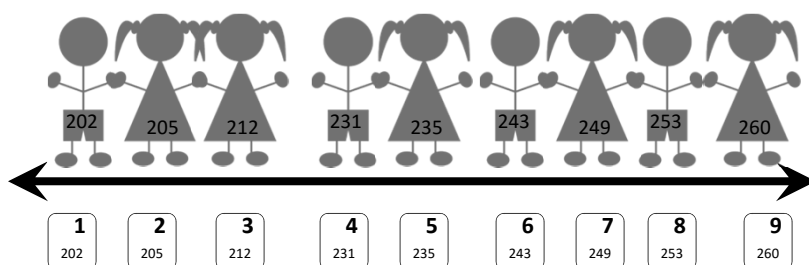
The Level 3 student is expected to demonstrate mastery of all of the items before the Level 3 Bookmark

Some Level 4 students may answer these correctly



Test Scale

- Items are ordered by difficulty
- Students are ordered by achievement



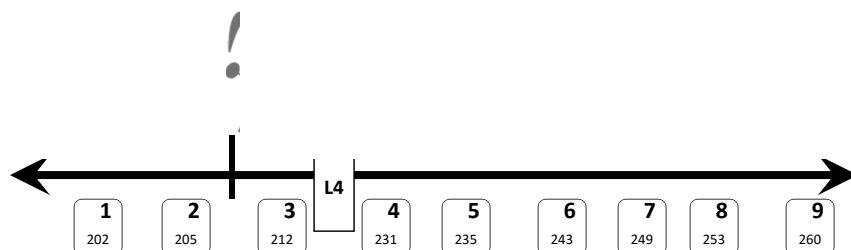
PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

Bookmarks and Cut Scores

- Bookmark: Separates items
- Cut Scores: Separate students



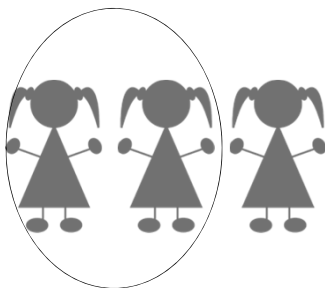
PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

What is meant by Mastery?

- Mastery: the point where a student has a $\frac{2}{3}$ chance of answering the item correctly



20

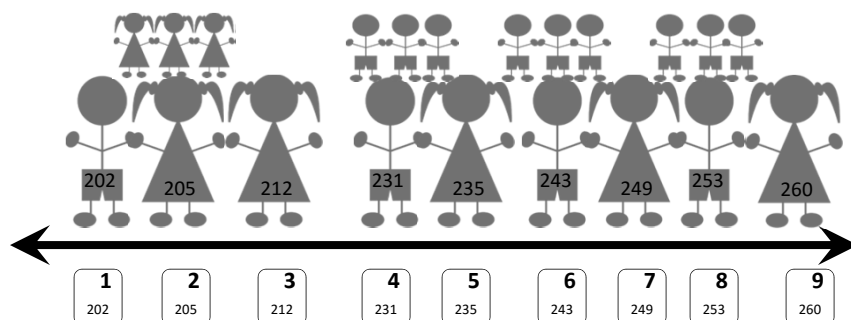


PACIFIC METRICS™
CORPORATION



Mastery & Item Location

- Item Location: Scale score necessary for a student to have $\frac{2}{3}$ chance of answering an item correctly

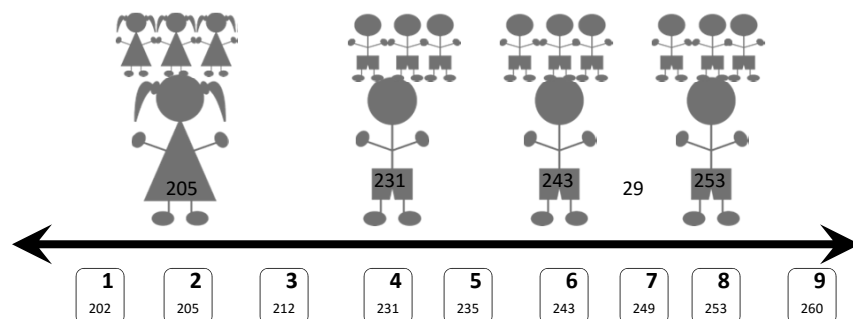


PACIFIC METRICS™
CORPORATION



Mastery & Item Location

- Item Location: Scale score necessary for a student to have 2/3 chance of answering correctly



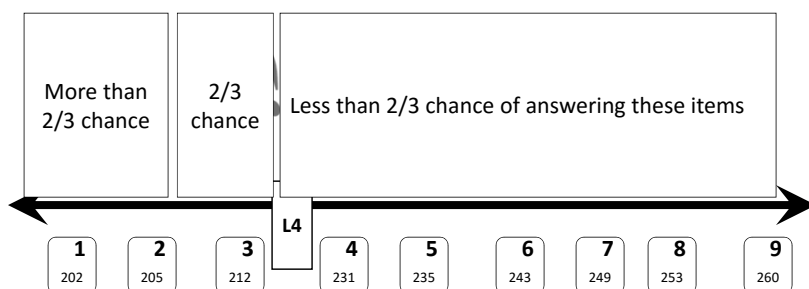
PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

Mastery and the Target Student

- Target Student: 2/3 chance of correctly answering the item just before the Bookmark



PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

Ask yourself: Would a student at the threshold of Level 4 have at least a 2/3 chance of earning this point?

No. Stop and place your bookmark.

Yes. Move onto the next item.

Ordered Item Booklet

PACIFIC METRICS CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Bookmark Placement

- Individual activity
- Use worksheet
- Table Facilitator will enter results in Excel File

ELPA21 Standard Setting Bookmark Rating Form

Name: _____

Circle Domains: Reading Listening Writing Speaking

Grade: _____

Tablet: _____

Panellet ID: _____

Round 1

Level 3 Bookmark	Level 4 Bookmark
------------------	------------------

Round 2

Level 3 Bookmark	Level 4 Bookmark
------------------	------------------

Round 3

Level 3 Bookmark	Level 4 Bookmark
------------------	------------------

Bookmark Placement

- Individual activity
- Use worksheet
- Table Facilitator will enter results in Excel File

Round 1
Bookmark Rating Form

Grade: 1
Domain: Reading
Table: 2

Panelist Number	Enter Panelists' Level 3 Bookmarks Below	Enter Panelists' Level 4 Bookmarks Below
6		
7		
8		
9		
10		



Round 2

Primary Activities

- View Round 1 results
- Discuss Round 1 results at your table
- Place Round 2 Bookmarks

	Review and Discuss Round 1 Level 4 Bookmarks
1	Place a green post-it in your ordered item booklet on the following pages to represent panelists' Round 1 Level 4 Bookmarks: Pages , , , ,
2	Skip this step
3	Discuss your rationales for any differences in your Round 1 Level 4 bookmark ratings.
4	Panelists record their Round 2 Level 4 Bookmark Ratings



Questions?



PACIFIC METRICS™
CORPORATION



Readiness Survey



ELPA21 Standard Setting Post-Bookmark Training Readiness Survey

Please consider the statements below and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble <i>only one</i> of the four options for each statement.				
	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors.	①	②	③	④
2. I participated in bookmark training and had an opportunity to ask questions and discuss the meaning of the Level 3 and Level 4 bookmarks.	①	②	③	④
3. I understand how to place my bookmarks.	①	②	③	④
4. I understand I will have opportunities to change my bookmarks in Rounds 2 and 3.	①	②	③	④

If you answered Disagree or Strongly Disagree to any of questions 1-4, then please answer the next two Yes/No questions.			No	Yes
5. I would like additional training on placing my bookmarks for Round 1.	①	②	①	②
6. I have additional questions that I would like to ask before placing my Round 1 bookmarks.	①	②	①	②



PACIFIC METRICS™
CORPORATION



Appendix D4. Introduction to Round 3 and Impact Data Slides

Important

- Please leave your computers set to the desktop
- Leave your materials in their packets (for now)
 - Exception: Facilitators hold your panelists' rating form
 - No sharing across tables yet!
- Today we begin with Round 3 of Reading
 - We form a single grade group so both tables join together
 - Get comfortable being a bit snug so we can all hear each other and all participate in the conversation



Panelist Training for Next Domains

- After Round 3 Reading discussion this morning we will begin activities for Listening
 - ✓ Already reviewed a form of Listening in the secure browser
- You see that we need to sync up Tables 1 and 2 (per grade group) for Round 3
 - Validity is supported by having each table spend about the same time studying the items
 - After you begin studying the OIB
 - Do a quick check-in across tables when one table hits the 10th (20th, 30th, etc.) item
 - Where is the other table?
 - Meet in the middle
 - One group speeds up just a bit (takes a bit less time per item)
 - The other group slows down just a bit (takes a bit more time per item)
 - See if we can sync up our pace to support similar experience of items



Panelist Training for Next Domains

- Kindergarten *may* move more quickly than HS
 - However, Listening, Writing and Speaking should be more similar in task duration than was Reading (where there was a difference in passage length)
- Because you know the routine you may be tempted to “anticipate.”
 - Do not discuss bookmarks during study of the OIB
 - We want you to make Round 1 bookmarks independently
 - This allows us to see the diversity of ratings
 - Do not discuss bookmarks across tables until Round 3
 - We want tables to make Round 2 bookmarks independently
 - We want to see the diversity of ratings



Today

- When you complete Reading Round 3 today
 - Each table enters their bookmarks
 - QA your work!
 - Collect all secure materials as yesterday.
 - Complete evaluation
 - Bundle and take a break
 - Reading materials collected
 - Listening materials delivered

Listening

- Secure materials sign-out
- Review Listening ALIs
 - Questions for Carolyn?
- Study Listening ordered item booklet (OIB)
 - Take notes on item map, etc.



Round 3 Display (of Round2 Results)

This data is a rough estimate at the moment. It will change when all student data is scored. Do not over-interpret

1. Visual Check of Bookmarks

2. Medians

3. Impact data

	D	E	F	G
			Level 3	Level 4 Bookmark
2	Data based on R2	Table 1	Median	
3	Table 2	Median		
4	Grade Band (T1 & T2)	Median		
5				
6				
7				
8				
9	Table Facilitator Visual Check			
10	Table 1 Level 3 Bookmarks:			
11	Table 1 Level 4 Bookmarks:			
12				
13	Table 2 Level 3 Bookmarks:			
14	Table 2 Level 4 Bookmarks:			

Estimated Percent of students based on R2 Results

	Grade Band (T1 & T2)	Levels 1 + 2	Level 3	Levels 4 + 5
		??%	??%	??%

PACIFIC METRICS™ CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Round 3 Display (of Round2 Results)

- This is primarily a criterion referenced standard setting
- Skills associated with level are the key
- We provide it as a reality check
 - Considered best practice
- You know
 - Students
 - Skills you are associating with each level
 - Data should not be surprising
- If impact data are consistent with your expectations
 - Discuss differences between Table 1 and Table 2 medians
 - What are the rationales for your differences
 - Discuss appropriate expectations for students in each level

Round 3 Display (of Round2 Results)

- If impact data are not consistent with your expectations
 - Table Facilitators! Raise your hands so a Workshop Facilitator can help frame the conversation
 - Go back to OIBs—skills are the key
 - Evaluate your bookmarks with respect to data. For example:
 - If percentage in Level 4 + Level 5 seem high, then perhaps additional skills should be expected for the Level 4 bookmark
 - Is it reasonable to set your Level 4 bookmark further in the OIB?
 - If percentage in Level 4 + Level 5 seem low, then perhaps fewer skills should be expected for the Level 4 bookmark
 - Is it reasonable to set your Level 4 bookmark at a lower page in the OIB?
 - Target student descriptor should be considered but does not dictate your bookmarks—that is your decision
 - Discuss impact data
 - Discuss differences between Table 1 and Table 2 medians
 - What are the rationales for your differences
 - Discuss appropriate expectations for students in each level



PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

Round 3

- When Discussion is complete
 - Make Round 3 ratings individually on rating forms
 - Pass rating forms to Table Facilitators
 - Table Facilitators enter ratings into R3 Ratings tab in Excel
 - SAVE!
 - Collect materials systematically
 - Take a break
- Overheads will provide activity reminders for Listening



PACIFIC METRICS™
CORPORATION



ELPA21
English Language Proficiency
Assessment for the 21st Century

Appendix D5. Proficiency Determination Training Slides

ELPA21 Standard Setting

Proficiency Determination



Achievement Level Policy Definitions

A STUDENT AT THIS LEVEL...	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of developing grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits superior English language skills, as measured by ELPA21.



Proficiency Determination Definitions

Proficient	ELPA21 recommends that students be considered <i>Proficient</i> when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once <i>Proficient</i> on ELPA21, ELLs can be considered for reclassification.
Progressing	ELPA21 recommends that students be considered <i>Progressing</i> when they are approaching a level of English language necessary to produce, interpret, and collaborate, with support, on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ...(<i>Profile to be determined</i>). Students scoring <i>Progressing</i> on ELPA21 are recommended to receive program support.
Emerging	ELPA21 recommends that students be considered <i>Emerging</i> when they have not yet attained a level of English language necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ...(<i>Profile to be determined</i>). Students scoring <i>Emerging</i> on ELPA21 are eligible for ongoing program support.

Note: The shaded text in the description of Progressing and Emerging cannot be finalized until after standard setting based on the outcomes of the workshop.



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

Proficiency Determination

- The Proficiency Determination evaluates performance across domains
- Profile-based
 - What is the Proficiency Determination for various profiles?
 - Proficient
 - Progressing
 - Emerging



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

What Proficiency Determination Should be Associated with the Following Profiles?

ID	First	Last	Grade	Reading	Writing	Listening	Speaking
1	Laura	Jones	K	4	4	4	4
2	Dan	Lewis	K	3	3	4	4
3	Mark	Bloch	HS	3	3	4	4
4	Sam	Laszlo	K	4	4	3	3
5	Andy	Winder	HS	4	4	3	3



Process

- Systematic process to support appropriate rules guiding the Proficiency Determination
 - Contrasting Groups Study
 - One piece of evidence for you to *consider*
 - EL Expert Advisory panel recommendations
 - A second piece of evidence for you to *consider*
 - This standard setting Proficiency Determination group's recommendations
 - CCSSO/ELPA21 sets policy based on these sources of information



Proficiency Determination

Summary of Contrasting Groups Study



Contrasting Groups Study Results

- Teachers in ELPA21 states read the policy descriptors and completed surveys
 - They classified their students into one of five levels for Reading, Writing, Listening, Speaking, and Overall based on the policy descriptors and their knowledge of the students' classroom achievement.



Profile	N	%	Cum %	Profile	N	%	Cum %	Profile	N	%	Cum %
33333	460	7.0%	7.0%	11122	60	0.9%	65.8%	44354	24	0.4%	78.6%
11111	405	6.2%	13.2%	32244	52	0.8%	66.6%	32343	23	0.4%	79.0%
22222	368	5.6%	18.8%	43444	50	0.8%	67.4%	43343	23	0.4%	79.3%
44444	352	5.4%	24.2%	44355	50	0.8%	68.1%	43454	23	0.4%	79.7%
33344	275	4.2%	28.4%	32243	46	0.7%	68.8%	21133	21	0.3%	80.0%
55555	249	3.8%	32.2%	43354	42	0.6%	69.5%	23222	20	0.3%	80.3%
44455	185	2.8%	35.1%	33354	41	0.6%	70.1%	34334	20	0.3%	80.6%
22233	167	2.6%	37.6%	34344	39	0.6%	70.7%	45444	20	0.3%	80.9%
44344	165	2.5%	40.1%	43355	37	0.6%	71.3%	22133	19	0.3%	81.2%
33233	153	2.3%	42.5%	54555	34	0.5%	71.8%	23223	19	0.3%	81.5%
32233	149	2.3%	44.8%	34333	33	0.5%	72.3%	34343	19	0.3%	81.8%
43344	145	2.2%	47.0%	45455	33	0.5%	72.8%	22243	18	0.3%	82.0%
22223	131	2.0%	49.0%	21222	32	0.5%	73.3%	32234	18	0.3%	82.3%
33343	123	1.9%	50.9%	33243	30	0.5%	73.7%	44345	18	0.3%	82.6%
33334	115	1.8%	52.6%	43455	30	0.5%	74.2%	33444	17	0.3%	82.8%
55455	113	1.7%	54.3%	21132	29	0.4%	74.6%	21112	16	0.2%	83.1%
22232	92	1.4%	55.7%	32344	28	0.4%	75.1%	22112	16	0.2%	83.3%
54455	83	1.3%	57.0%	11121	27	0.4%	75.5%	22244	16	0.2%	83.6%
11112	81	1.2%	58.3%	44434	27	0.4%	75.9%	33223	16	0.2%	83.8%
21122	81	1.2%	59.5%	21123	26	0.4%	76.3%	22123	15	0.2%	84.1%
44454	75	1.1%	60.6%	32232	26	0.4%	76.7%	22212	15	0.2%	84.3%
33244	74	1.1%	61.8%	33234	26	0.4%	77.1%	22234	15	0.2%	84.5%
22122	71	1.1%	62.9%	33323	26	0.4%	77.5%	23233	14	0.2%	84.7%
32333	71	1.1%	63.9%	32223	25	0.4%	77.9%	33332	14	0.2%	84.9%
44445	62	0.9%	64.9%	33355	24	0.4%	78.2%	33345	14	0.2%	85.2%

Grade Level Results



Summary of CGS Results

- Profiles associated with a Proficient Proficiency Determination is indicated by
 - Writing = 3, if other domains are 4 or 5
 - Writing and Reading = 3 and S and L = 4 or 5
 - any combination of 4s and 5s



Summary of CGS Results

- Profiles associated with a Proficient Proficiency Determination is indicated by
 - A student with a 1 or 2 in any domain can not be Proficient
 - Two profiles conflict across most grades (33344 & 43344, and 22233 & 32233).



Proficiency Determination

Summary of EL Expert Advisory Panel Recommendations



EL Expert Advisory Panel Discussion

- EL Expert Panels met and we reviewed CGS results
- Discussed various profile rules
- The next week they completed a survey
- Responses from 6 panel members



EL Expert Advisory Panel Proficiency Determination Survey Results

What profiles result in a Proficiency Determination of Proficient?

Majority rule: A student is Proficient if

- Reading, Writing, Listening & Speaking are all 4's and 5's
- Reading and Writing are 4 and Listening & Speaking are at least 3
- No compensatory model
 - A 3 cannot be compensated with a 5



Considerations from CGS and EL Expert Panel Surveys

- 3344 versus 4433
- Across grade band differences in profiles
- Compensatory models
 - CGS sometimes used compensatory models
 - EL Expert Panel did not recommend a compensatory model for Level 4



Proficiency Determination

Today's Activities



Today's Discussion

- Should the profiles associated with Proficient, Progressing, and Emerging be the same for all grade bands?
 - Can they differ by grade band?
 - Seemed to for CGS
 - Why or why not?



Today's Discussion and Activities

- We will discuss issues, options, considerations across grade bands
- We will break into grade band groups for focused discussions
- Grade band groups will summarize their thinking and report to whole group
- We would like a recommendation from each grade group
 - Note if you have modest disagreement
- Recommendations should be sensible when viewed from grade to grade



Today's Discussion

- There are $5 \times 5 \times 5 \times 5 = 625$ unique profiles *per grade band*
- Profile rules that are easy to communicate and interpret are desirable.
 - **Hypothetical examples:** Students receive a *Proficient* Proficiency determination if they have
 - 4 or 5 in every domain
 - This is a conjunctive model
 - $R \geq 4$ and $W \geq 4$ and $L \geq 4$ and $S \geq 4$
 - » Overall percent limited by most rigorous domain
 - 4 in three domains and 3 in one domain
 - Mixed model
 - $R \geq 4$ and $W \geq 4$ and $L \geq 4$ and $S \geq 3$
 - Or $R \geq 4$ and $W \geq 4$ and $L \geq 3$ and $S \geq 4$
 - Or $R \geq 4$ and $W \geq 3$ and $L \geq 4$ and $S \geq 4$
 - Or $R \geq 3$ and $W \geq 4$ and $L \geq 4$ and $S \geq 4$
 - 4 in three domains and 3 in Listening or Speaking
 - 4 in two domains and 3 in two domains



Today's Discussion

- *Hypothetical examples (continued)*: Students receive a *Proficient* Proficiency determination if they have
 - Compensatory profiles:
 - (4 in two domains) and (3 and 5 in the other two domains)
 - » A 5 compensates for a 3
 - (3 in two domains) and (5 in two domains)



What Profiles are Associated with a “Proficient” Proficiency Determination

- All 4s or 5s?
- Can you have one 3?
 - 3444
 - 4344
 - 4434
 - 4443
 - Does it matter which domain has the 3?
- Can you have two 3s?
 - Does it matter which domains have the 3s?



What Profiles are Associated with a “Proficient” Proficiency Determination

- Compensatory model:
 - Can you have one 3 if it is compensated by a 5?
 - 3544
 - 3454
 - 3445
 - Etc...
 - Does it matter which domain has the 3 and the 5?



What Profiles are Associated with a “Proficient” Proficiency Determination

- Compensatory model
 - Can you have two 3s if they are compensated by 2 5s?
 - 3345
 - 3354
 - Etc...
 - Does it matter which domain has the 3s and the 5s?



What Profiles are Associated with a “Proficient” Proficiency Determination

- Can you have any 2s or 1s and be Proficient?



What Profiles are Associated with an “Emerging” Proficiency Determination

- Consider profiles that differentiate Emerging and Progressing
- Which profiles might be associated with Emerging?
 - All 1s: 1111
 - Three 1s and one 2
 - Does it matter which domain the 2 is in?
 - Two 1s and two 2s?
 - Does it matter which domain the 2s are in?
 - One 1 and three 2s?
 - Does it matter which domain the 1 is in?
 - 2222?
 - Are you Emerging if you have one 3 (one 4?)
- What profile moves the student into the entry of Developing?



ELPA21 Standard Setting

Whole Group Discussion



Final Proficiency Determination Definitions

Proficient	ELPA21 recommends that students be considered <i>Proficient</i> when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once <i>Proficient</i> on ELPA21, ELLs can be considered for reclassification.
Progressing	ELPA21 recommends that students be considered <i>Progressing</i> when they are approaching a level of English language necessary to produce, interpret, and collaborate, with support, on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... <i>(Profile to be determined)</i> . Students scoring <i>Progressing</i> on ELPA21 are recommended to receive program support.
Emerging	ELPA21 recommends that students be considered <i>Emerging</i> when they have not yet attained a level of English language necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... <i>(Profile to be determined)</i> . Students scoring <i>Emerging</i> on ELPA21 are eligible for ongoing program support.

Note: The shaded text in the description of Progressing and Emerging cannot be finalized until after standard setting based on the outcomes of the workshop.



Today's Discussion

- Should the profiles associated with Proficient, Progressing, and Emerging be the same for all grade bands?
 - Can they differ by grade band?
 - Seemed to for CGS
 - Why or why not?



Grade Band Activities

Thought questions will be displayed on each monitor

1. Spend some time formulating your own opinion before moving to group discussion
2. Open up the conversation to group discussion
3. Capture various options discussed on worksheets provided
4. Enter group recommendations on computer worksheet
 - Recommendations should be interpretable and captured with both examples and documentation of the “guiding principle.”

Guiding Principle Here (Rule and rationale): To be Proficient a student must have a profile of ...because....				
Overall	Reading	Writing	Listening	Speaking
4	?	?	?	?



Grade Band Activities

Thought questions will be displayed on each monitor

5. We will reassemble as a cross grade group where grade bands will report on their thinking
 - Discuss different approaches
 - Review articulation of profile recommendations across grade bands
6. If necessary or desired we will move back into grade bands for further consideration
7. We will ask for final recommendations and document them, noting the degree of consensus



ELPA21 Standard Setting

Thought Questions for Grade Groups
(Displayed on Laptops and Monitors)



Final Proficiency Determination Definitions

Proficient	ELPA21 recommends that students be considered <i>Proficient</i> when they attain a level of English language necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of proficiency across all non-exempted domains. Once <i>Proficient</i> on ELPA21, ELLs can be considered for reclassification.
Progressing	ELPA21 recommends that students be considered <i>Progressing</i> when they are approaching a level of English language necessary to produce, interpret, and collaborate, with support, on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... <i>(Profile to be determined)</i> . Students scoring <i>Progressing</i> on ELPA21 are recommended to receive program support.
Emerging	ELPA21 recommends that students be considered <i>Emerging</i> when they have not yet attained a level of English language necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile ... <i>(Profile to be determined)</i> . Students scoring <i>Emerging</i> on ELPA21 are eligible for ongoing program support.

Note: The shaded text in the description of Progressing and Emerging cannot be finalized until after standard setting based on the outcomes of the workshop.



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

What Profiles are Associated with a “Proficient” Proficiency Determination: Thought Questions:

- Is a student with all 4s or 5s Proficient?
- If the rest of the domains are at least 4:
 - Can you have one 3?
 - Does it matter which domain has the 3?
 - Can you have two 3s?
 - Does it matter which domains have the 3s?
 - Can you have one 3 if it is compensated by a 5?
 - Does it matter which domain has the 3 and the 5?
- Can you have two 3s if they are compensated by two 5s?
 - Does it matter which domain has the 3s and the 5s?
- Can you have any 2s or 1s and be Proficient?



PACIFIC METRICS™
CORPORATION

ELPA21
English Language Proficiency
Assessment for the 21st Century

What Profiles are Associated with an “Emerging” Proficiency Determination: Thought Questions

- Consider profiles that differentiate Emerging and Progressing
- Which profiles might be associated with Emerging?
 - All 1s: 1111
 - Three 1s and one 2
 - Does it matter which domain the 2 is in?
 - Two 1s and two 2s?
 - Does it matter which domain the 2s are in?
 - One 1 and three 2s?
 - Does it matter which domain the 1 is in?
 - 2222?
 - Are you Emerging if you have one 3 (one 4?)
- What profile moves the student into the entry of Developing?



Worksheet

Guiding Principle Here (Rule and rationale): To be Proficient a student must have a profile of ...because....				
Overall	Reading	Writing	Listening	Speaking
4	?	?	?	?



Appendix E. In-Person Bookmark Standard Setting Workshop Materials

Appendix E1. Table Facilitator Agenda

Note: Times are approximate and will be adjusted as needed. Breaks will be provided throughout.

Tuesday Morning, July 19th

Morning Agenda and Goals:

- Introductions
- Distribute and track secure materials
- Identify a volunteer scribe to take notes on the computer for your table
- EL Expert leads Q & A of the ELP standards and ALIs
- Support the review of one form of your grade band assessment online
- Study the ordered item booklet (OIB) at your table. This is intended to help panelists attain a deeper understanding of what the test measures

What You Need (all materials provided by Pacific Metrics):

- Panelist materials assignment and tracking form
- Secure materials packets (printed copies of ALIs, OIBs, item maps, and stimulus booklets)
- Laptop

Schedule and Description of Activities and Roles:

- **7:30 AM:** Table Facilitators meet with Workshop Facilitators in Cupples Ballroom
- **8:00 AM:** Opening Session (all workshop attendees)
- **9:30 AM:** Introductions and secure materials sign-out
 - A packet of secure materials will be provided for each panelist. Each secure piece has a Panelist ID number in the top right hand corner. The Table Facilitators for a grade band should assign secure packets so that they are consecutively numbered at your tables (e.g., Table 1 assigns sets 1–4 and Table 2 assigns sets 5–8). The Table Facilitators work together to complete the tracking form for your grade band. Have panelists sign and print their names on the tracking form by the ID number of the materials they receive. They should also put their names on each piece of secure material.
 - Introduce yourself and support brief introductions around the table. There will be many opportunities to get to know your fellow panelists so keep initial introductions brief (modeling this with your own brief introduction). One minute per panelist is a good target.
 - Ask for a volunteer to act as scribe at the table (this may be a rotating role). The scribe will take notes on the computer about items when studying the OIB. The scribe should sit next to you so that you can share the laptop controls (there is a large monitor at each table for panelists to view). You may choose to be the scribe but you have enough responsibility and it is probably better for someone else to take that on. There will be a modest amount of opening different files (all contained in a single folder on the

desktop) so the scribe should have strong keyboarding skills and be able to work with you to move to the various files.

- **9:45 AM:** Q & A for the ELP Standards and ALIs for Reading. The Workshop Facilitators and EL Expert will direct panelists to the appropriate materials and open the floor up for questions. ELP standards and ALIs should be reviewed prior to the workshop and this Q & A is intended to answer any questions panelists may have.
- **10:00 AM:** Table-level review of Reading online operational form. The Workshop Facilitators will demonstrate how to access the online operational forms on your computer. This activity is intended to provide an opportunity to see a full form of the Reading assessment in its online format as a student sees it.
 - Follow the Workshop Facilitators' prompts and let them know if you have challenges logging in. Move through each item as a student would; perhaps a bit quicker. You will have many opportunities to discuss these items so use this one to experience the assessment as a student and not as a means to prompt discussion.
 - Take a 10-minute break after completion.
- **10:30 AM:** OIB Review
 - Upon return from break, begin studying the OIB.
 - Panelists should have their OIBs, item maps, and stimulus booklets out.
 - Ask the scribe to open the item map so it appears on the monitor and take notes with input from the table.
 - Panelists turn to the first item in their OIBs and locate it on their item maps. Review the item:
 - Table Facilitator prompts panelists with the first question: *What does this item measure? That is, what do you know about a student who responds successfully to this item?*
 - The scribe should be taking notes summarizing the responses, which will be used during achievement level descriptor (ALD) writing. Notes that are informative, succinct responses will be most useful for creating valid ALDs to help teachers and parents understand students' KSPs based on their test results.
 - Table Facilitator asks the second question: *Why is this item more difficult than the preceding items?*
 - Scribe takes notes.
 - When discussion ceases to provide new information, move to the next question.
 - Repeat for all items in the OIB.
 - Considerations:
 - Stimulus booklet. Longer stimuli that are common to multiple items are presented in a separate booklet to maintain the flow of the OIB. This booklet is ordered by Set Leader, which can be found on each associated page in the OIB. To view an item's stimulus, identify the Set Leader in the OIB and find that number in the table of contents of the stimulus booklet. Then find the page on which the stimulus appears.
 - Partial credit item rubric: A paper copy of the simple rubrics used for the automated scoring of Technology Enabled items will be provided. The scoring rubric for these items is frequently intuitive. When it is not, please refer to the paper copy of the rubrics for these items.

- Monitor time. Encourage thoughtful discussion, and do not rush through items. Remember, you will be discussing the items in this activity and after Rounds 1 and 2 before your final round of ratings. You *do not* need to exhaustively dissect every item. It is important to the validity of the process that panelists have enough time and don't feel rushed. It is also important to complete the process in our 4-day workshop. These activities must be completed for all four domains in that time period.
- Support equitable and diverse discussion by encouraging all panelists to contribute to the discussion.
- The items should be familiar. Typically, most of the items are on the form you viewed online in the previous activity. If panelists don't feel comfortable discussing the item without seeing it online, then you may view it on the online item viewer. You will be provided training on the item viewer. The link to the online viewer is in the folder on your computer desktop and will be opened and activated for the current domain before you begin.
 - To find an item in the item viewer, use the number after the underscore in the Item ID column in the item map. (See Figure 4 in the Appendix for a sample item map.)
- Caveats
 - Do not spend time critiquing items. While this is natural, this is not an item review workshop. If panelists feel the need to provide item input, have them write their concerns on an index card provided at the table for ELPA21 staff review.
 - If you look up every item in the online viewer, you will need much more time than we have. Most items have a structure similar to items in the online form you viewed and panelists should be able to review the questions without viewing each item in the online viewer.
- **Noon:** Lunch
 - All secure materials are color coded. Be sure all color-coded materials are left in the room. Panelists should never remove these materials from the room.

For Help:

- If you have questions about content, call the EL Expert to your table.
- If you have questions about process or policy, call the Workshop Facilitator to your table. Policy questions will be answered by ELPA21 staff, but will be collected by the Workshop Facilitators and answered promptly if vital to proceed or if not vital but the answer will benefit all panelists in a group setting.

Tuesday Afternoon, July 19th

Afternoon Agenda and Goals:

- Review Target Student Descriptors
- Bookmark Training: Receive training to make Bookmark recommendations
- Round 1: Panelists make their first cut score recommendations
- Round 2: Panelists discuss differences in cut score recommendations within their tables
- Round 3: Panelists review the reasonableness of impact data associated with the grade band median recommendation and discuss differences between the grade band Table 1 and Table 2 cut score recommendations
- View final results
- Distribute, complete, and collect Domain evaluations
- Support systematic collection of secure materials

What You Need:

- OIBs, item maps, stimulus booklets
- Rating forms
- Excel file
- Readiness surveys
- Evaluations
- Secure materials collection checklist

Schedule and Description of Activities and Roles:

- **1:00 PM:** Reading Round 1 (R1)
 - Workshop Facilitator and EL Expert review Target Student Descriptors
 - Workshop Facilitator provides Bookmark training
 - Readiness surveys. You will pass out the Bookmark readiness surveys following training. If someone at your table marks “yes” to the last two questions, then raise your hand that additional training is needed for your table.
 - Round 1 ratings
 - Panelists make their ratings on their paper rating forms independently and without discussion. If there are questions, ask a Workshop Facilitator.
 - Collect panelists’ rating forms.
 - Review each panelist’s rating form to see that the Level 4 Bookmark is higher than the Level 3 Bookmark.
 - When all are collected open the Excel “Enter R1 Ratings” tab and enter all data in the orange highlighted cells for Level 3 and the green cells for Level 4. See Figure 1 for a screenshot of the “Enter R1 Ratings” tab.
 - Ask another panelist to review your entries for accuracy. When you have entered them and the entries have gone through QA, save the excel file.
 - Have panelists take a short break after providing you their completed rating forms.

	A	B	C	D	E	F	G	H	I
2							Enter Panelists' Level 3 Bookmarks Below		Enter Panelists' Level 4 Bookmarks Below
3									
4									
5					Panelist Number				
6		Grade: 3			1				
7		Domain: Reading							
8		Table: 1			2				
9									
10					3				
11									
12					4				
13									
14					5				

Figure 1. Excel “Enter R1 Ratings” tab

- **2:30 PM:** Reading Round 2 (R2): Discussion of Round 1 results and Round 2 ratings
 - Open the Excel file to “R2 Display” tab and follow steps 1–9. See Figure 2 for a screenshot of the “R2 Display” tab.
 1. Each panelist places a green post-representing your table’s panelists’ Level 4 bookmarks. We start with Level 4, not Level 3, because Level 4 is the “anchor level.”
 2. Panelists place a blue post-it bookmark representing the Level 4 Contrasting Groups Study (CGS) results.
 - Remind panelists of the CGS training from the opening session. The CGS results are another reference point, described below.
 - Contrasting Groups Study. The CGS is a study conducted by Pacific Metrics for ELPA21. EL teachers in ELPA21 states classified their EL students into 5 levels of achievement for each domain and overall based on the ELPA21 Achievement Level Policy Descriptors. Pacific Metrics analyzed this data and by associating teachers’ classifications with the operational ELPA21 test results for these students we are able to estimate cut scores and bookmarks associated with the teachers’ judgments.
 - The results of the CGS will be provided to panelists by translating the CGS cut score estimates into bookmarks in the OIBs so panelists can see the CGS results expressed in the same metric used to express their own ratings.
 3. Discussion of Round 1 Level 4 bookmarks. Ask panelists to discuss the reasons for their various bookmark placements. One way to do this is to begin by having the panelists with the lowest and highest bookmarks discuss their rationales for their bookmark placement.

- Remind panelists that there are no wrong bookmark placements, just differences of opinion. Round 2 allows panelists to reflect on their own and others' bookmark placements.
4. Panelists record Round 2 ratings for Level 4 based on their updated perspective.
 5. Remove Level 4 bookmarks. Each panelist now places an orange post-it for each of the panelist's Round 1 Level 3 bookmarks.
 6. Place a blue post-it representing the Level 3 CGS results.
 7. Discuss Round 1 Level 3 differences.
 8. Panelists record Round 2 ratings for Level 3.
 9. Collect the rating forms, enter them into the Excel "R2 Ratings" tab, and QA them.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF		
1			Results from Round 1 Displayed at the Beginning of Round 2																															
2			Review and Discuss Round 1 Level 4 Bookmarks																	Review and Discuss Round 1 Level 3 Bookmarks														
3																																		
4																																		
5			1	Place a green post-it in your ordered item booklet on the following pages to represent panelists' Round 1 Level 4 Bookmarks:																5	Place a blue post-it in your ordered item booklet on the following pages to represent panelists' Round 1 Level 3 Bookmarks:													
6				Pages 12 , 14 , 16 , 18 , 20																	Pages 2 , 4 , 6 , 8 , 10													
7																																		
8			2	Place a blue bookmark on page 13 for the Level 4 CGS results.																6	Place a blue bookmark on page ? for the Level 3 CGS results.													
9																																		
10			3	Discuss your rationales for any differences in your Round 1 Level 4 bookmark ratings.																7	Discuss your rationales for any differences in your Round 1 Level 3 bookmark ratings.													
11																																		
12			4	Panelists record their Round 2 Level 4 Bookmark Ratings																8	Panelists record their Round 2 Level 3 Bookmark Ratings													
13																																		
14																				9	Record panelists' new Level 3 and Level 4 bookmarks in "Enter R2 Ratings" tab, QA the entries, and save the file.													
<div><div><div>Enter R1 Ratings</div><div>R2 Display</div><div>Enter R2 Ratings</div><div>R3 Display</div><div>Enter R3 Ratings</div><div>Post R3 Display</div><div>domain grade table</div><div>FDs (hide)</div><div><div>+</div></div></div></div>																																		

Figure 2. Excel "R2 Display" tab

- **3:30 PM** (Time tentative, depending on R2 duration. May carry over to the next day if necessary.): Reading Round 3 (R3)
 - The two tables join together as one group representing the grade band. You no longer work as separate, independent tables.
 - The Workshop Facilitator will introduce Round 3 activities, which include reviewing
 - Each table's median bookmark
 - The grade band combined bookmark (median of all panelists at Table 1 and Table 2)
 - The impact data—the percentage of students in the various levels based on the grade band median bookmark
 - The CGS cut score

- Open the Excel “R3 Display” tab when instructed by the Workshop Facilitator. Figure 3 provides a screenshot of the Excel “R3 Display” tab.

	A	B	C	D	E	F	G
1				Data based on R2 Bookmarks		Level 3 Bookmark	Level 4 Bookmark
2				Table 1 Median		6	16
3				Table 2 Median		5	17
4				Grade Band (T1 & T2) Median		5.5	16.5
5				CGS		?	?
6							
7					Estimated percent of students based on R2 Results		
8					Levels 1 + 2	Level 3	Levels 4 + 5
9				Grade Band (T1 & T2)	x	x	x
10				Table Faciliator Visual Check			
11				Table 1 Level 3 Bookmarks:	2	4	6
12				Table 1 Level 4 Bookmarks:	12	14	16
13							
14				Table 2 Level 3 Bookmarks:	1	3	5
15				Table 2 Level 4 Bookmarks	13	15	17
16							

Figure 3. Excel “R3 Display” tab

- Table Faciliator Visual Check. We have linked the Table 1 and Table 2 Excel files. Both grade band Table Facilitators should check the following:
 - The same data is displayed in each of their Excel files.
 - The bookmarks for their respective tables are correctly represented (you can look to the “Enter R3 Ratings” tab to check your own table’s bookmarks).
 - If the Table Faciliator Visual Check indicates correct data, then proceed. If there are any discrepancies call a Workshop Facilitator to update your spreadsheet.
- Observe the impact data. If you are very surprised by the data, call a Workshop Facilitator to your table to help frame the discussion.
- As a single group (Tables 1 and 2) discuss the differences between the two tables’ median bookmarks and the CGS bookmarks.
 - Begin with Level 4
 - Discuss differences between each table’s Level 4 Bookmarks.
 - Panelists at both tables record their Round 3 Level 4 ratings on the rating form.
 - Repeat for Level 3
- Collect the rating forms and enter them in the Excel “R3 Ratings” tab
- **4:45 PM:** Session close. Reading Evaluations and collection of secure materials. Sessions may run longer if we are near completion of an important task.
 - Workshop Facilitators will initiate close of the daily session
 - Panelists complete Reading evaluations
 - Table Facilitators collect evaluations and submit to Workshop Facilitators

- Table Facilitators follow systematic secure materials collection as described by Workshop Facilitators
- **5:15 PM: Daily Debrief**
 - The Table Facilitators are asked to attend the first part of the daily debrief in order to provide feedback:
 - Challenges that should be shared
 - What went well and what could be improved
 - Share any useful information associated with the quality of the workshop, table dynamics, etc.

Note that we have provided detailed directions for one domain. We will begin with Reading and when have completed all three rounds for Reading we will proceed and repeat this process for Listening, Writing, and Speaking. The only difference between domains will be the nature of the content and that Writing and Speaking have rubrics and examples of student work to help understand the scoring of some items (written work for writing and mp3 sound files for Speaking). We will provide training on the review of rubrics and examples of student work for Writing and Speaking at the workshop.

Appendix: Additional Details

Studying the Ordered Item Booklet


Panelists develop a comprehensive understanding of what each domain and the associated tests measure by studying OIBs—a set of test items selected to be representative of the construct measured by each form of the test, ordered by difficulty.

The following three sets of materials are used to study the OIBs

1. Ordered Item Booklet. The OIB has one item per page with the easiest item first and the most difficult item last. Each page includes the item and information such as ELP Standards and Sub-Claims measured by the item.
2. Stimulus Booklet. Longer stimuli that are common to multiple items are presented in a separate booklet to maintain the flow of the OIB. This booklet is ordered by Set Leader ID, which can be found on each associated page in the OIB.
3. Item Map. Item maps support the review of the OIB for each domain and grade band. The item maps include the following information, as illustrated in Figure 4:
 - Order of Difficulty: OIB Page number
 - Location: The scale score at which a student has a 2/3 likelihood of a correct response
 - Score Point (1 for multiple choice, the score point considered for partial credit items)
 - Item and Set Leader ID (the numbers after the underscore in Item ID are used to identify an item in the item viewer)
 - Task Type and Task Sub Type

Panelists at the standard setting workshop will complete the final two columns of the item map:

- What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?
- Why is this item or score point more difficult than the items that precede it?



Use this number to look up the item in the item viewer

ELPA21 Standard Setting July 2016

Grade 3 Reading Item Map

Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID	AIR ITS Item ID	Set Leader	AIR Set Leader ID	Task Type	Task Sub Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	416	1	VH1X5899_152				Read-Along Sentence			
2	424	1	VH1X0543_126				Read and Match	Word		
3	425	1	VH180X28_184				Read and Match	Sentence		
4	440	1 of 2	VH20X001_114	VH201024_112			Procedural Text			
5	500	1	VHX98893_172	VH181843_113			Short Correspondence			
6	500	1	VH18X625_198				Read and Match	Sentence		
7	500	2 of 2	VH203X01_114	VH201024_112			Procedural Text			
8	500	1	VH18X026_188	VH182011_116			Short Correspondence			

Figure 4. Sample Item Map

The panelists discuss each item in the OIB. In particular, they will discuss the knowledge, skills, and processes being measured by each item as well as why the item is more difficult than the items that precede it, as indicated in Figure 5.

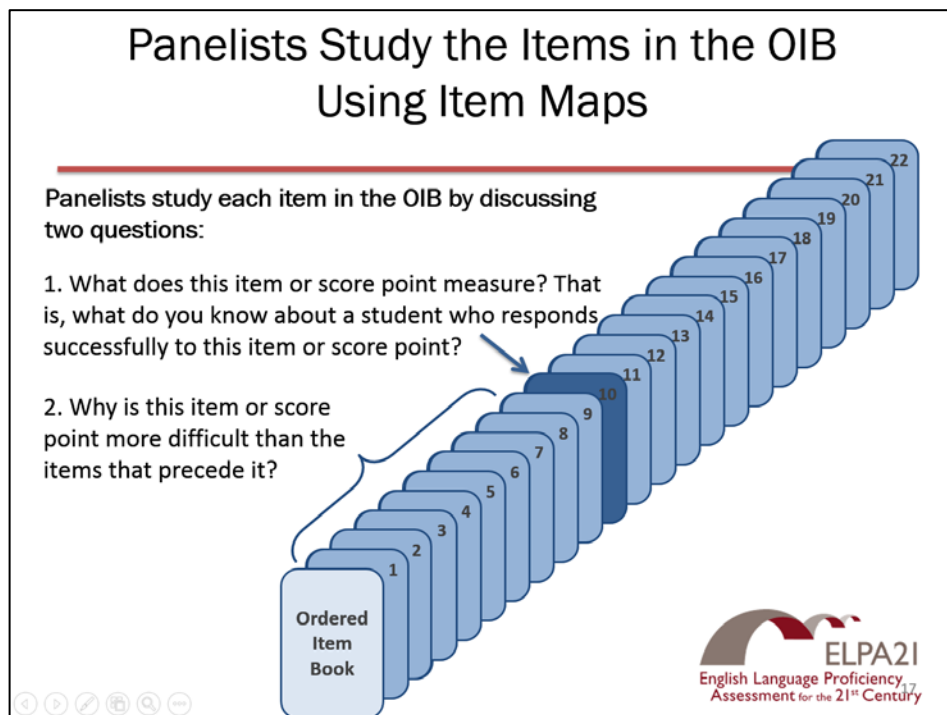


Figure 5. Studying the Ordered Item Booklet

Training panelists on the Bookmarking task

Training begins by focusing panelists on the student just at the Level 4 cut score—the Target Student and follows with training on placing bookmarks.

Target Student Discussion. The Level 4 Target Student is the student on the threshold of an achievement level. Understanding the Level 4 Target Student helps differentiate the skills associated with just leaving Level 3 and just entering Level 4. Panelists will review Target Student descriptors for Level 4 and Level 3 and will have an opportunity to discuss them with the EL Expert.

When the Target Student discussion is complete, panelists will engage in bookmark training.

Bookmark Training. Panelists will be trained to place bookmarks to reflect their recommendations. Figure 6 describes the fundamental task for Level 4. Level 4 is the anchor level—Early Advanced.

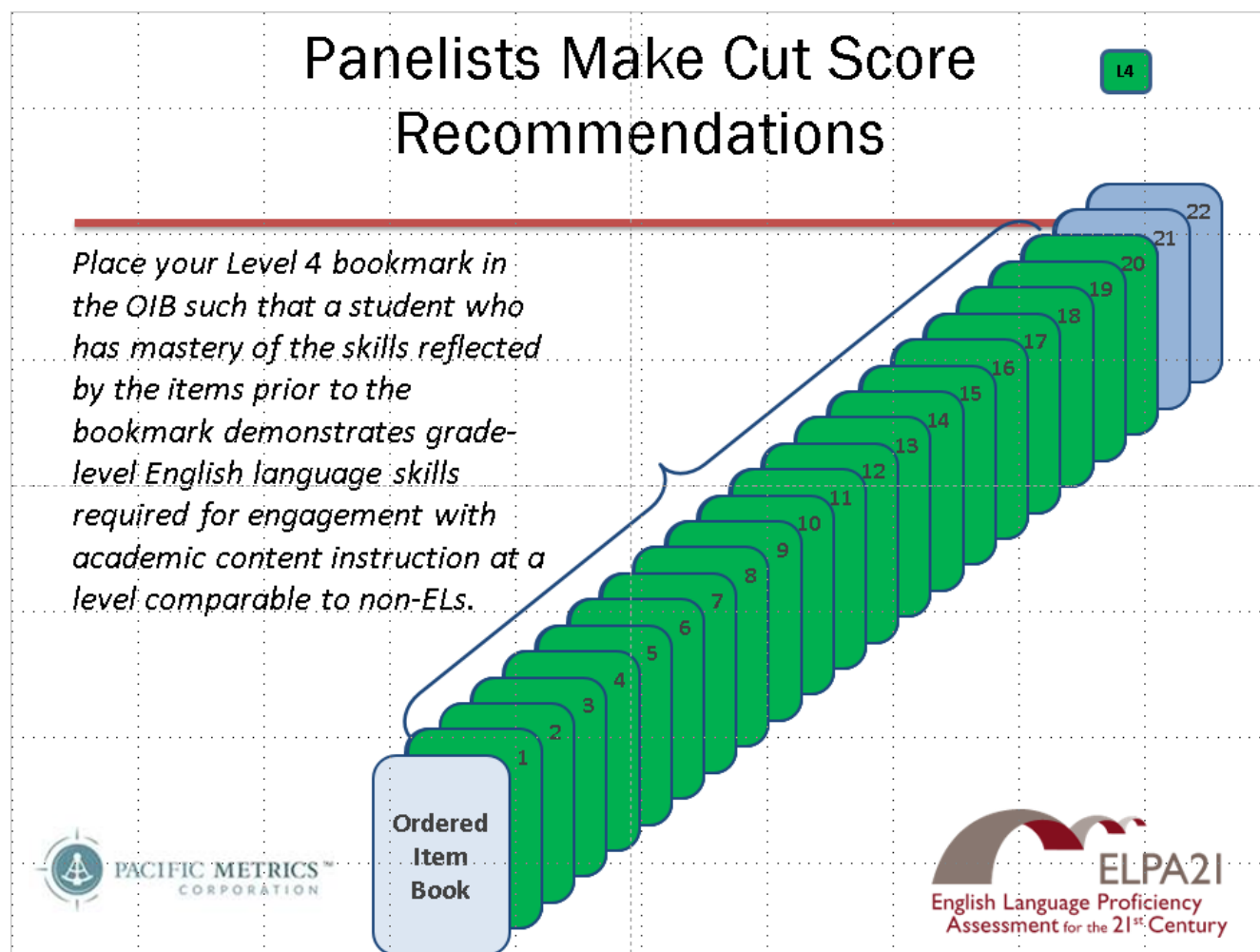


Figure 6. Bookmark Training Slide

Multiple methods are used to train panelists. Using Figure 6 as a starting point, the Workshop Facilitators will take an ordered item booklet and place a bookmark in the OIB and interpret the bookmark.

“If you place your Level 4 bookmark here you will notice that it divides the OIB into two parts. Your bookmark indicates that you believe that a student who has mastery of the skills reflected by the items before the bookmark demonstrates grade-level English language skills required for engagement with academic content instruction at a level comparable to non-ELs.

You know it is in the right location if you go a few items back and think ‘No, they need those few additional items to meet the stated criteria to be in Level 4,’ or you go a few items beyond the bookmark and think ‘It would be nice if they had these additional skills, but if a student didn’t have strong mastery of those items they would still meet the criteria.’

You will have opportunities to adjust your bookmarks in Rounds 2 and 3 so place your bookmark at the location you think is best and fine tune your thinking after discussions with your fellow panelists and after seeing other relevant data.”

recommendations either reinforces a panelist's commitment to their Round 1 ratings or provides a perspective that results in modifying their original ratings.

The panelists will discuss the information provided. They will be instructed to ground their conversations in terms of content in the OIB, specifically with respect to the skills students should have to be in each level. Panelists are not required to reach consensus during this process. Following discussion, the panelists make their Round 2 ratings.

Procedural Details. Table Facilitators open the Round 2 Display tab in the Excel worksheet shown in Figure 7 so it is displayed on the monitor and follow the directions. The Table Facilitator calls out each indicated page in the OIB for the Level 4 bookmark, making sure that each panelist places a green post-it on each page called out.

Sample Interpretation: Notice that we did not all have the same bookmark placements. We all agreed that the items before the first of our bookmarks (Page 15 in Figure 7) should be mastered to be in Level 4, and we all agreed that no items after the last of our bookmarks (Page 30 in Figure 7) need to be mastered to be in Level 4. So we only need to discuss our differences for items between the first and last of our bookmarks (items 15–29 using data from Figure 7). Let's discuss our rationales for our bookmarks.

Let's also note the CGS bookmark and consider that in our discussion. We can discuss what was said about the CGS study in the opening session if you like and ask the Workshop Facilitators for clarification if needed.

Contrasting Groups Study. The CGS is a study conducted by Pacific Metrics for ELPA21. EL teachers in ELPA21 states classified their EL students into 5 levels of achievement based on the ELPA21 Achievement Level Policy Descriptors. These teachers classified each of their students into one of five levels for each domain and overall. Pacific Metrics analyzed this data and by associating teachers' classifications with the operational ELPA21 test results for these students we are able to estimate cut scores and bookmarks associated with the teachers' judgments.

The results of the CGS are provided to panelists by translating the CGS cut score estimates into bookmarks in the OIBs so that panelists can see the CGS results—how educators throughout ELPA21 states placed their students—expressed in the same metric used to express their own ratings.

Panelists discuss their differences and make a Round 2 Level 4 bookmark rating based on their updated understanding following discussion. Then they repeat this activity for Level 3.

Panelists enter their Levels 3 and 4 bookmarks on their rating forms and provide them to the Table Facilitator, who enters them into the Excel Round 2 Bookmark Ratings Worksheet as they did in Round 1.

The display materials needed for Round 3 are populated in the next tab in the Excel file. A screenshot of a draft Round 3 Display Worksheet is provided in Figure 8.

	A	B	C	D	E	F	G	H
1				Data based on R2 Bookmarks		Level 3 Bookmark	Level 4 Bookmark	
2				Table 1 Median		4	13	
3				Table 2 Median		3	12	
4				Grade Band (T1 & T2) Median		3.5	12	
5				CGS		?	?	
6								
7					Estimated percent of students based on R2 Results			
8					Levels 1 + 2	Level 3	Levels 4 + 5	
9			Grade Band (T1 & T2)		x	x	x	
10		Table Faciliator Visual Check						
11		Table 1 Level 3 Bookmarks:	4	4	4	3	3	
12		Table 1 Level 4 Bookmarks:	12	12	13	13	13	
13								
14		Table 2 Level 3 Bookmarks:	3	3	3	4	4	
15		Table 2 Level 4 Bookmarks:	11	11	12	12	12	
16								

Figure 8. Draft Round 3 Display Worksheet

Round 3

Round 3 begins with a presentation by the Workshop Facilitators with respect to the tables joining together and the impact data being presented. The Workshop Facilitators will convey to the panelists that they have made their decisions based purely on a criterion-referenced basis using their own judgment and listening to the opinions of their fellow panelists. Now additional information will be presented. First, the two tables in each grade band come together to form a single grade band group to make their final ratings. As such, they will see the median bookmarks for each table and for all panelists in the grade band.

Next, they will see the impact data—the percentage of students who would be classified in each achievement level associated with the median bookmark for all panelists. The Bookmark Procedure is considered a criterion-referenced exercise; however, it is considered best practice to bring norm-referenced information into consideration prior to the final round of ratings to provide a reality check to panelists.

“If the data seem reasonable based on the median bookmark location, then it supports the validity of that bookmark. If the data seem unusual to you—too rigorous or not rigorous enough—then it should prompt you to look again at the bookmarks, and it provides a direction for further discussion that may result in reconsidering your judgments for Round 3. In either case, you should make decisions based on content and not solely on impact data. We do not want you to ‘chase numbers.’ However, the impact data should also not be a surprise—you know the students and what they can do based on your experience. You placed bookmarks based on your expectations and should have some idea of how well students will meet them. Of course this data is from multiple states and none of you is familiar with students across all the states.

As your Table Facilitator shows you this information for your grade band, if you are very surprised by the data, call a Workshop Facilitator to your table to help frame the discussion. Otherwise, discuss the differences between the table’s median bookmarks and the CGS bookmarks. Begin with Level 4, indicate your Round 3 ratings on the rating form, and then repeat the process for Level 3.”

As described earlier, you, the Table Facilitator, will do a visual check with the other Table Facilitator in your grade band to be sure you are looking at the same information on your screens. Your excel files are linked and we want to make sure that the links have update correctly. Thus, the data at the bottom of the screen should correctly represent your tables' bookmarks. If they do not, call a Workshop Facilitator to provide a quick fix.

Workshop Facilitators will float between the tables during panelists' discussions. Following discussion, the panelists will make their Round 3 ratings, which Table Facilitators will enter into the Excel Round 3 Ratings Worksheet. Data will be compiled and panelists will be provided with their results (bookmarks and associated impact data) when they are available, but there is no further discussion of these results until the vertical articulation session.

This process is repeated for each of the domains.

Daily Secure Materials Collection

Materials will be collected in a systematic manner at the close of each domain and/or day to support the tracking of secure materials. Table Facilitators will observe their table's panelists stack each piece of color coded (secure) materials in a specified order called out by the Table Facilitator. The Table Facilitators will initial the tracking form after their review, and materials will be collected by Pacific Metrics staff and stored in the secure Operations Room. A list of all secure materials will be provided to the Table Facilitator to support this systematic collection of materials.

Appendix E2. Bookmark Rating Form



ELPA21 Standard Setting

Bookmark Rating Form

Name: _____

Circle Domain: **Reading** **Listening** **Writing** **Speaking**

Grade: _____

Table: _____

Panelist #: _____

Round 1

Level 3 Bookmark

Level 4 Bookmark

Round 2

Level 3 Bookmark

Level 4 Bookmark

Round 3

Level 3 Bookmark

Level 4 Bookmark

Appendix E3. Bookmarks

Level 3 Bookmark

Level 3 Bookmark

Level 3 Bookmark



**ELPA21
Standard Setting**

**ELPA21
Standard Setting**

**ELPA21
Standard Setting**

**Level 3
Bookmark**

**Level 3
Bookmark**

**Level 3
Bookmark**

I am placing my Level 3 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark applies **some** grade-level English language skills and will benefit from EL Program support.

I am placing my Level 3 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark applies **some** grade-level English language skills and will benefit from EL Program support.

I am placing my Level 3 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark applies **some** grade-level English language skills and will benefit from EL Program support.

Level 3 Bookmark

Level 3 Bookmark

Level 3 Bookmark

Level 4 Bookmark



**ELPA21
Standard Setting**

Level 4 Bookmark

I am placing my Level 4 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark should be able to demonstrate English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.

Level 4 Bookmark

Level 4 Bookmark



**ELPA21
Standard Setting**

Level 4 Bookmark

I am placing my Level 4 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark should be able to demonstrate English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.

Level 4 Bookmark

Level 4 Bookmark



**ELPA21
Standard Setting**

Level 4 Bookmark

I am placing my Level 4 bookmark here because a student who has mastery of the English language skills reflected by the items before the bookmark should be able to demonstrate English language skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs.

Level 4 Bookmark

Appendix E4. Sample Item Map

ELPA21 Standard Setting July 2016

Grade 3 Reading Item Map

Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID__AIR ITS Item ID	Set Leader__ AIR Set Leader ID	Task Type	Task Sub Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	381	1	VHxxxxxx__xxxx		Read-Along Sentence			
2	440	1 of 2	VHxxxxx__ _xxxx	VHxxxxx__ _xxx	Procedural Text			
3	453	1	VHxxxxx__ _xxxx		Read and Match	Sentence		
4	481	2 of 2	VHxxxxx__ _xxxx	VHxxxxx__ _xxx	Procedural Text			
5	485	1	VHxxxxx__ _xxxx	VHxxxxx__ _xxx	Short Correspondence			
6	510	1	VHxxxxx__ _xxxx		Read and Match	Word		
7	539	1	VHxxxxx__ _xxxx	VHxxxxx__ _xxx	Short Correspondence			
8	552	1	VHxxxxx__ _xxxx	VHxxxxx__ _xxx	Short Correspondence			

Appendix E5. Materials Summary

Materials provided to each panelist:

- Brief agenda of workshop events
- Achievement Level Indicators (for the associated grade band, for each domain)
- Policy Descriptors and Performance Targets by Domain (appended to this section)
- Target Student Descriptors (for the grade being set within the grade band, for each domain)
 - Target achievement level descriptors summarize the knowledge and skills held by students at the threshold of Level 4 within a domain and grade. The target ALDs will be developed by EL Experts and will be associated with the specific grade for which cut scores are being set directly by panelists.
- Ordered Item Booklet
- Item Map
- Stimulus Booklet
- Exemplars and Rubrics for Writing and Speaking (Speaking mp3 exemplars on table laptop)
- Bookmarks
 - Each panelist receives a set of Level 4 and Level 5 bookmarks to support their bookmark placement
- Rating Form
 - Paper rating forms are provided for panelists to will note their recommended bookmarks during each round for each domain.
- Readiness surveys
 - Readiness surveys are used to ascertain whether panelists are feel prepared to engage in subsequent activities based on training received. Readiness surveys are administered at the close of the opening session and again before the first round of ratings.
- Evaluations.
 - Evaluations provide an opportunity to gain feedback and assess the validity of the process. That is, the validity of the standard setting is supported by evidence for “procedural validity.” Evidence of procedural validity will be gathered through process evaluation questionnaires administered to panelists after they have completed the work for each domain including a summative/cross-grade activity evaluation at the close of the workshop.

At each table

- English Language Proficiency Standards
- Excel files used by the Table Facilitators to transfer panelists’ ratings to the laptop provided at each table by Pacific Metrics. The excel file, upon saving, will be accessible to the Pacific Metrics Operations Room Manager via Dropbox.
- Materials tracking list
 - A list associating panelists with coded materials
- Supplies (e.g., pencils, post-its, etc.)

Policy Definitions for Five Performance Levels

A student at this level...	
Level 1: Beginning	Displays few grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of developing grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies some grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates skills required for the autonomous engagement with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits superior English language skills, as measured by ELPA21.

Performance Targets by Domain

Domain	Definition
Listening	An EL can listen and comprehend spoken English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Reading	An EL can read and comprehend written English at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can produce speech at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can write texts at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

Appendix F. In-Person Bookmark Standard Setting Workshop Results

Appendix F1. Round by Round Results

<i>Round 1 Results for Kindergarten Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	22	34	467	524
1	2	27	34	495	524
1	3	22	34	467	524
1	4	27	34	495	524
2	6	17	31	458	513
2	7	27	41	495	566
2	8	22	38	467	544
2	9	30	42	507	572
Table 1	Median	24.5	34	482*	524*
	Minimum	22	34	467	524
	Maximum	27	34	495	524
Table 2	Median	24.5	39.5	482*	548*
	Minimum	17	31	458	513
	Maximum	30	42	507	572
Grade	Median	24.5	34	482*	524*
	Minimum	17	31	458	513
	Maximum	30	42	507	572

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Kindergarten Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	25	36	487	530		
1	2	27	34	495	524		
1	3	27	34	495	524		
1	4	27	34	495	524		
2	6	24	39	482	548		
2	7	25	41	487	566		
2	8	24	41	482	566		
2	9	27	39	495	548		
Table 1	Median	27	34	495*	524*		
	Minimum	25	34	487	524		
	Maximum	27	36	495	530		
Table 2	Median	24.5	40	482*	553*		
	Minimum	24	39	482	548		
	Maximum	27	41	495	566		
Grade	Median	26	37.5	489*	539*		
	Minimum	24	34	482	524		
	Maximum	27	41	495	566		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.39	0.35	0.26

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Kindergarten Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	24	38	482	544		
1	2	24	38	482	544		
1	3	24	36	482	530		
1	4	24	35	482	529		
2	6	24	39	482	548		
2	7	24	36	482	530		
2	8	24	40	482	553		
2	9	27	39	495	548		
Table 1	Median	24	37	482*	539*		
	Minimum	24	35	482	529		
	Maximum	24	38	482	544		
Table 2	Median	24	39	482*	548*		
	Minimum	24	36	482	530		
	Maximum	27	40	495	553		
Grade	Median	24	38	482*	544*		
	Minimum	24	35	482	529		
	Maximum	27	40	495	553		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.32	0.48	0.19

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Kindergarten Reading</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	23	30	469	495
1	2	22	37	468	541
1	3	29	36	491	532
1	4	27	33	479	513
2	6	23	32	469	505
2	7	16	16	446	446
2	8	15	33	442	513
2	9	12	33	440	513
Table 1	Median	25	34.5	473*	523*
	Minimum	22	30	468	495
	Maximum	29	37	491	541
Table 2	Median	15.5	32.5	442*	505*
	Minimum	12	16	440	446
	Maximum	23	33	469	513
Grade	Median	22.5	33	468*	513*
	Minimum	12	16	440	446
	Maximum	29	37	491	541

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Kindergarten Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	26	33	473	513		
1	2	27	37	479	541		
1	3	23	33	469	513		
1	4	26	33	473	513		
2	6	23	37	469	541		
2	7	23	37	469	541		
2	8	23	33	469	513		
2	9	23	39	469	551		
Table 1	Median	26	33	473*	513*		
	Minimum	23	33	469	513		
	Maximum	27	37	479	541		
Table 2	Median	23	37	469*	541*		
	Minimum	23	33	469	513		
	Maximum	23	39	469	551		
Grade	Median	23	35	469*	527*		
	Minimum	23	33	469	513		
	Maximum	27	39	479	551		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.23	0.43	0.34

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Kindergarten Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	23	35	469	527		
1	2	23	37	469	541		
1	3	25	35	473	527		
1	4	26	34	473	523		
2	6	26	37	473	541		
2	7	23	39	469	551		
2	8	23	37	469	541		
2	9	23	39	469	551		
Table 1	Median	24	35	469*	527*		
	Minimum	23	34	469	523		
	Maximum	26	37	473	541		
Table 2	Median	23	38	469*	542*		
	Minimum	23	37	469	541		
	Maximum	26	39	473	551		
Grade	Median	23	37	469*	541*		
	Minimum	23	34	469	523		
	Maximum	26	39	473	551		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.23	0.51	0.26

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Kindergarten Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	22	36	479	507
1	2	23	46	480	558
1	3	22	38	479	517
1	4	27	43	488	541
2	6	42	48	537	566
2	7	42	50	537	569
2	8	40	50	521	569
2	9	36	54	507	622
Table 1	Median	22.5	40.5	478*	521*
	Minimum	22	36	479	507
	Maximum	27	46	488	558
Table 2	Median	41	50	533*	568*
	Minimum	36	48	507	566
	Maximum	42	54	537	622
Grade	Median	31.5	47	499*	560*
	Minimum	22	36	479	507
	Maximum	42	54	537	622

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Kindergarten Speaking								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	26	42	486	537			
1	2	36	46	507	558			
1	3	26	42	486	537			
1	4	26	45	486	556			
2	6	29	42	494	537			
2	7	33	43	501	541			
2	8	29	42	494	537			
2	9	34	44	506	551			
Table 1	Median	26	43.5	485*	541*			
	Minimum	26	42	486	537			
	Maximum	36	46	507	558			
Table 2	Median	31	42.5	499*	537*			
	Minimum	29	42	494	537	Associated Impact Data*		
	Maximum	34	44	506	551	% L1 + L2	% L3	% L4 + L5
Grade	Median	29	42.5	493*	537*	0.36	0.37	0.27
	Minimum	26	42	486	537			
	Maximum	36	46	507	558			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Kindergarten Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	29	43	494	541		
1	2	30	43	498	541		
1	3	29	43	494	541		
1	4	28	45	493	556		
2	6	29	43	494	541		
2	7	29	43	494	541		
2	8	33	46	501	558		
2	9	30	44	498	551		
Table 1	Median	29	43	493*	541*		
	Minimum	28	43	493	541		
	Maximum	30	45	498	556		
Table 2	Median	29.5	43.5	493*	541*		
	Minimum	29	43	494	541		
	Maximum	33	46	501	558		
Grade	Median	29	43	493*	541*		
	Minimum	28	43	493	541		
	Maximum	33	46	501	558		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.36	0.41	0.23

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Kindergarten Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	25	33	491	503
1	2	37	47	511	554
1	3	31	42	499	527
1	4	24	36	490	506
2	6	33	43	503	539
2	7	41	48	526	573
2	8	14	40	458	521
2	9	25	42	491	527
Table 1	Median	28	39	493*	514*
	Minimum	24	33	490	503
	Maximum	37	47	511	554
Table 2	Median	29	42.5	496*	527*
	Minimum	14	40	458	521
	Maximum	41	48	526	573
Grade	Median	28	42	493*	527*
	Minimum	14	33	458	503
	Maximum	41	48	526	573

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Kindergarten Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	30	42	499	527		
1	2	31	43	499	539		
1	3	31	42	499	527		
1	4	26	37	492	511		
2	6	33	48	503	573		
2	7	35	48	506	573		
2	8	33	48	503	573		
2	9	33	48	503	573		
Table 1	Median	30.5	42	499*	527*		
	Minimum	26	37	492	511		
	Maximum	31	43	499	539		
Table 2	Median	33	48	503*	573*		
	Minimum	33	48	503	573		
	Maximum	35	48	506	573		
Grade	Median	32	45.5	501*	545*		
	Minimum	26	37	492	511		
	Maximum	35	48	506	573		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.59	0.26	0.15

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Kindergarten Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	33	47	503	554		
1	2	33	47	503	554		
1	3	33	48	503	573		
1	4	27	39	492	514		
2	6	33	48	503	573		
2	7	33	48	503	573		
2	8	33	48	503	573		
2	9	33	48	503	573		
Table 1	Median	33	47	503*	554*		
	Minimum	27	39	492	514		
	Maximum	33	48	503	573		
Table 2	Median	33	48	503*	573*		
	Minimum	33	48	503	573		
	Maximum	33	48	503	573		
Grade	Median	33	48	503*	573*		
	Minimum	27	39	492	514		
	Maximum	33	48	503	573		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.60	0.30	0.10

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 1 Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	20	36	428	457
1	2	24	38	431	465
1	3	13	30	417	437
1	4	19	41	428	485
2	6	26	39	433	468
2	7	24	36	431	457
2	8	15	33	420	442
2	9	24	36	431	457
Table 1	Median	19.5	37	428*	462*
	Minimum	13	30	417	437
	Maximum	24	41	431	485
Table 2	Median	24	36	431*	457*
	Minimum	15	33	420	442
	Maximum	26	39	433	468
Grade	Median	22	36	431*	457*
	Minimum	13	30	417	437
	Maximum	26	41	433	485

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 1 Listening						
		Bookmarks		Associated Cut Scores		
Table	Panelist	Level 3	Level 4	Level 3	Level 4	
1	1	20	36	428	457	
1	2	24	39	431	468	
1	3	24	38	431	465	
1	4	24	36	431	457	
2	6	24	36	431	457	
2	7	24	38	431	465	
2	8	25	39	432	468	
2	9	24	38	431	465	
Table 1	Median	24	37	431*	462*	
	Minimum	20	36	428	457	
	Maximum	24	39	431	468	
Table 2	Median	24	38	431*	465*	
	Minimum	24	36	431	457	
	Maximum	25	39	432	468	
Grade	Median	24	38	431*	465*	
	Minimum	20	36	428	457	
	Maximum	25	39	432	468	

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.07	0.11	0.81

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 1 Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	34	42	444	504		
1	2	36	48	457	567		
1	3	34	48	444	567		
1	4	36	48	457	567		
2	6	36	48	457	567		
2	7	34	42	444	504		
2	8	34	42	444	504		
2	9	26	42	433	504		
Table 1	Median	35	48	447*	567*		
	Minimum	34	42	444	504		
	Maximum	36	48	457	567		
Table 2	Median	34	42	444*	504*		
	Minimum	26	42	433	504		
	Maximum	36	48	457	567		
Grade	Median	34	45	444*	527*		
	Minimum	26	42	433	504		
	Maximum	36	48	457	567		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.11	0.44	0.45

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 1 Results for Grade 1 Reading					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	13	33	436	509
1	2	23	36	472	527
1	3	19	33	462	509
1	4	23	29	472	499
2	6	29	46	499	581
2	7	13	32	436	509
2	8	20	27	466	488
2	9	13	23	436	472
Table 1	Median	21	33	468*	509*
	Minimum	13	29	436	499
	Maximum	23	36	472	527
Table 2	Median	16.5	29.5	453*	499*
	Minimum	13	23	436	472
	Maximum	29	46	499	581
Grade	Median	19.5	32.5	462*	509*
	Minimum	13	23	436	472
	Maximum	29	46	499	581

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 1 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	23	35	472	521		
1	2	23	35	472	521		
1	3	23	35	472	521		
1	4	23	35	472	521		
2	6	23	36	472	527		
2	7	13	32	436	509		
2	8	20	29	466	499		
2	9	20	29	466	499		
Table 1	Median	23	35	472*	521*		
	Minimum	23	35	472	521		
	Maximum	23	35	472	521		
Table 2	Median	20	30.5	466*	501*		
	Minimum	13	29	436	499		
	Maximum	23	36	472	527		
Grade	Median	23	35	472*	521*		
	Minimum	13	29	436	499		
	Maximum	23	36	472	527		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.28	0.40	0.32

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 1 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	23	35	472	521		
1	2	23	36	472	527		
1	3	23	35	472	521		
1	4	23	36	472	527		
2	6	23	36	472	527		
2	7	20	35	466	521		
2	8	21	34	468	520		
2	9	23	32	472	509		
Table 1	Median	23	35.5	472*	521*		
	Minimum	23	35	472	521		
	Maximum	23	36	472	527		
Table 2	Median	22	34.5	470*	520*		
	Minimum	20	32	466	509		
	Maximum	23	36	472	527		
Grade	Median	23	35	472*	521*		
	Minimum	20	32	466	509		
	Maximum	23	36	472	527		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.28	0.40	0.32

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 1 Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	31	51	450	530
1	2	38	56	479	574
1	3	33	55	458	554
1	4	41	58	483	626
2	6	56	59	574	640
2	7	56	59	574	640
2	8	56	59	574	640
2	9	43	58	500	626
Table 1	Median	35.5	55.5	465*	554*
	Minimum	31	51	450	530
	Maximum	41	58	483	626
Table 2	Median	56	59	574*	640*
	Minimum	43	58	500	626
	Maximum	56	59	574	640
Grade	Median	42	58	491*	626*
	Minimum	31	51	450	530
	Maximum	56	59	574	640

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 1 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	38	56	479	574		
1	2	38	56	479	574		
1	3	33	55	458	554		
1	4	38	58	479	626		
2	6	56	59	574	640		
2	7	56	59	574	640		
2	8	56	59	574	640		
2	9	56	58	574	626		
Table 1	Median	38	56	479*	574*		
	Minimum	33	55	458	554		
	Maximum	38	58	479	626		
Table 2	Median	56	59	574*	640*		
	Minimum	56	58	574	626		
	Maximum	56	59	574	640		
Grade	Median	47	58	515*	626*		
	Minimum	33	55	458	554		
	Maximum	56	59	574	640		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.55	0.45	0.00

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 1 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	49	56	525	574		
1	2	38	56	479	574		
1	3	33	55	458	554		
1	4	56	59	574	640		
2	6	56	59	574	640		
2	7	48	56	522	574		
2	8	56	59	574	640		
2	9	50	58	525	626		
Table 1	Median	43.5	56	500*	574*		
	Minimum	33	55	458	554		
	Maximum	56	59	574	640		
Table 2	Median	53	58.5	540*	626*		
	Minimum	48	56	522	574		
	Maximum	56	59	574	640		
Grade	Median	49.5	57	525*	600*		
	Minimum	33	55	458	554		
	Maximum	56	59	574	640		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.68	0.30	0.02

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 1 Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	42	55	475	600
1	2	33	48	457	519
1	3	33	48	457	519
1	4	42	55	475	600
2	6	48	55	519	600
2	7	39	51	468	534
2	8	35	46	461	491
2	9	33	48	457	519
Table 1	Median	37.5	51.5	465*	534*
	Minimum	33	48	457	519
	Maximum	42	55	475	600
Table 2	Median	37	49.5	465*	522*
	Minimum	33	46	457	491
	Maximum	48	55	519	600
Grade	Median	37	49.5	465*	522*
	Minimum	33	46	457	491
	Maximum	48	55	519	600

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 1 Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	42	55	475	600		
1	2	33	51	457	534		
1	3	33	48	457	519		
1	4	42	55	475	600		
2	6	48	55	519	600		
2	7	48	55	519	600		
2	8	48	51	519	534		
2	9	48	51	519	534		
Table 1	Median	37.5	53	465*	540*		
	Minimum	33	48	457	519		
	Maximum	42	55	475	600		
Table 2	Median	48	53	519*	540*		
	Minimum	48	51	519	534		
	Maximum	48	55	519	600		
Grade	Median	45	53	484*	540*	% L1 + L2	% L3
	Minimum	33	48	457	519		
	Maximum	48	55	519	600		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.31	0.52	0.17

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 1 Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	48	55	519	600		
1	2	39	51	468	534		
1	3	35	48	461	519		
1	4	48	55	519	600		
2	6	48	55	519	600		
2	7	48	55	519	600		
2	8	48	51	519	534		
2	9	48	51	519	534		
Table 1	Median	43.5	53	478*	540*		
	Minimum	35	48	461	519		
	Maximum	48	55	519	600		
Table 2	Median	48	53	519*	540*		
	Minimum	48	51	519	534		
	Maximum	48	55	519	600		
Grade	Median	48	53	519*	540*	% L1 + L2	% L3
	Minimum	35	48	461	519		
	Maximum	48	55	519	600		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.42	0.41	0.17

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 3 Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	27	35	459	494
1	2	17	34	429	485
1	3	18	30	436	472
1	4	26	30	458	472
2	6	15	30	424	472
2	7	15	27	424	459
2	8	18	30	436	472
2	9	22	41	444	510
Table 1	Median	22	32	444*	474*
	Minimum	17	30	429	472
	Maximum	27	35	459	494
Table 2	Median	16.5	30	428*	472*
	Minimum	15	27	424	459
	Maximum	22	41	444	510
Grade	Median	18	30	436*	472*
	Minimum	15	27	424	459
	Maximum	27	41	459	510

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 3 Listening								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	19	34	437	485			
1	2	19	34	437	485			
1	3	19	34	437	485			
1	4	19	34	437	485			
2	6	18	39	436	505			
2	7	18	39	436	505			
2	8	18	37	436	496			
2	9	19	39	437	505			
Table 1	Median	19	34	437*	485*			
	Minimum	19	34	437	485			
	Maximum	19	34	437	485			
Table 2	Median	18	39	436*	505*			
	Minimum	18	37	436	496	Associated Impact Data*		
	Maximum	19	39	437	505	% L1 + L2	% L3	% L4 + L5
Grade	Median	19	35.5	437*	494*	0.05	0.14	0.81
	Minimum	18	34	436	485			
	Maximum	19	39	437	505			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 3 Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	21	43	442	521		
1	2	21	43	442	521		
1	3	21	40	442	507		
1	4	22	41	444	510		
2	6	21	43	442	521		
2	7	21	43	442	521		
2	8	21	43	442	521		
2	9	21	43	442	521		
Table 1	Median	21	42	442*	510*		
	Minimum	21	40	442	507		
	Maximum	22	43	444	521		
Table 2	Median	21	43	442*	521*		
	Minimum	21	43	442	521		
	Maximum	21	43	442	521		
Grade	Median	21	43	442*	521*		
	Minimum	21	40	442	507		
	Maximum	22	43	444	521		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.05	0.31	0.64

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 3 Reading</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	29	41	517	552
1	2	11	21	443	495
1	3	28	41	514	552
1	4	18	33	485	532
2	6	12	33	453	532
2	7	33	37	532	544
2	8	15	32	474	528
2	9	20	34	493	539
Table 1	Median	23	37	505*	544*
	Minimum	11	21	443	495
	Maximum	29	41	517	552
Table 2	Median	17.5	33.5	481*	532*
	Minimum	12	32	453	528
	Maximum	33	37	532	544
Grade	Median	19	33.5	487*	532*
	Minimum	11	21	443	495
	Maximum	33	41	532	552

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 3 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	24	41	510	552		
1	2	23	41	505	552		
1	3	21	41	495	552		
1	4	21	41	495	552		
2	6	20	33	493	532		
2	7	14	33	467	532		
2	8	15	33	474	532		
2	9	20	33	493	532		
Table 1	Median	22	41	501*	552*		
	Minimum	21	41	495	552		
	Maximum	24	41	510	552		
Table 2	Median	17.5	33	481*	532*		
	Minimum	14	33	467	532		
	Maximum	20	33	493	532		
Grade	Median	20.5	37	493*	544*		
	Minimum	14	33	467	532		
	Maximum	24	41	510	552		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.31	0.42	0.27

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 3 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	21	38	495	547		
1	2	20	37	493	544		
1	3	21	38	495	547		
1	4	20	33	493	532		
2	6	21	38	495	547		
2	7	21	33	495	532		
2	8	21	33	495	532		
2	9	21	36	495	543		
Table 1	Median	20.5	37.5	493*	544*		
	Minimum	20	33	493	532		
	Maximum	21	38	495	547		
Table 2	Median	21	34.5	495*	539*		
	Minimum	21	33	495	532		
	Maximum	21	38	495	547		
Grade	Median	21	36.5	495*	543*		
	Minimum	20	33	493	532		
	Maximum	21	38	495	547		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.35	0.35	0.30

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 3 Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	26	41	456	500
1	2	21	38	427	483
1	3	29	43	459	514
1	4	29	41	459	500
2	6	30	40	464	499
2	7	35	43	478	514
2	8	40	48	499	535
2	9	31	45	466	519
Table 1	Median	27.5	41	457*	500*
	Minimum	21	38	427	483
	Maximum	29	43	459	514
Table 2	Median	33	44	474*	516*
	Minimum	30	40	464	499
	Maximum	40	48	499	535
Grade	Median	29.5	42	459*	507*
	Minimum	21	38	427	483
	Maximum	40	48	499	535

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 3 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	26	45	456	519		
1	2	25	39	456	488		
1	3	25	40	456	499		
1	4	25	39	456	488		
2	6	35	45	478	519		
2	7	35	44	478	516		
2	8	35	45	478	519		
2	9	35	45	478	519		
Table 1	Median	25	39.5	456*	488*		
	Minimum	25	39	456	488		
	Maximum	26	45	456	519		
Table 2	Median	35	45	478*	519*		
	Minimum	35	44	478	516		
	Maximum	35	45	478	519		
Grade	Median	30.5	44.5	464*	516*		
	Minimum	25	39	456	488		
	Maximum	35	45	478	519		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.11	0.41	0.48

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 3 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	32	45	472	519		
1	2	34	44	474	516		
1	3	31	43	466	514		
1	4	35	45	478	519		
2	6	35	45	478	519		
2	7	35	43	478	514		
2	8	33	45	474	519		
2	9	35	45	478	519		
Table 1	Median	33	44.5	474*	516*		
	Minimum	31	43	466	514		
	Maximum	35	45	478	519		
Table 2	Median	35	45	478*	519*		
	Minimum	33	43	474	514		
	Maximum	35	45	478	519		
Grade	Median	34.5	45	474*	519*		
	Minimum	31	43	466	514		
	Maximum	35	45	478	519		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.18	0.36	0.46

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 3 Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	24	39	488	540
1	2	21	43	468	556
1	3	21	43	468	556
1	4	21	44	468	560
2	6	28	33	498	521
2	7	22	33	484	521
2	8	28	34	498	529
2	9	27	36	493	530
Table 1	Median	21	43	468*	556*
	Minimum	21	39	468	540
	Maximum	24	44	488	560
Table 2	Median	27.5	33.5	493*	521*
	Minimum	22	33	484	521
	Maximum	28	36	498	530
Grade	Median	23	37.5	487*	533*
	Minimum	21	33	468	521
	Maximum	28	44	498	560

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 3 Writing								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	24	43	488	556			
1	2	22	43	484	556			
1	3	22	43	484	556			
1	4	23	43	487	556			
2	6	28	33	498	521			
2	7	28	33	498	521			
2	8	28	33	498	521			
2	9	28	34	498	529			
Table 1	Median	22.5	43	484*	556*			
	Minimum	22	43	484	556			
	Maximum	24	43	488	556			
Table 2	Median	28	33	498*	521*			
	Minimum	28	33	498	521			
	Maximum	28	34	498	529			
Grade	Median	26	38.5	493*	537*			
	Minimum	22	33	484	521			
	Maximum	28	43	498	556			

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.32	0.39	0.28

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 3 Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	26	39	493	540		
1	2	27	38	493	537		
1	3	26	37	493	533		
1	4	23	39	487	540		
2	6	28	36	498	530		
2	7	28	37	498	533		
2	8	28	35	498	530		
2	9	26	37	493	533		
Table 1	Median	26	38.5	493*	537*		
	Minimum	23	37	487	533		
	Maximum	27	39	493	540		
Table 2	Median	28	36.5	498*	530*		
	Minimum	26	35	493	530		
	Maximum	28	37	498	533		
Grade	Median	26.5	37	493*	533*		
	Minimum	23	35	487	530		
	Maximum	28	39	498	540		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.32	0.37	0.31

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 5 Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	24	31	431	450
1	2	17	30	425	448
1	3	20	37	429	471
1	4	26	37	436	471
2	6	30	39	448	487
2	7	22	37	430	471
2	8	15	39	423	487
2	9	13	36	407	467
Table 1	Median	22	34	430*	465*
	Minimum	17	30	425	448
	Maximum	26	37	436	471
Table 2	Median	18.5	38	426*	479*
	Minimum	13	36	407	467
	Maximum	30	39	448	487
Grade	Median	21	37	429*	471*
	Minimum	13	30	407	448
	Maximum	30	39	448	487

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 5 Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	24	37	431	471		
1	2	24	37	431	471		
1	3	23	37	430	471		
1	4	24	37	431	471		
2	6	22	39	430	487		
2	7	22	39	430	487		
2	8	22	39	430	487		
2	9	22	39	430	487		
Table 1	Median	24	37	431*	471*		
	Minimum	23	37	430	471		
	Maximum	24	37	431	471		
Table 2	Median	22	39	430*	487*		
	Minimum	22	39	430	487		
	Maximum	22	39	430	487		
Grade	Median	22.5	38	430*	479*		
	Minimum	22	37	430	471		
	Maximum	24	39	431	487		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.06	0.14	0.79

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 5 Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	24	39	431	487		
1	2	24	39	431	487		
1	3	24	39	431	487		
1	4	24	39	431	487		
2	6	24	39	431	487		
2	7	24	39	431	487		
2	8	24	39	431	487		
2	9	24	39	431	487		
Table 1	Median	24	39	431*	487*		
	Minimum	24	39	431	487		
	Maximum	24	39	431	487		
Table 2	Median	24	39	431*	487*		
	Minimum	24	39	431	487		
	Maximum	24	39	431	487		
Grade	Median	24	39	431*	487*		
	Minimum	24	39	431	487		
	Maximum	24	39	431	487		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.07	0.19	0.74

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 5 Reading</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	17	32	455	508
1	2	16	31	455	508
1	3	14	26	450	490
1	4	21	34	476	515
2	6	21	42	476	533
2	7	23	39	482	524
2	8	24	34	485	515
2	9	32	54	508	599
Table 1	Median	16.5	31.5	455*	508*
	Minimum	14	26	450	490
	Maximum	21	34	476	515
Table 2	Median	23.5	40.5	482*	526*
	Minimum	21	34	476	515
	Maximum	32	54	508	599
Grade	Median	21	34	476*	515*
	Minimum	14	26	450	490
	Maximum	32	54	508	599

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 5 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	18	31	460	508		
1	2	18	31	460	508		
1	3	18	31	460	508		
1	4	18	31	460	508		
2	6	23	42	482	533		
2	7	23	43	482	533		
2	8	23	45	482	541		
2	9	23	44	482	538		
Table 1	Median	18	31	460*	508*		
	Minimum	18	31	460	508		
	Maximum	18	31	460	508		
Table 2	Median	23	43.5	482*	533*		
	Minimum	23	42	482	533		
	Maximum	23	45	482	541		
Grade	Median	20.5	36.5	467*	518*		
	Minimum	18	31	460	508		
	Maximum	23	45	482	541		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.24	0.40	0.36

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 5 Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	21	35	476	515		
1	2	21	35	476	515		
1	3	20	35	467	515		
1	4	22	35	478	515		
2	6	21	35	476	515		
2	7	20	35	467	515		
2	8	21	35	476	515		
2	9	23	35	482	515		
Table 1	Median	21	35	476*	515*		
	Minimum	20	35	467	515		
	Maximum	22	35	478	515		
Table 2	Median	21	35	476*	515*		
	Minimum	20	35	467	515		
	Maximum	23	35	482	515		
Grade	Median	21	35	476*	515*		
	Minimum	20	35	467	515		
	Maximum	23	35	482	515		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.27	0.37	0.36

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 5 Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	28	47	439	530
1	2	34	48	464	534
1	3	33	51	455	561
1	4	34	48	464	534
2	6	34	48	464	534
2	7	32	47	445	530
2	8	34	48	464	534
2	9	36	47	469	530
Table 1	Median	33.5	48	455*	534*
	Minimum	28	47	439	530
	Maximum	34	51	464	561
Table 2	Median	34	47.5	464*	530*
	Minimum	32	47	445	530
	Maximum	36	48	469	534
Grade	Median	34	48	464*	534*
	Minimum	28	47	439	530
	Maximum	36	51	469	561

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 5 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	33	48	455	534		
1	2	34	48	464	534		
1	3	33	50	455	537		
1	4	34	50	464	537		
2	6	34	48	464	534		
2	7	34	48	464	534		
2	8	34	48	464	534		
2	9	36	48	469	534		
Table 1	Median	33.5	49	455*	537*		
	Minimum	33	48	455	534		
	Maximum	34	50	464	537		
Table 2	Median	34	48	464*	534*		
	Minimum	34	48	464	534		
	Maximum	36	48	469	534		
Grade	Median	34	48	464*	534*		
	Minimum	33	48	455	534		
	Maximum	36	50	469	537		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.15	0.64	0.22

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 5 Speaking								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	33	43	455	507			
1	2	33	43	455	507			
1	3	33	43	455	507			
1	4	33	43	455	507			
2	6	34	44	464	508			
2	7	34	43	464	507			
2	8	34	43	464	507			
2	9	35	45	466	513			
Table 1	Median	33	43	455*	507*			
	Minimum	33	43	455	507			
	Maximum	33	43	455	507			
Table 2	Median	34	43.5	464*	507*			
	Minimum	34	43	464	507	Associated Impact Data*		
	Maximum	35	45	466	513	% L1 + L2	% L3	% L4 + L5
Grade	Median	33.5	43	455*	507*	0.11	0.38	0.50
	Minimum	33	43	455	507			
	Maximum	35	45	466	513			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 5 Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	28	41	463	538
1	2	23	41	433	538
1	3	25	34	443	482
1	4	27	39	455	517
2	6	18	41	423	538
2	7	22	39	431	517
2	8	27	41	455	538
2	9	27	39	455	517
Table 1	Median	26	40	447*	522*
	Minimum	23	34	433	482
	Maximum	28	41	463	538
Table 2	Median	24.5	40	440*	522*
	Minimum	18	39	423	517
	Maximum	27	41	455	538
Grade	Median	26	40	447*	522*
	Minimum	18	34	423	482
	Maximum	28	41	463	538

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 5 Writing								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	27	41	455	538			
1	2	26	41	447	538			
1	3	27	40	455	522			
1	4	27	41	455	538			
2	6	27	39	455	517			
2	7	27	39	455	517			
2	8	27	39	455	517			
2	9	27	39	455	517			
Table 1	Median	27	41	455*	538*			
	Minimum	26	40	447	522			
	Maximum	27	41	455	538			
Table 2	Median	27	39	455*	517*	Associated Impact Data*		
	Minimum	27	39	455	517			
	Maximum	27	39	455	517	% L1 + L2	% L3	% L4 + L5
Grade	Median	27	39.5	455*	517*	0.14	0.49	0.37
	Minimum	26	39	447	517			
	Maximum	27	41	455	538			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 5 Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	27	41	455	538		
1	2	26	41	447	538		
1	3	27	40	455	522		
1	4	27	41	455	538		
2	6	27	39	455	517		
2	7	27	39	455	517		
2	8	27	39	455	517		
2	9	27	39	455	517		
Table 1	Median	27	41	455*	538*		
	Minimum	26	40	447	522		
	Maximum	27	41	455	538		
Table 2	Median	27	39	455*	517*		
	Minimum	27	39	455	517		
	Maximum	27	39	455	517		
Grade	Median	27	39.5	455*	517*		
	Minimum	26	39	447	517		
	Maximum	27	41	455	538		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.14	0.49	0.37

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 7 Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	12	29	387	445
1	2	17	35	411	458
1	3	18	38	414	471
1	4	17	35	411	458
1	5	25	39	440	471
2	6	19	42	415	477
2	7	37	45	467	511
2	8	26	37	442	467
2	9	24	43	439	482
2	10	37	43	467	482
Table 1	Median	17	35	411*	458*
	Minimum	12	29	387	445
	Maximum	18	38	414	471
Table 2	Median	25	42.5	440*	477*
	Minimum	19	37	415	467
	Maximum	37	45	467	511
Grade	Median	19	38	415*	471*
	Minimum	12	29	387	445
	Maximum	37	45	467	511

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 7 Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	18	37	414	467		
1	2	17	37	411	467		
1	3	18	38	414	471		
1	4	17	38	411	471		
1	5	18	37	414	467		
2	6	26	42	442	477		
2	7	29	45	445	511		
2	8	26	45	442	511		
2	9	27	45	443	511		
2	10	26	43	442	482		
Table 1	Median	17.5	37.5	411*	467*		
	Minimum	17	37	411	467		
	Maximum	18	38	414	471		
Table 2	Median	26.5	45	442*	511*		
	Minimum	26	42	442	477		
	Maximum	29	45	445	511		
Grade	Median	18	38	414*	471*	% L1 + L2	% L3
	Minimum	17	37	411	467		
	Maximum	29	45	445	511		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.11	0.21	0.68

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 7 Listening								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	26	45	442	511			
1	2	24	42	439	477			
1	3	25	43	440	482			
1	4	24	45	439	511			
1	5	26	43	442	482			
2	6	26	42	442	477			
2	7	27	45	443	511			
2	8	26	45	442	511			
2	9	27	45	443	511			
2	10	27	44	443	502			
Table 1	Median	24.5	44	439*	502*			
	Minimum	24	42	439	477			
	Maximum	26	45	442	511			
Table 2	Median	26.5	45	442*	511*			
	Minimum	26	42	442	477			
	Maximum	27	45	443	511			
Grade	Median	26	45	442*	511*			
	Minimum	24	42	439	477			
	Maximum	27	45	443	511			
						Associated Impact Data*		
						% L1 + L2	% L3	% L4 + L5
						0.15	0.26	0.59

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 7 Reading</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	18	33	476	515
1	2	19	32	477	512
1	3	16	32	473	512
1	4	11	19	462	477
1	5	20	33	484	515
2	6	10	43	459	560
2	7	11	27	462	499
2	8	23	32	492	512
2	9	25	44	494	561
2	10	25	35	494	517
Table 1	Median	17	32	475*	512*
	Minimum	11	19	462	477
	Maximum	19	33	477	515
Table 2	Median	17	37.5	475*	537*
	Minimum	10	27	459	499
	Maximum	25	44	494	561
Grade	Median	18	32	476*	512*
	Minimum	10	19	459	477
	Maximum	25	44	494	561

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 7 Reading								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	17	33	475	515			
1	2	19	32	477	512			
1	3	17	34	475	516			
1	4	17	32	475	512			
1	5	17	30	475	509			
2	6	17	35	475	517			
2	7	17	35	475	517			
2	8	17	35	475	517			
2	9	17	35	475	517			
2	10	17	35	475	517			
Table 1	Median	17	32.5	475*	512*			
	Minimum	17	32	475	512			
	Maximum	19	34	477	516			
Table 2	Median	17	35	475*	517*			
	Minimum	17	35	475	517			
	Maximum	17	35	475	517	% L1 + L2	% L3	% L4 + L5
Grade	Median	17	34	475*	516*	0.42	0.33	0.25
	Minimum	17	30	475	509			
	Maximum	19	35	477	517			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 7 Reading								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	17	32	475	512			
1	2	19	32	477	512			
1	3	17	34	475	516			
1	4	17	35	475	517			
1	5	17	33	475	515			
2	6	17	35	475	517			
2	7	17	34	475	516			
2	8	17	35	475	517			
2	9	17	36	475	526			
2	10	17	33	475	515			
Table 1	Median	17	33	475*	515*			
	Minimum	17	32	475	512			
	Maximum	19	35	477	517			
Table 2	Median	17	35	475*	517*			
	Minimum	17	34	475	516	Associated Impact Data*		
	Maximum	17	36	475	526	% L1 + L2	% L3	% L4 + L5
Grade	Median	17	34	475*	516*	0.42	0.33	0.25
	Minimum	17	32	475	512			
	Maximum	19	36	477	526			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 7 Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	31	50	466	568
1	2	31	48	466	548
1	3	31	45	466	539
1	4	29	48	454	548
1	5	29	47	454	548
2	6	32	51	473	585
2	7	42	51	524	585
2	8	41	51	517	585
2	9	36	53	492	607
2	10	29	53	454	607
Table 1	Median	31	48	466*	548*
	Minimum	29	45	454	539
	Maximum	31	50	466	568
Table 2	Median	38.5	51	504*	585*
	Minimum	32	51	473	585
	Maximum	42	53	524	607
Grade	Median	31	50	466*	568*
	Minimum	29	45	454	539
	Maximum	42	53	524	607

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 7 Speaking								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	30	48	464	548			
1	2	31	48	466	548			
1	3	30	46	464	545			
1	4	29	48	454	548			
1	5	29	47	454	548			
2	6	36	51	492	585			
2	7	42	51	524	585			
2	8	41	51	517	585			
2	9	40	53	511	607			
2	10	42	52	524	587			
Table 1	Median	30	48	464*	548*			
	Minimum	29	46	454	545			
	Maximum	31	48	466	548			
Table 2	Median	40.5	51	511*	585*			
	Minimum	36	51	492	585	Associated Impact Data*		
	Maximum	42	53	524	607	% L1 + L2	% L3	% L4 + L5
Grade	Median	31	48	466*	548*	0.27	0.64	0.09
	Minimum	29	46	454	545			
	Maximum	42	53	524	607			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 7 Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	30	48	464	548		
1	2	31	48	466	548		
1	3	34	46	486	545		
1	4	40	48	511	548		
1	5	36	48	492	548		
2	6	36	51	492	585		
2	7	42	51	524	585		
2	8	41	51	517	585		
2	9	42	53	524	607		
2	10	40	51	511	585		
Table 1	Median	32.5	48	473*	548*		
	Minimum	30	46	464	545		
	Maximum	40	48	511	548		
Table 2	Median	41.5	51	517*	585*		
	Minimum	36	51	492	585		
	Maximum	42	53	524	607		
Grade	Median	36	48	492*	548*	% L1 + L2	% L3
	Minimum	30	46	464	545	0.46	0.46
	Maximum	42	53	524	607		0.09

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for Grade 7 Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	26	38	431	489
1	2	26	43	431	514
1	3	32	43	462	514
1	4	28	52	438	589
1	5	27	42	436	509
2	6	32	48	462	542
2	7	33	49	464	549
2	8	39	50	490	557
2	9	38	52	489	589
2	10	24	47	422	528
Table 1	Median	27	43	436*	514*
	Minimum	26	38	431	489
	Maximum	32	52	462	589
Table 2	Median	35.5	49.5	469*	549*
	Minimum	32	48	462	542
	Maximum	39	52	490	589
Grade	Median	32	48	462*	542*
	Minimum	26	38	431	489
	Maximum	39	52	490	589

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for Grade 7 Writing							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	28	52	438	589		
1	2	31	48	455	542		
1	3	28	48	438	542		
1	4	28	52	438	589		
1	5	30	49	446	549		
2	6	33	49	464	549		
2	7	39	52	490	589		
2	8	38	52	489	589		
2	9	39	52	490	589		
2	10	39	52	490	589		
Table 1	Median	28	50	438*	557*		
	Minimum	28	48	438	542		
	Maximum	31	52	455	589		
Table 2	Median	38.5	52	489*	589*		
	Minimum	33	49	464	549		
	Maximum	39	52	490	589		
Grade	Median	31	52	455*	589*	% L1 + L2	% L3
	Minimum	28	48	438	542	0.22	0.77
	Maximum	39	52	490	589		0.01

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for Grade 7 Writing								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	33	48	464	542			
1	2	32	48	462	542			
1	3	35	48	469	542			
1	4	36	52	470	589			
1	5	32	44	462	516			
2	6	38	49	489	549			
2	7	39	52	490	589			
2	8	38	52	489	589			
2	9	39	52	490	589			
2	10	39	52	490	589			
Table 1	Median	34	48	469*	542*			
	Minimum	32	48	462	542			
	Maximum	36	52	470	589			
Table 2	Median	38.5	52	489*	589*			
	Minimum	38	49	489	549			
	Maximum	39	52	490	589			
Grade	Median	36	49	470*	549*	% L1 + L2	% L3	% L4 + L5
	Minimum	32	44	462	516	0.32	0.62	0.06
	Maximum	39	52	490	589			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for High School Listening</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	11	31	432	492
1	2	16	35	443	503
1	3	15	31	439	492
1	4	14	31	439	492
1	5	15	36	439	507
2	6	27	47	481	527
2	7	16	37	443	507
2	8	22	36	453	507
2	9	31	43	492	519
Table 1	Median	14.5	31	439*	492*
	Minimum	11	31	432	492
	Maximum	16	35	443	503
Table 2	Median	24.5	40	461*	514*
	Minimum	16	36	443	507
	Maximum	31	47	492	527
Grade	Median	16	36	443*	507*
	Minimum	11	31	432	492
	Maximum	31	47	492	527

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for High School Listening								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	15	31	439	492			
1	2	16	32	443	499			
1	3	16	31	443	492			
1	4	15	32	439	499			
1	5	16	31	443	492			
2	6	25	40	468	514			
2	7	25	42	468	517			
2	8	25	40	468	514			
2	9	25	41	468	516			
Table 1	Median	15.5	31.5	439*	492*			
	Minimum	15	31	439	492			
	Maximum	16	32	443	499			
Table 2	Median	25	40.5	468*	514*			
	Minimum	25	40	468	514			
	Maximum	25	42	468	517	% L1 + L2	% L3	% L4 + L5
Grade	Median	16	32	443*	499*	0.15	0.40	0.44
	Minimum	15	31	439	492			
	Maximum	25	42	468	517			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for High School Listening							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	19	32	448	499		
1	2	18	35	444	503		
1	3	16	31	443	492		
1	4	19	36	448	507		
1	5	19	31	448	492		
2	6	24	36	461	507		
2	7	18	36	444	507		
2	8	18	36	444	507		
2	9	21	36	450	507		
Table 1	Median	18.5	33.5	444*	499*		
	Minimum	16	31	443	492		
	Maximum	19	36	448	507		
Table 2	Median	19.5	36	448*	507*		
	Minimum	18	36	444	507		
	Maximum	24	36	461	507		
Grade	Median	19	36	448*	507*		
	Minimum	16	31	443	492		
	Maximum	24	36	461	507		

Associated Impact Data*		
% L1 + L2	% L3	% L4 + L5
0.18	0.47	0.35

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for High School Reading</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	15	22	481	508
1	2	15	30	481	535
1	3	18	35	490	541
1	4	13	27	475	526
1	5	18	28	490	530
2	6	20	41	503	577
2	7	18	34	490	539
2	8	19	33	494	538
2	9	18	28	490	530
Table 1	Median	15	28.5	481*	530*
	Minimum	13	22	475	508
	Maximum	18	35	490	541
Table 2	Median	18.5	33.5	490*	538*
	Minimum	18	28	490	530
	Maximum	20	41	503	577
Grade	Median	18	30	490*	535*
	Minimum	13	22	475	508
	Maximum	20	41	503	577

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for High School Reading							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	18	28	490	530		
1	2	18	28	490	530		
1	3	15	29	481	532		
1	4	15	26	481	524		
1	5	16	35	484	541		
2	6	18	34	490	539		
2	7	18	34	490	539		
2	8	18	32	490	537		
2	9	18	33	490	538		
Table 1	Median	16.5	28	484*	530*		
	Minimum	15	26	481	524		
	Maximum	18	29	490	532		
Table 2	Median	18	33.5	490*	538*		
	Minimum	18	32	490	537		
	Maximum	18	34	490	539		
Grade	Median	18	32	490*	537*		
	Minimum	15	26	481	524		
	Maximum	18	35	490	541		
						Associated Impact Data*	
						% L1 + L2	% L3
						% L4 + L5	
						0.56	0.29
							0.15

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for High School Reading								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	15	30	481	535			
1	2	16	31	484	535			
1	3	16	30	484	535			
1	4	16	30	484	535			
1	5	16	32	484	537			
2	6	16	33	484	538			
2	7	16	31	484	535			
2	8	16	33	484	538			
2	9	16	31	484	535			
Table 1	Median	16	30	484*	535*			
	Minimum	15	30	481	535			
	Maximum	16	31	484	535			
Table 2	Median	16	32	484*	537*			
	Minimum	16	31	484	535			
	Maximum	16	33	484	538	% L1 + L2	% L3	% L4 + L5
Grade	Median	16	31	484*	535*	0.49	0.35	0.15
	Minimum	15	30	481	535			
	Maximum	16	33	484	538			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for High School Speaking</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	28	38	469	509
1	2	37	46	506	533
1	3	29	43	472	524
1	4	28	39	469	515
1	5	29	43	472	524
2	6	32	49	499	564
2	7	30	48	482	556
2	8	34	43	502	524
2	9	32	48	499	556
Table 1	Median	28.5	41	469*	519*
	Minimum	28	38	469	509
	Maximum	37	46	506	533
Table 2	Median	32	48	499*	556*
	Minimum	30	43	482	524
	Maximum	34	49	502	564
Grade	Median	30	43	482*	524*
	Minimum	28	38	469	509
	Maximum	37	49	506	564

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for High School Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	28	43	469	524		
1	2	28	38	469	509		
1	3	29	43	472	524		
1	4	28	43	469	524		
1	5	28	43	469	524		
2	6	32	46	499	533		
2	7	32	46	499	533		
2	8	32	43	499	524		
2	9	32	46	499	533		
Table 1	Median	28	43	469*	524*		
	Minimum	28	38	469	509		
	Maximum	29	43	472	524		
Table 2	Median	32	46	499*	533*		
	Minimum	32	43	499	524		
	Maximum	32	46	499	533		
Grade	Median	29	43	472*	524*	0.24	0.46
	Minimum	28	38	469	509		
	Maximum	32	46	499	533		

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for High School Speaking							
		Bookmarks		Associated Cut Scores			
Table	Panelist	Level 3	Level 4	Level 3	Level 4		
1	1	29	41	472	519		
1	2	32	41	499	519		
1	3	28	41	469	519		
1	4	29	43	472	524		
1	5	29	43	472	524		
2	6	32	43	499	524		
2	7	30	43	482	524		
2	8	30	43	482	524		
2	9	30	45	482	529		
Table 1	Median	29	41	472*	519*		
	Minimum	28	41	469	519		
	Maximum	32	43	499	524		
Table 2	Median	30	43	482*	524*		
	Minimum	30	43	482	524		
	Maximum	32	45	499	529		
Grade	Median	30	43	482*	524*		
	Minimum	28	41	469	519		
	Maximum	32	45	499	529		
						Associated Impact Data*	
						% L1 + L2	% L3
						% L4 + L5	
						0.25	0.45
						0.30	

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

<i>Round 1 Results for High School Writing</i>					
		Bookmarks		Associated Cut Scores	
Table	Panelist	Level 3	Level 4	Level 3	Level 4
1	1	23	34	460	512
1	2	24	36	465	523
1	3	28	35	470	517
1	4	19	34	433	512
1	5	13	37	417	531
2	6	29	44	472	566
2	7	25	40	467	545
2	8	22	36	454	523
2	9	30	38	483	534
Table 1	Median	23.5	34.5	460*	512*
	Minimum	13	34	417	512
	Maximum	28	36	470	523
Table 2	Median	27	39	468*	536*
	Minimum	22	36	454	523
	Maximum	30	44	483	566
Grade	Median	24	36	465*	523*
	Minimum	13	34	417	512
	Maximum	30	44	483	566

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 2 Results for High School Writing								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	27	35	468	517			
1	2	27	35	468	517			
1	3	28	35	470	517			
1	4	23	35	460	517			
1	5	28	35	470	517			
2	6	29	42	472	563			
2	7	29	42	472	563			
2	8	28	42	470	563			
2	9	29	40	472	545			
Table 1	Median	27	35	468*	517*			
	Minimum	23	35	460	517			
	Maximum	28	35	470	517			
Table 2	Median	29	42	472*	563*			
	Minimum	28	40	470	545			
	Maximum	29	42	472	563			
Grade	Median	28	35	470*	517*	0.30	0.39	0.31
	Minimum	23	35	460	517			
	Maximum	29	42	472	563			

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Round 3 Results for High School Writing								
		Bookmarks		Associated Cut Scores				
Table	Panelist	Level 3	Level 4	Level 3	Level 4			
1	1	27	38	468	534			
1	2	28	38	470	534			
1	3	28	40	470	545			
1	4	27	38	468	534			
1	5	28	38	470	534			
2	6	28	38	470	534			
2	7	28	40	470	545			
2	8	28	41	470	546			
2	9	28	39	470	536			
Table 1	Median	27.5	38	468*	534*			
	Minimum	27	38	468	534			
	Maximum	28	40	470	545			
Table 2	Median	28	39.5	470*	536*			
	Minimum	28	38	470	534			
	Maximum	28	41	470	546			
Grade	Median	28	38	470*	534*			
	Minimum	27	38	468	534			
	Maximum	28	41	470	546			
						Associated Impact Data*		
						% L1 + L2	% L3	% L4 + L5
						0.30	0.54	0.16

*the scaled score associated with the median bookmark, which may differ from the median of the scaled scores

Appendix F2. Calculating a Meaningful Standard Error for the Bookmark Cut Score

In the Bookmark Standard Setting Procedure for a given grade and content area, participants are assigned to roughly equivalent small groups that work independently through Round 2. Thus, the set of Round 2 cut scores provide some information about the stability of consensus in Bookmark cut scores across independent small group replications. To quantify this degree of consensus, we calculate the cluster sample standard error (Cochran, 1963, p. 210) of the Round 2 mean cut score. Cluster sample standard errors are appropriate when, as may be reasonably assumed here, data are collected from groups and independence can be assumed between groups but not within groups.

For the Bookmark Procedure, the standard error of the Bookmark cut score (SE_{cut}) is based on the cluster sample standard error of the Round 2 mean cut score. Because the final Bookmark cut scores are based on the *median* of the group instead of the mean, this cluster sample standard error (SE_{cut}) is adjusted by $\sqrt{\frac{\pi}{2}}$ (Huynh, 2003). The standard error of the Bookmark cut score is:

$$SE_{cut} = \left(\sqrt{\frac{\pi}{2}} \right) \left(\sqrt{\frac{S^2}{N} \left[1 + \left(\frac{N}{n} - 1 \right) r \right]} \right),$$

where S^2 is the sample variance of individual Round 2 cut scores, r is the Round 2 intraclass correlation, N is the number of participants, and n is the number of groups. To be precise, if Y_{ik} is the cut score from the i^{th} participant in the k^{th} group, \bar{Y}_k is the average cut score for group k , and $\bar{\bar{Y}}$ is the average of all Round 2 cut scores, then

$$r = \frac{Var(\bar{Y}_k)}{Var(\bar{Y}_k) + Var(Y_{ik} - \bar{Y}_k)} \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum_{n,k} (Y_{nk} - \bar{\bar{Y}})^2$$

If we have only two groups ($n=2$) and perfect dependence (agreement) within groups ($r=1$), then the cluster sample standard error simplifies to $SE_{cut} = \left(\sqrt{\frac{\pi}{2}} \right) \left(\frac{|Y_1 - Y_2|}{2} \right)$, which is the standard error formula employed by NAEP

for two independent replications of a modified Angoff procedure (ACT, 1983, pp. 4-8). If, on the other hand, individual participants acted independently of their groups ($r=0$), then the cluster sample standard error simplifies to the traditional standard error of the mean for independent observations, $SE_{cut} = \left(\sqrt{\frac{\pi}{2}} \right) \left(\sqrt{\frac{S^2}{N}} \right)$. In this

manner, SE_{cut} provides a simple, flexible, and general way to quantify the amount of uncertainty associated with final Bookmark cut scores.

It is appropriate (if statistically imprecise) to say that repeated replications of this very standard setting procedure with different judges sampled from the same population of potential judges would result in a range of cut scores, most of which would fall in a band of width $4 * SE_{cut}$. In the graphical displays of participant data, we depict such an interval centered at the median of the Round 3 cut score. The purpose of calculating statistics like SE_{cut} and producing graphs of the types displayed here is to effectively communicate the complex information that is gathered during a Bookmark Standard Setting Procedure.

Adapted from: Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998, April). The bookmark standard setting procedure: Methodology and recent implementations. Paper presented at the annual meeting of the National Council for Measurement in Education, San Diego, CA.

© THE REGENTS OF THE UNIVERSITY OF CALIFORNIA. For permission to use this work, contact ELPA21 at UCLA.

References

ACT (1993). Setting achievement levels on the 1992 National Assessment of Educational Progress in Mathematics, Reading, and Writing: A technical report on reliability and validity.

Cochran, W. G. (1963). *Sampling techniques*. New York: John Wiley & Sons.

Huynh, H. (2003, August). Technical Memorandum for Computing Standard Error in Bookmark Standard Setting. (The South Carolina PACT 2003 Standard Setting Support Project). Columbia: University of South Carolina.

Appendix G. Evaluations

Appendix G1. Post-Opening Session Readiness

Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
1	The orientation session provided a clear overview of the standard setting process	0	0	30	21	51	0%	0%	59%	41%
2	I understand the goals of the standard setting workshop	0	0	26	25	51	0%	0%	51%	49%
3	I understand my role in the standard setting workshop	0	0	24	27	51	0%	0%	47%	53%
4	The orientation session provided a clear overview of the Contrasting Groups Study	0	1	30	20	51	0%	2%	59%	39%
5	I understand how the Contrasting Groups Study data will be used in the standard setting process	1	2	27	21	51	2%	4%	53%	41%
6	The orientation session provided a clear explanation of the development of ELPA21	0	5	28	18	51	0%	10%	55%	35%
7	I understand how the results of the standard setting will be used to support the reporting of ELPA21 results	0	1	28	22	51	0%	2%	55%	43%
8	I understand how to study the items in the ordered item booklet	0	0	27	24	51	0%	0%	53%	47%
Sum		1	9	220	178	408				
percentage of total number of responses		0.2%	2.2%	53.9%	43.6%					

	If you answered Disagree or Strongly Disagree to any of questions 1-8, then please answer the next two Yes/No questions.	No	Yes	0
9	I would like additional training on studying the ordered item booklet.	28	3	31
10	I have additional questions on material presented during the opening session that I would like answered before I begin the next task.	29	2	31
	Sum	57	5	62
	percentage of total number of responses	92%	8%	

Appendix G2. Post-Bookmark Training Readiness

Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
		1	2	3	4		#1	#2	#3	#4
1	I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors	0	1	17	33	51	0%	2%	33%	65%
2	I participated in bookmark training and had an opportunity to ask questions and discuss the meaning of the Level 3 and Level 4 bookmarks	0	0	19	32	51	0%	0%	37%	63%
3	I understand how to place my bookmarks	0	0	19	32	51	0%	0%	37%	63%
4	I understand I will have opportunities to change my bookmarks in Rounds 2 and 3	0	0	17	34	51	0%	0%	33%	67%
Sum		0	1	72	131	204				
percentage of total number of responses		0.0%	0.5%	35.3%	64.2%	1				
If you answered Disagree or Strongly Disagree to any of questions 1-8, then please answer the next two Yes/No questions.		No	Yes			0				
5	I would like additional training on placing my bookmarks for Round 1.	19	0			19				
6	I have additional questions that I would like to ask before placing my Round 1 bookmarks.	19	0			19				
Sum		38	0			38				
percentage of total number of responses		100%	0%							

Appendix G3. Reading Evaluation Results

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
K	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	1	7	8	0%	0%	13%	88%
K	2	The training materials were helpful	0	0	4	4	8	0%	0%	50%	50%
K	3	Taking the student test was helpful and informative	0	2	5	1	8	0%	25%	63%	13%
K	4	My group shared a common understanding of the Target Students	0	0	4	4	8	0%	0%	50%	50%
K	5	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%
K	6	During Round 1, I placed my bookmarks independently	0	0	0	8	8	0%	0%	0%	100%
K	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	1	7	8	0%	0%	13%	88%
K	8	The policy definitions were clearly communicated	0	0	6	2	8	0%	0%	75%	25%
K	9	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
K	10	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	11	I feel the recommended standards that resulted from this process are reasonable	0	0	2	6	8	0%	0%	25%	75%
K	12	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
K	13	I understood how to interpret the impact data	0	0	2	6	8	0%	0%	25%	75%
K	14	The impact data influenced where I placed my final bookmarks	0	1	4	3	8	0%	13%	50%	38%
K	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
K	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	2	6	8	0%	0%	25%	75%
K	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
K	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%
K		Sum	0	3	52	89	144				
K		percentage of total number of responses	0%	2%	36%	62%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
1	2	The training materials were helpful	0	0	4	4	8	0%	0%	50%	50%
1	3	Taking the student test was helpful and informative	0	0	5	3	8	0%	0%	63%	38%
1	4	My group shared a common understanding of the Target Students	0	0	4	4	8	0%	0%	50%	50%
1	5	Discussing the Target Students helped me place my bookmarks	0	1	2	5	8	0%	13%	25%	63%
1	6	During Round 1, I placed my bookmarks independently	0	0	3	5	8	0%	0%	38%	63%
1	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	8	The policy definitions were clearly communicated	0	0	4	4	8	0%	0%	50%	50%
1	9	I understood how to place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
1	10	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
1	11	I feel the recommended standards that resulted from this process are reasonable	0	0	1	7	8	0%	0%	13%	88%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
1	12	The impact data helped me evaluate my group's final bookmarks	0	0	2	6	8	0%	0%	25%	75%
1	13	I understood how to interpret the impact data	0	0	4	4	8	0%	0%	50%	50%
1	14	The impact data influenced where I placed my final bookmarks	0	1	4	3	8	0%	13%	50%	38%
1	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
1	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
1	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
1	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
1	Sum		0	2	59	83	144				
1	percentage of total number of responses		0%	1%	41%	58%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
3	2	The training materials were helpful	0	0	2	6	8	0%	0%	25%	75%
3	3	Taking the student test was helpful and informative	0	1	2	5	8	0%	13%	25%	63%
3	4	My group shared a common understanding of the Target Students	0	0	4	4	8	0%	0%	50%	50%
3	5	Discussing the Target Students helped me place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	6	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
3	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	8	The policy definitions were clearly communicated	0	0	2	6	8	0%	0%	25%	75%
3	9	I understood how to place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
3	10	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
3	11	I feel the recommended standards that resulted from this process are reasonable	0	0	2	6	8	0%	0%	25%	75%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
3	12	The impact data helped me evaluate my group's final bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	13	I understood how to interpret the impact data	0	0	4	4	8	0%	0%	50%	50%
3	14	The impact data influenced where I placed my final bookmarks	0	1	2	5	8	0%	13%	25%	63%
3	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	1	0	4	3	8	13%	0%	50%	38%
3	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	1	4	3	8	0%	13%	50%	38%
3	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	1	4	3	8	0%	13%	50%	38%
3	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	4	3	8	0%	13%	50%	38%
3	Sum		1	5	48	90	144				
3	percentage of total number of responses		1%	3%	33%	63%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	4	4	8	0%	0%	50%	50%
5	2	The training materials were helpful	0	0	6	2	8	0%	0%	75%	25%
5	3	Taking the student test was helpful and informative	0	2	5	1	8	0%	25%	63%	13%
5	4	My group shared a common understanding of the Target Students	0	0	5	3	8	0%	0%	63%	38%
5	5	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%
5	6	During Round 1, I placed my bookmarks independently	0	0	1	7	8	0%	0%	13%	88%
5	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	1	7	8	0%	0%	13%	88%
5	8	The policy definitions were clearly communicated	0	0	4	4	8	0%	0%	50%	50%
5	9	I understood how to place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
5	10	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
5	11	I feel the recommended standards that resulted from this process are reasonable	0	0	4	4	8	0%	0%	50%	50%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
5	12	The impact data helped me evaluate my group's final bookmarks	0	1	4	3	8	0%	13%	50%	38%
5	13	I understood how to interpret the impact data	0	0	4	4	8	0%	0%	50%	50%
5	14	The impact data influenced where I placed my final bookmarks	0	2	5	1	8	0%	25%	63%	13%
5	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	5	3	8	0%	0%	63%	38%
5	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	5	3	8	0%	0%	63%	38%
5	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	5	3	8	0%	0%	63%	38%
5	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	5	3	8	0%	0%	63%	38%
5	Sum		0	5	71	68	144				
5	percentage of total number of responses		0%	3%	49%	47%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	1	4	5	10	0%	10%	40%	50%
7	2	The training materials were helpful	0	1	4	5	10	0%	10%	40%	50%
7	3	Taking the student test was helpful and informative	0	1	3	6	10	0%	10%	30%	60%
7	4	My group shared a common understanding of the Target Students	0	2	4	4	10	0%	20%	40%	40%
7	5	Discussing the Target Students helped me place my bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	6	During Round 1, I placed my bookmarks independently	0	0	2	8	10	0%	0%	20%	80%
7	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	8	10	0%	0%	20%	80%
7	8	The policy definitions were clearly communicated	0	0	5	5	10	0%	0%	50%	50%
7	9	I understood how to place my bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	10	I had enough time to consider my bookmark placement	0	0	2	8	10	0%	0%	20%	80%
7	11	I feel the recommended standards that resulted from this process are reasonable	0	0	5	5	10	0%	0%	50%	50%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
7	12	The impact data helped me evaluate my group's final bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	13	I understood how to interpret the impact data	0	0	6	4	10	0%	0%	60%	40%
7	14	The impact data influenced where I placed my final bookmarks	0	1	7	2	10	0%	10%	70%	20%
7	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	7	10	0%	0%	30%	70%
7	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	1	3	6	10	0%	10%	30%	60%
7	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	7	10	0%	0%	30%	70%
7	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	6	10	0%	0%	40%	60%
7	Sum		0	7	66	107	180				
7	percentage of total number of responses		0%	4%	37%	59%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	7	9	0%	0%	22%	78%
HS	2	The training materials were helpful	0	0	1	8	9	0%	0%	11%	89%
HS	3	Taking the student test was helpful and informative	0	0	0	9	9	0%	0%	0%	100%
HS	4	My group shared a common understanding of the Target Students	0	0	2	7	9	0%	0%	22%	78%
HS	5	Discussing the Target Students helped me place my bookmarks	0	0	0	9	9	0%	0%	0%	100%
HS	6	During Round 1, I placed my bookmarks independently	0	0	0	9	9	0%	0%	0%	100%
HS	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	7	9	0%	0%	22%	78%
HS	8	The policy definitions were clearly communicated	0	0	3	6	9	0%	0%	33%	67%
HS	9	I understood how to place my bookmarks	0	0	0	9	9	0%	0%	0%	100%
HS	10	I had enough time to consider my bookmark placement	0	0	0	9	9	0%	0%	0%	100%
HS	11	I feel the recommended standards that resulted from this process are reasonable	0	0	2	7	9	0%	0%	22%	78%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
HS	12	The impact data helped me evaluate my group's final bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	13	I understood how to interpret the impact data	0	0	0	9	9	0%	0%	0%	100%
HS	14	The impact data influenced where I placed my final bookmarks	0	0	7	2	9	0%	0%	78%	22%
HS	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	6	9	0%	0%	33%	67%
HS	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	3	6	9	0%	0%	33%	67%
HS	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	2	7	9	0%	0%	22%	78%
HS	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	2	7	9	0%	0%	22%	78%
HS		Sum	0	0	30	132	162				
HS		percentage of total number of responses	0%	0%	19%	81%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	1	17	33	51	0%	2%	33%	65%
summary	2	The training materials were helpful	0	1	21	29	51	0%	2%	41%	57%
summary	3	Taking the student test was helpful and informative	0	6	20	25	51	0%	12%	39%	49%
summary	4	My group shared a common understanding of the Target Students	0	2	23	26	51	0%	4%	45%	51%
summary	5	Discussing the Target Students helped me place my bookmarks	0	1	15	35	51	0%	2%	29%	69%
summary	6	During Round 1, I placed my bookmarks independently	0	0	8	43	51	0%	0%	16%	84%
summary	7	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	11	40	51	0%	0%	22%	78%
summary	8	The policy definitions were clearly communicated	0	0	24	27	51	0%	0%	47%	53%
summary	9	I understood how to place my bookmarks	0	0	11	40	51	0%	0%	22%	78%
summary	10	I had enough time to consider my bookmark placement	0	0	10	41	51	0%	0%	20%	80%
summary	11	I feel the recommended standards that resulted from this process are reasonable	0	0	16	35	51	0%	0%	31%	69%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
summary	12	The impact data helped me evaluate my group's final bookmarks	0	1	15	35	51	0%	2%	29%	69%
summary	13	I understood how to interpret the impact data	0	0	20	31	51	0%	0%	39%	61%
summary	14	The impact data influenced where I placed my final bookmarks	0	6	29	16	51	0%	12%	57%	31%
summary	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	1	0	22	28	51	2%	0%	43%	55%
summary	16	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	2	21	28	51	0%	4%	41%	55%
summary	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	1	21	29	51	0%	2%	41%	57%
summary	18	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	22	28	51	0%	2%	43%	55%
summary	Sum		1	22	326	569	918				
summary	percentage of total number of responses		0%	2%	36%	62%					

Appendix G4. Listening Evaluation Results

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
K	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	1	7	8	0%	0%	13%	88%
K	2	Taking the student test was helpful and informative	0	1	5	2	8	0%	13%	63%	25%
K	3	My group shared a common understanding of the Target Students	0	0	3	5	8	0%	0%	38%	63%
K	4	Discussing the Target Students helped me place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
K	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
K	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	4	4	8	0%	0%	50%	50%
K	7	The policy definitions were clearly communicated	0	0	3	5	8	0%	0%	38%	63%
K	8	I understood how to place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
K	9	I had enough time to consider my bookmark placement	0	0	1	7	8	0%	0%	13%	88%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	10	I feel the recommended standards that resulted from this process are reasonable	0	0	5	3	8	0%	0%	63%	38%
K	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
K	12	I understood how to interpret the impact data	0	0	1	7	8	0%	0%	13%	88%
K	13	The impact data influenced where I placed my final bookmarks	0	0	3	5	8	0%	0%	38%	63%
K	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	5	3	8	0%	0%	63%	38%
K	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	5	3	8	0%	0%	63%	38%
K	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	5	3	8	0%	0%	63%	38%
K	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	5	3	8	0%	0%	63%	38%
K		Sum	0	1	54	81	136	0%	1%	40%	60%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
		percentage of total number of responses	0%	1%	40%	60%					
1	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	4	4	8	0%	0%	50%	50%
1	2	Taking the student test was helpful and informative	0	0	3	5	8	0%	0%	38%	63%
1	3	My group shared a common understanding of the Target Students	0	0	3	5	8	0%	0%	38%	63%
1	4	Discussing the Target Students helped me place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	5	During Round 1, I placed my bookmarks independently	0	0	1	7	8	0%	0%	13%	88%
1	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
1	7	The policy definitions were clearly communicated	0	0	2	6	8	0%	0%	25%	75%
1	8	I understood how to place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
1	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	10	I feel the recommended standards that resulted from this process are reasonable	0	0	6	2	8	0%	0%	75%	25%
1	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	12	I understood how to interpret the impact data	0	0	5	3	8	0%	0%	63%	38%
1	13	The impact data influenced where I placed my final bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
1	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
1	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	5	3	8	0%	0%	63%	38%
1	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
1	Sum		0	0	54	82	136	0%	0%	40%	60%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
		percentage of total number of responses	0%	0%	40%	60%					
3	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
3	2	Taking the student test was helpful and informative	0	0	4	4	8	0%	0%	50%	50%
3	3	My group shared a common understanding of the Target Students	0	1	3	4	8	0%	13%	38%	50%
3	4	Discussing the Target Students helped me place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
3	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	7	The policy definitions were clearly communicated	0	0	3	5	8	0%	0%	38%	63%
3	8	I understood how to place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	10	I feel the recommended standards that resulted from this process are reasonable	0	0	3	5	8	0%	0%	38%	63%
3	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%
3	13	The impact data influenced where I placed my final bookmarks	0	1	4	3	8	0%	13%	50%	38%
3	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
3	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	1	2	5	8	0%	13%	25%	63%
3	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
3	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	2	5	8	0%	13%	25%	63%
3		Sum	0	4	48	84	136	0%	3%	35%	62%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
		percentage of total number of responses	0%	3%	35%	62%					
5	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
5	2	Taking the student test was helpful and informative	0	2	5	1	8	0%	25%	63%	13%
5	3	My group shared a common understanding of the Target Students	0	0	5	3	8	0%	0%	63%	38%
5	4	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%
5	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
5	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	1	7	8	0%	0%	13%	88%
5	7	The policy definitions were clearly communicated	0	0	5	3	8	0%	0%	63%	38%
5	8	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
5	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	10	I feel the recommended standards that resulted from this process are reasonable	0	0	3	5	8	0%	0%	38%	63%
5	11	The impact data helped me evaluate my group's final bookmarks	0	0	4	4	8	0%	0%	50%	50%
5	12	I understood how to interpret the impact data	0	0	2	6	8	0%	0%	25%	75%
5	13	The impact data influenced where I placed my final bookmarks	0	1	4	3	8	0%	13%	50%	38%
5	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
5	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
5	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
5	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
5	Sum		0	3	59	74	136	0%	2%	43%	54%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
		percentage of total number of responses	0%	2%	43%	54%					
7	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	4	6	10	0%	0%	40%	60%
7	2	Taking the student test was helpful and informative	0	1	4	5	10	0%	10%	40%	50%
7	3	My group shared a common understanding of the Target Students	0	0	4	6	10	0%	0%	40%	60%
7	4	Discussing the Target Students helped me place my bookmarks	0	0	4	6	10	0%	0%	40%	60%
7	5	During Round 1, I placed my bookmarks independently	0	0	3	7	10	0%	0%	30%	70%
7	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	7	The policy definitions were clearly communicated	0	0	4	6	10	0%	0%	40%	60%
7	8	I understood how to place my bookmarks	0	0	2	8	10	0%	0%	20%	80%
7	9	I had enough time to consider my bookmark placement	0	0	2	8	10	0%	0%	20%	80%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	6	10	0%	0%	40%	60%
7	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	12	I understood how to interpret the impact data	0	0	4	6	10	0%	0%	40%	60%
7	13	The impact data influenced where I placed my final bookmarks	0	1	6	3	10	0%	10%	60%	30%
7	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	6	10	0%	0%	40%	60%
7	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	4	6	10	0%	0%	40%	60%
7	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	6	10	0%	0%	40%	60%
7	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	6	10	0%	0%	40%	60%
7		Sum	0	2	63	105	170	0%	1%	37%	62%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
		percentage of total number of responses	0%	1%	37%	62%					
HS	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	0	9	9	0%	0%	0%	100%
HS	2	Taking the student test was helpful and informative	0	0	1	8	9	0%	0%	11%	89%
HS	3	My group shared a common understanding of the Target Students	0	0	0	9	9	0%	0%	0%	100%
HS	4	Discussing the Target Students helped me place my bookmarks	0	0	0	9	9	0%	0%	0%	100%
HS	5	During Round 1, I placed my bookmarks independently	0	0	0	9	9	0%	0%	0%	100%
HS	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	7	The policy definitions were clearly communicated	0	0	0	9	9	0%	0%	0%	100%
HS	8	I understood how to place my bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	9	I had enough time to consider my bookmark placement	0	0	2	7	9	0%	0%	22%	78%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	5	9	0%	0%	44%	56%
HS	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	6	9	0%	0%	33%	67%
HS	12	I understood how to interpret the impact data	0	0	3	6	9	0%	0%	33%	67%
HS	13	The impact data influenced where I placed my final bookmarks	0	1	7	1	9	0%	11%	78%	11%
HS	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	6	3	9	0%	0%	67%	33%
HS	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	6	3	9	0%	0%	67%	33%
HS	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	5	4	9	0%	0%	56%	44%
HS	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	5	4	9	0%	0%	56%	44%
HS	Sum		0	1	44	108	153	0%	1%	29%	71%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
		percentage of total number of responses	0%	1%	29%	71%					
summary	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	15	36	51	0%	0%	29%	71%
summary	2	Taking the student test was helpful and informative	0	4	22	25	51	0%	8%	43%	49%
summary	3	My group shared a common understanding of the Target Students	0	1	18	32	51	0%	2%	35%	63%
summary	4	Discussing the Target Students helped me place my bookmarks	0	0	15	36	51	0%	0%	29%	71%
summary	5	During Round 1, I placed my bookmarks independently	0	0	10	41	51	0%	0%	20%	80%
summary	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	13	38	51	0%	0%	25%	75%
summary	7	The policy definitions were clearly communicated	0	0	17	34	51	0%	0%	33%	67%
summary	8	I understood how to place my bookmarks	0	0	11	40	51	0%	0%	22%	78%
summary	9	I had enough time to consider my bookmark placement	0	0	11	40	51	0%	0%	22%	78%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	10	I feel the recommended standards that resulted from this process are reasonable	0	0	25	26	51	0%	0%	49%	51%
summary	11	The impact data helped me evaluate my group's final bookmarks	0	0	19	32	51	0%	0%	37%	63%
summary	12	I understood how to interpret the impact data	0	0	18	33	51	0%	0%	35%	65%
summary	13	The impact data influenced where I placed my final bookmarks	0	4	27	20	51	0%	8%	53%	39%
summary	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	26	25	51	0%	0%	51%	49%
summary	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	1	25	25	51	0%	2%	49%	49%
summary	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	26	25	51	0%	0%	51%	49%
summary	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	24	26	51	0%	2%	47%	51%
summary		Sum	0	11	322	534	867	0%	1%	37%	62%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
		percentage of total number of responses	0%	1%	37%	62%					

Appendix G5. Speaking Evaluation Results

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
K	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
K	2	Taking the student test was helpful and informative	0	1	6	1	8	0%	13%	75%	13%
K	3	My group shared a common understanding of the Target Students	0	2	5	1	8	0%	25%	63%	13%
K	4	Discussing the Target Students helped me place my bookmarks	0	2	5	1	8	0%	25%	63%	13%
K	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
K	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	4	4	8	0%	0%	50%	50%
K	7	The policy definitions were clearly communicated	0	1	5	2	8	0%	13%	63%	25%
K	8	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
K	9	I had enough time to consider my bookmark placement	0	2	1	5	8	0%	25%	13%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	4	8	0%	0%	50%	50%
K	11	The impact data helped me evaluate my group's final bookmarks	0	1	4	3	8	0%	13%	50%	38%
K	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%
K	13	The impact data influenced where I placed my final bookmarks	0	1	4	3	8	0%	13%	50%	38%
K	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
K	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0	4	4	8	0%	0%	50%	50%
K	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
K	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	0	4	4	8	0%	0%	50%	50%
K	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	5	3	8	0%	0%	63%	38%
K	20	I feel that my grade group as a whole is credible	0	0	1	7	8	0%	0%	13%	88%
K	21	Overall, I believe that my opinions were considered and valued by my group	0	0	2	6	8	0%	0%	25%	75%
K	22	Overall, I valued the workshop as a professional development experience	0	0	0	8	8	0%	0%	0%	100%
K	23	This experience will help me target instruction for the students in my classroom	0	1	3	4	8	0%	13%	38%	50%
K	24	The food and service at the facility met my expectations.	0	1	5	2	8	0%	13%	63%	25%
K	25	The workspace had accommodations appropriate to facilitate our work	0	4	2	2	8	0%	50%	25%	25%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
K	26	Participating in the workshop increased my understanding of ELPA 21	0	0	2	6	8	0%	0%	25%	75%
K	27	The workshop was well organized	0	0	4	4	8	0%	0%	50%	50%
K		Sum	0	16	92	108	216				
K		percentage of total number of responses	0%	7%	43%	50%					
1	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	1	7	8	0%	0%	13%	88%
1	2	Taking the student test was helpful and informative	0	0	1	7	8	0%	0%	13%	88%
1	3	My group shared a common understanding of the Target Students	0	1	1	6	8	0%	13%	13%	75%
1	4	Discussing the Target Students helped me place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
1	5	During Round 1, I placed my bookmarks independently	0	0	1	7	8	0%	0%	13%	88%
1	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	1	7	8	0%	0%	13%	88%
1	7	The policy definitions were clearly communicated	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	8	I understood how to place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
1	9	I had enough time to consider my bookmark placement	0	0	1	7	8	0%	0%	13%	88%
1	10	I feel the recommended standards that resulted from this process are reasonable	1	1	2	4	8	13%	13%	25%	50%
1	11	The impact data helped me evaluate my group's final bookmarks	1	2	2	3	8	13%	25%	25%	38%
1	12	I understood how to interpret the impact data	0	0	0	8	8	0%	0%	0%	100%
1	13	The impact data influenced where I placed my final bookmarks	2	3	1	2	8	25%	38%	13%	25%
1	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	1	1	6	8	0%	13%	13%	75%
1	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0	3	5	8	0%	0%	38%	63%
1	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%
1	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	1	3	4	8	0%	13%	38%	50%
1	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	3	5	8	0%	0%	38%	63%
1	20	I feel that my grade group as a whole is credible	1	2	3	2	8	13%	25%	38%	25%
1	21	Overall, I believe that my opinions were considered and valued by my group	0	1	1	6	8	0%	13%	13%	75%
1	22	Overall, I valued the workshop as a professional development experience	0	0	1	7	8	0%	0%	13%	88%
1	23	This experience will help me target instruction for the students in my classroom	0	0	1	7	8	0%	0%	13%	88%
1	24	The food and service at the facility met my expectations.	0	1	1	6	8	0%	13%	13%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	25	The workspace had accommodations appropriate to facilitate our work	0	0	0	8	8	0%	0%	0%	100%
1	26	Participating in the workshop increased my understanding of ELPA 21	0	0	0	8	8	0%	0%	0%	100%
1	27	The workshop was well organized	0	0	0	8	8	0%	0%	0%	100%
1		Sum	5	13	38	160	216				
1		percentage of total number of responses	2%	6%	18%	74%					
3	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	6	8	0%	0%	25%	75%
3	2	Taking the student test was helpful and informative	0	0	3	5	8	0%	0%	38%	63%
3	3	My group shared a common understanding of the Target Students	0	0	3	5	8	0%	0%	38%	63%
3	4	Discussing the Target Students helped me place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	7	The policy definitions were clearly communicated	0	0	3	5	8	0%	0%	38%	63%
3	8	I understood how to place my bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
3	10	I feel the recommended standards that resulted from this process are reasonable	0	0	3	5	8	0%	0%	38%	63%
3	11	The impact data helped me evaluate my group's final bookmarks	0	0	2	6	8	0%	0%	25%	75%
3	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%
3	13	The impact data influenced where I placed my final bookmarks	0	0	4	4	8	0%	0%	50%	50%
3	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
3	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	1	2	5	8	0%	13%	25%	63%
3	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
3	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	2	5	8	0%	13%	25%	63%
3	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	0	3	5	8	0%	0%	38%	63%
3	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	3	5	8	0%	0%	38%	63%
3	20	I feel that my grade group as a whole is credible	0	0	2	6	8	0%	0%	25%	75%
3	21	Overall, I believe that my opinions were considered and valued by my group	0	0	2	6	8	0%	0%	25%	75%
3	22	Overall, I valued the workshop as a professional development experience	0	0	1	7	8	0%	0%	13%	88%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	23	This experience will help me target instruction for the students in my classroom	0	0	1	7	8	0%	0%	13%	88%
3	24	The food and service at the facility met my expectations.	0	1	2	5	8	0%	13%	25%	63%
3	25	The workspace had accommodations appropriate to facilitate our work	0	0	2	6	8	0%	0%	25%	75%
3	26	Participating in the workshop increased my understanding of ELPA 21	0	0	0	8	8	0%	0%	0%	100%
3	27	The workshop was well organized	0	0	0	8	8	0%	0%	0%	100%
3		Sum	0	3	58	155	216				
3		percentage of total number of responses	0%	1%	27%	72%					
5	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
5	2	Taking the student test was helpful and informative	0	2	4	2	8	0%	25%	50%	25%
5	3	My group shared a common understanding of the Target Students	0	0	3	4	7	0%	0%	43%	57%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	4	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%
5	5	During Round 1, I placed my bookmarks independently	0	0	1	7	8	0%	0%	13%	88%
5	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	3	5	8	0%	0%	38%	63%
5	7	The policy definitions were clearly communicated	0	0	4	4	8	0%	0%	50%	50%
5	8	I understood how to place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
5	9	I had enough time to consider my bookmark placement	0	0	1	7	8	0%	0%	13%	88%
5	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	4	8	0%	0%	50%	50%
5	11	The impact data helped me evaluate my group's final bookmarks	0	0	2	6	8	0%	0%	25%	75%
5	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%
5	13	The impact data influenced where I placed my final bookmarks	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
5	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0	4	4	8	0%	0%	50%	50%
5	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
5	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	4	4	8	0%	0%	50%	50%
5	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	0	4	4	8	0%	0%	50%	50%
5	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	3	5	8	0%	0%	38%	63%
5	20	I feel that my grade group as a whole is credible	0	0	3	5	8	0%	0%	38%	63%
5	21	Overall, I believe that my opinions were considered and valued by my group	0	0	1	7	8	0%	0%	13%	88%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	22	Overall, I valued the workshop as a professional development experience	0	0	1	7	8	0%	0%	13%	88%
5	23	This experience will help me target instruction for the students in my classroom	0	0	3	5	8	0%	0%	38%	63%
5	24	The food and service at the facility met my expectations.	0	0	1	7	8	0%	0%	13%	88%
5	25	The workspace had accommodations appropriate to facilitate our work	0	0	1	7	8	0%	0%	13%	88%
5	26	Participating in the workshop increased my understanding of ELPA 21	0	0	1	7	8	0%	0%	13%	88%
5	27	The workshop was well organized	0	0	1	7	8	0%	0%	13%	88%
5		Sum	0	2	70	143	215				
5		percentage of total number of responses	0%	1%	33%	67%					
7	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	7	10	0%	0%	30%	70%
7	2	Taking the student test was helpful and informative	0	0	4	6	10	0%	0%	40%	60%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	3	My group shared a common understanding of the Target Students	0	1	4	5	10	0%	10%	40%	50%
7	4	Discussing the Target Students helped me place my bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	5	During Round 1, I placed my bookmarks independently	0	0	1	9	10	0%	0%	10%	90%
7	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	3	7	10	0%	0%	30%	70%
7	7	The policy definitions were clearly communicated	0	0	4	6	10	0%	0%	40%	60%
7	8	I understood how to place my bookmarks	0	0	1	9	10	0%	0%	10%	90%
7	9	I had enough time to consider my bookmark placement	0	0	1	9	10	0%	0%	10%	90%
7	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	6	10	0%	0%	40%	60%
7	11	The impact data helped me evaluate my group's final bookmarks	0	2	3	5	10	0%	20%	30%	50%
7	12	I understood how to interpret the impact data	0	1	3	6	10	0%	10%	30%	60%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	13	The impact data influenced where I placed my final bookmarks	1	2	4	3	10	10%	20%	40%	30%
7	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	7	10	0%	0%	30%	70%
7	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0	2	8	10	0%	0%	20%	80%
7	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	7	10	0%	0%	30%	70%
7	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	2	8	10	0%	0%	20%	80%
7	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	0	4	6	10	0%	0%	40%	60%
7	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	5	5	10	0%	0%	50%	50%
7	20	I feel that my grade group as a whole is credible	0	0	4	6	10	0%	0%	40%	60%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	21	Overall, I believe that my opinions were considered and valued by my group	0	0	3	7	10	0%	0%	30%	70%
7	22	Overall, I valued the workshop as a professional development experience	0	0	0	10	10	0%	0%	0%	100%
7	23	This experience will help me target instruction for the students in my classroom	0	0	1	9	10	0%	0%	10%	90%
7	24	The food and service at the facility met my expectations.	0	0	1	9	10	0%	0%	10%	90%
7	25	The workspace had accommodations appropriate to facilitate our work	0	0	1	9	10	0%	0%	10%	90%
7	26	Participating in the workshop increased my understanding of ELPA 21	0	0	0	10	10	0%	0%	0%	100%
7	27	The workshop was well organized	0	0	0	10	10				
7		Sum	1	6	67	196	270				
7		percentage of total number of responses	0%	2%	25%	73%					
HS	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	7	9	0%	0%	22%	78%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	2	Taking the student test was helpful and informative	0	0	4	5	9	0%	0%	44%	56%
HS	3	My group shared a common understanding of the Target Students	0	0	6	3	9	0%	0%	67%	33%
HS	4	Discussing the Target Students helped me place my bookmarks	0	1	4	4	9	0%	11%	44%	44%
HS	5	During Round 1, I placed my bookmarks independently	0	0	0	9	9	0%	0%	0%	100%
HS	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	7	The policy definitions were clearly communicated	0	0	0	9	9	0%	0%	0%	100%
HS	8	I understood how to place my bookmarks	0	0	0	9	9	0%	0%	0%	100%
HS	9	I had enough time to consider my bookmark placement	0	0	0	9	9	0%	0%	0%	100%
HS	10	I feel the recommended standards that resulted from this process are reasonable	0	0	5	4	9	0%	0%	56%	44%
HS	11	The impact data helped me evaluate my group's final bookmarks	0	0	4	5	9	0%	0%	44%	56%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	12	I understood how to interpret the impact data	0	0	2	7	9	0%	0%	22%	78%
HS	13	The impact data influenced where I placed my final bookmarks	0	1	6	2	9	0%	11%	67%	22%
HS	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	1	3	5	9	0%	11%	33%	56%
HS	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	0	3	6	9	0%	0%	33%	67%
HS	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	5	9	0%	0%	44%	56%
HS	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	6	9	0%	0%	33%	67%
HS	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	0	5	4	9	0%	0%	56%	44%
HS	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	3	6	9	0%	0%	33%	67%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	20	I feel that my grade group as a whole is credible	0	0	1	8	9	0%	0%	11%	89%
HS	21	Overall, I believe that my opinions were considered and valued by my group	0	0	1	8	9	0%	0%	11%	89%
HS	22	Overall, I valued the workshop as a professional development experience	0	0	0	9	9	0%	0%	0%	100%
HS	23	This experience will help me target instruction for the students in my classroom	0	0	1	8	9	0%	0%	11%	89%
HS	24	The food and service at the facility met my expectations.	0	0	4	5	9	0%	0%	44%	56%
HS	25	The workspace had accommodations appropriate to facilitate our work	0	0	0	9	9	0%	0%	0%	100%
HS	26	Participating in the workshop increased my understanding of ELPA 21	0	0	0	9	9	0%	0%	0%	100%
HS	27	The workshop was well organized	0	0	1	8	9	0%	0%	11%	89%
HS		Sum	0	3	63	177	243				
HS		percentage of total number of responses	0%	1%	26%	73%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	14	37	51	0%	0%	27%	73%
summary	2	Taking the student test was helpful and informative	0	3	22	26	51	0%	6%	43%	51%
summary	3	My group shared a common understanding of the Target Students	0	4	22	24	50	0%	8%	44%	48%
summary	4	Discussing the Target Students helped me place my bookmarks	0	3	19	29	51	0%	6%	37%	57%
summary	5	During Round 1, I placed my bookmarks independently	0	0	7	44	51	0%	0%	14%	86%
summary	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	14	37	51	0%	0%	27%	73%
summary	7	The policy definitions were clearly communicated	0	1	18	32	51	0%	2%	35%	63%
summary	8	I understood how to place my bookmarks	0	0	8	43	51	0%	0%	16%	84%
summary	9	I had enough time to consider my bookmark placement	0	2	6	43	51	0%	4%	12%	84%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	10	I feel the recommended standards that resulted from this process are reasonable	1	1	22	27	51	2%	2%	43%	53%
summary	11	The impact data helped me evaluate my group's final bookmarks	1	5	17	28	51	2%	10%	33%	55%
summary	12	I understood how to interpret the impact data	0	1	14	36	51	0%	2%	27%	71%
summary	13	The impact data influenced where I placed my final bookmarks	3	7	21	20	51	6%	14%	41%	39%
summary	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	2	16	33	51	0%	4%	31%	65%
summary	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.	0	1	18	32	51	0%	2%	35%	63%
summary	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	21	30	51	0%	0%	41%	59%
summary	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	18	32	51	0%	2%	35%	63%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
summary	18	I am confident that the Bookmark Standard Setting Procedure produced valid cut scores.	0	1	23	27	51	0%	2%	45%	53%
summary	19	I feel the procedure resulted in recommendations for performance standards that are defensible	0	0	22	29	51	0%	0%	43%	57%
summary	20	I feel that my grade group as a whole is credible	1	2	14	34	51	2%	4%	27%	67%
summary	21	Overall, I believe that my opinions were considered and valued by my group	0	1	10	40	51	0%	2%	20%	78%
summary	22	Overall, I valued the workshop as a professional development experience	0	0	3	48	51	0%	0%	6%	94%
summary	23	This experience will help me target instruction for the students in my classroom	0	1	10	40	51	0%	2%	20%	78%
summary	24	The food and service at the facility met my expectations.	0	3	14	34	51	0%	6%	27%	67%
summary	25	The workspace had accommodations appropriate to facilitate our work	0	4	6	41	51	0%	8%	12%	80%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
summary	26	Participating in the workshop increased my understanding of ELPA 21	0	0	3	48	51	0%	0%	6%	94%
summary	27	The workshop was well organized	0	0	6	45	51	0%	0%	12%	88%
summary		Sum	6	43	388	939	1376				
summary		percentage of total number of responses	0%	3%	28%	68%	1				

Appendix G6. Writing Evaluation Results

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
K	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	3	5	8	0%	0%	38%	63%
K	2	Taking the student test was helpful and informative	0	0	7	1	8	0%	0%	88%	13%
K	3	My group shared a common understanding of the Target Students	0	0	4	3	7	0%	0%	57%	43%
K	4	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%
K	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
K	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	5	3	8	0%	0%	63%	38%
K	7	The policy definitions were clearly communicated	0	1	4	3	8	0%	13%	50%	38%
K	8	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%

1 no response

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	9	I had enough time to consider my bookmark placement	0	0	3	5	8	0%	0%	38%	63%
K	10	I feel the recommended standards that resulted from this process are reasonable	0	0	4	4	8	0%	0%	50%	50%
K	11	The impact data helped me evaluate my group's final bookmarks	0	0	5	3	8	0%	0%	63%	38%
K	12	I understood how to interpret the impact data	0	0	4	4	8	0%	0%	50%	50%
K	13	The impact data influenced where I placed my final bookmarks	0	1	2	5	8	0%	13%	25%	63%
K	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	1	2	5	8	0%	13%	25%	63%
K	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	5	8	0%	0%	38%	63%
K	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%
K		Sum	0	3	61	71	135	0%	2%	45%	53%
		percentage of total number of responses	0%	2%	45%	53%					
1	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	6	8	0%	0%	25%	75%
1	2	Taking the student test was helpful and informative	0	0	4	4	8	0%	0%	50%	50%
1	3	My group shared a common understanding of the Target Students	0	1	4	3	8	0%	13%	50%	38%
1	4	Discussing the Target Students helped me place my bookmarks	0	0	3	5	8	0%	0%	38%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
1	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
1	7	The policy definitions were clearly communicated	0	0	3	5	8	0%	0%	38%	63%
1	8	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
1	10	I feel the recommended standards that resulted from this process are reasonable	0	0	3	5	8	0%	0%	38%	63%
1	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
1	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
1	13	The impact data influenced where I placed my final bookmarks	0	1	3	4	8	0%	13%	38%	50%
1	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	2	6	8	0%	0%	25%	75%
1	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%
1	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	2	6	8	0%	0%	25%	75%
1	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	5	8	0%	0%	38%	63%
1	Sum		0	2	47	87	136	0%	1%	35%	64%
percentage of total number of responses			0%	1%	35%	64%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	6	8	0%	0%	25%	75%
3	2	Taking the student test was helpful and informative	0	0	3	5	8	0%	0%	38%	63%
3	3	My group shared a common understanding of the Target Students	0	0	4	4	8	0%	0%	50%	50%
3	4	Discussing the Target Students helped me place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	5	During Round 1, I placed my bookmarks independently	0	0	2	6	8	0%	0%	25%	75%
3	6	I considered the ELP Proficiency Standards and ALLs when I placed my bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	7	The policy definitions were clearly communicated	0	0	3	5	8	0%	0%	38%	63%
3	8	I understood how to place my bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	10	I feel the recommended standards that resulted from this process are reasonable	0	0	3	5	8	0%	0%	38%	63%
3	11	The impact data helped me evaluate my group's final bookmarks	0	0	3	5	8	0%	0%	38%	63%
3	12	I understood how to interpret the impact data	0	0	3	5	8	0%	0%	38%	63%
3	13	The impact data influenced where I placed my final bookmarks	0	1	3	4	8	0%	13%	38%	50%
3	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
3	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	1	2	5	8	0%	13%	25%	63%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	4	4	8	0%	0%	50%	50%
3	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	3	4	8	0%	13%	38%	50%
3		Sum	0	3	50	83	136				
		percentage of total number of responses	0%	2%	37%	61%					
5	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	1	7	8	0%	0%	13%	88%
5	2	Taking the student test was helpful and informative	0	2	3	3	8	0%	25%	38%	38%
5	3	My group shared a common understanding of the Target Students	0	0	4	3	7	0%	0%	57%	43%
5	4	Discussing the Target Students helped me place my bookmarks	0	0	4	4	8	0%	0%	50%	50%

1 no response

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	5	During Round 1, I placed my bookmarks independently	0	0	1	7	8	0%	0%	13%	88%
5	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	6	8	0%	0%	25%	75%
5	7	The policy definitions were clearly communicated	0	0	4	4	8	0%	0%	50%	50%
5	8	I understood how to place my bookmarks	0	0	1	7	8	0%	0%	13%	88%
5	9	I had enough time to consider my bookmark placement	0	0	2	6	8	0%	0%	25%	75%
5	10	I feel the recommended standards that resulted from this process are reasonable	0	0	2	6	8	0%	0%	25%	75%
5	11	The impact data helped me evaluate my group's final bookmarks	0	1	2	5	8	0%	13%	25%	63%
5	12	I understood how to interpret the impact data	0	0	2	6	8	0%	0%	25%	75%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	13	The impact data influenced where I placed my final bookmarks	0	2	4	2	8	0%	25%	50%	25%
5	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	2	6	8	0%	0%	25%	75%
5	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	2	6	8	0%	0%	25%	75%
5	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	2	6	8	0%	0%	25%	75%
5	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	2	6	8	0%	0%	25%	75%
5	Sum		0	5	40	90	135	0%	4%	30%	67%
percentage of total number of responses			0%	4%	30%	67%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	2	8	10	0%	0%	20%	80%
7	2	Taking the student test was helpful and informative	0	1	4	5	10	0%	10%	40%	50%
7	3	My group shared a common understanding of the Target Students	0	0	3	7	10	0%	0%	30%	70%
7	4	Discussing the Target Students helped me place my bookmarks	0	0	2	8	10	0%	0%	20%	80%
7	5	During Round 1, I placed my bookmarks independently	0	0	1	9	10	0%	0%	10%	90%
7	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	1	9	10	0%	0%	10%	90%
7	7	The policy definitions were clearly communicated	0	0	2	8	10	0%	0%	20%	80%
7	8	I understood how to place my bookmarks	0	0	1	9	10	0%	0%	10%	90%
7	9	I had enough time to consider my bookmark placement	0	0	1	9	10	0%	0%	10%	90%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	10	I feel the recommended standards that resulted from this process are reasonable	0	0	2	8	10	0%	0%	20%	80%
7	11	The impact data helped me evaluate my group's final bookmarks	0	0	2	8	10	0%	0%	20%	80%
7	12	I understood how to interpret the impact data	0	0	2	8	10	0%	0%	20%	80%
7	13	The impact data influenced where I placed my final bookmarks	0	2	3	4	9	0%	22%	33%	44%
7	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	2	8	10	0%	0%	20%	80%
7	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	2	8	10	0%	0%	20%	80%

1 no response

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	2	8	10	0%	0%	20%	80%
7	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	2	8	10	0%	0%	20%	80%
7		Sum	0	3	34	132	169	0%	2%	20%	78%
		percentage of total number of responses	0%	2%	20%	78%					
HS	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	1	8	9	0%	0%	11%	89%
HS	2	Taking the student test was helpful and informative	0	0	4	5	9	0%	0%	44%	56%
HS	3	My group shared a common understanding of the Target Students	0	0	0	9	9	0%	0%	0%	100%
HS	4	Discussing the Target Students helped me place my bookmarks	0	0	0	9	9	0%	0%	0%	100%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	5	During Round 1, I placed my bookmarks independently	0	0	0	9	9	0%	0%	0%	100%
HS	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	2	7	9	0%	0%	22%	78%
HS	7	The policy definitions were clearly communicated	0	0	1	8	9	0%	0%	11%	89%
HS	8	I understood how to place my bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	9	I had enough time to consider my bookmark placement	0	0	0	9	9	0%	0%	0%	100%
HS	10	I feel the recommended standards that resulted from this process are reasonable	0	0	1	8	9	0%	0%	11%	89%
HS	11	The impact data helped me evaluate my group's final bookmarks	0	0	1	8	9	0%	0%	11%	89%
HS	12	I understood how to interpret the impact data	0	0	0	9	9	0%	0%	0%	100%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
HS	13	The impact data influenced where I placed my final bookmarks	0	1	3	5	9	0%	11%	33%	56%
HS	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	0	3	6	9	0%	0%	33%	67%
HS	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	0	3	6	9	0%	0%	33%	67%
HS	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	3	6	9	0%	0%	33%	67%
HS	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	0	3	6	9	0%	0%	33%	67%
HS	Sum		0	1	26	126	153	0%	1%	17%	82%
	percentage of total number of responses		0%	1%	17%	82%					

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking	0	0	11	40	51	0%	0%	22%	78%
summary	2	Taking the student test was helpful and informative	0	3	25	23	51	0%	6%	49%	45%
summary	3	My group shared a common understanding of the Target Students	0	1	22	27	50	0%	2%	44%	54%
summary	4	Discussing the Target Students helped me place my bookmarks	0	0	16	34	50	0%	0%	32%	68%
summary	5	During Round 1, I placed my bookmarks independently	0	0	10	41	51	0%	0%	20%	80%
summary	6	I considered the ELP Proficiency Standards and ALIs when I placed my bookmarks	0	0	12	39	51	0%	0%	24%	76%
summary	7	The policy definitions were clearly communicated	0	0	18	33	51	0%	0%	35%	65%
summary	8	I understood how to place my bookmarks	0	1	13	37	51	0%	2%	25%	73%
summary	9	I had enough time to consider my bookmark placement	0	0	10	41	51	0%	0%	20%	80%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
summary	10	I feel the recommended standards that resulted from this process are reasonable	0	0	14	37	51	0%	0%	27%	73%
summary	11	The impact data helped me evaluate my group's final bookmarks	0	1	15	35	51	0%	2%	29%	69%
summary	12	I understood how to interpret the impact data	0	0	15	36	51	0%	0%	29%	71%
summary	13	The impact data influenced where I placed my final bookmarks	0	7	20	23	50	0%	14%	40%	46%
summary	14	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high	0	1	15	35	51	0%	2%	29%	69%
summary	15	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low	0	2	14	35	51	0%	4%	27%	69%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
summary	16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high	0	0	16	35	51	0%	0%	31%	69%
summary	17	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low	0	1	16	34	51	0%	2%	31%	67%
summary		Sum	0	17	262	585	864	0%	2%	30%	68%
		percentage of total number of responses	0%	2%	30%	68%					

Appendix G7. Proficiency Determination

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
K	1	I understood the process used for the Proficiency Determination	0	0	3	4	7	0%	0%	43%	57%
K	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	4	3	7	0%	0%	57%	43%
K	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	0	0	3	4	7	0%	0%	43%	57%
K	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	2	5	7	0%	0%	29%	71%
K	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	0	2	5	7	0%	0%	29%	71%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
K	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	3	4	7	0%	0%	43%	57%
K		Sum	0	0	17	25	42				
K		percentage of total number of responses	0%	0%	40%	60%	1				
1	1	I understood the process used for the Proficiency Determination	0	0	1	7	8	0%	0%	13%	88%
1	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	1	7	8	0%	0%	13%	88%
1	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	0	2	1	5	8	0%	25%	13%	63%
1	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	0	8	8	0%	0%	0%	100%

Grade	Item No.	Question	strongly disagree 1	disagree 2	agree 3	strongly agree 4	Total	% #1	% #2	% #3	% #4
1	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	0	1	7	8	0%	0%	13%	88%
1	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	1	7	8	0%	0%	13%	88%
1		Sum	0	2	5	41	48				
		percentage of total number of responses	0%	4%	10%	85%	1				
3	1	I understood the process used for the Proficiency Determination	0	0	1	5	6	0%	0%	17%	83%
3	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	2	4	6	0%	0%	33%	67%
3	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	1	0	1	4	6	17%	0%	17%	67%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
3	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	1	5	6	0%	0%	17%	83%
3	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	1	1	4	6	0%	17%	17%	67%
3	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	1	5	6	0%	0%	17%	83%
3	Sum		1	1	7	27	36	3%	3%	19%	75%
3	percentage of total number of responses		3%	3%	19%	75%	1	3%	3%	19%	75%
5	1	I understood the process used for the Proficiency Determination	0	0	3	4	7	0%	0%	43%	57%
5	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	3	4	7	0%	0%	43%	57%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
5	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	0	0	4	3	7	0%	0%	57%	43%
5	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	3	4	7	0%	0%	43%	57%
5	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	1	1	5	7	0%	14%	14%	71%
5	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	3	4	7	0%	0%	43%	57%
5	Sum		0	1	17	24	42	0%	2%	40%	57%
5	percentage of total number of responses		0%	2%	40%	57%	0.99	0%	2%	40%	58%
7	1	I understood the process used for the Proficiency Determination	0	0	1	9	10	0%	0%	10%	90%

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	1	9	10	0%	0%	10%	90%
7	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	0	0	1	9	10	0%	0%	10%	90%
7	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	2	8	10	0%	0%	20%	80%
7	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	0	2	8	10	0%	0%	20%	80%
7	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	1	9	10	0%	0%	10%	90%
7	Sum		0	0	8	52	60				

Grade	Item No.	Question	strongly disagree	disagree	agree	strongly agree	Total	%	%	%	%
			1	2	3	4		#1	#2	#3	#4
7		percentage of total number of responses	0%	0%	13%	87%	1				
HS	1	I understood the process used for the Proficiency Determination	0	0	2	8	10	0%	0%	20%	80%
HS	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	6	4	10	0%	0%	60%	40%
HS	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	0	0	6	4	10	0%	0%	60%	40%
HS	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	3	7	10	0%	0%	30%	70%
HS	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	0	3	7	10	0%	0%	30%	70%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
HS	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	3	7	10	0%	0%	30%	70%
HS		Sum	0	0	23	37	60				
HS		percentage of total number of responses	0%	0%	38%	62%	1				
summary	1	I understood the process used for the Proficiency Determination	0	0	11	37	48	0%	0%	23%	77%
summary	2	I would defend the recommendations for Proficiency Determination rules against criticism that they are too rigorous	0	0	17	31	48	0%	0%	35%	65%
summary	3	I would defend the recommendations for Proficiency Determination rules against criticism that they are not rigorous enough	1	2	16	29	48	2%	4%	33%	60%
summary	4	I was given the opportunity to express my opinion about Proficiency Determination rules being considered	0	0	11	37	48	0%	0%	23%	77%

			strongly disagree	disagree	agree	strongly agree		%	%	%	%
Grade	Item No.	Question	1	2	3	4	Total	#1	#2	#3	#4
summary	5	I believe the process I participated in to recommend Proficiency Determination rules was fair and allowed me to recommend a Proficiency Determination that reflected my thinking	0	2	10	36	48	0%	4%	21%	75%
summary	6	I understand that the Proficiency Determination rules may be adjusted following the workshop	0	0	12	36	48	0%	0%	25%	75%
summary		Sum	1	4	77	206	288				
summary		percentage of total number of responses	0%	1%	27%	72%					

Appendix H. Auditor's Report

**Observation and Report on
English Language Proficiency Assessment for the 21st Century (ELPA21)**

Standard Setting Workshops Conducted

July 19-22, 2016

Observation and Report to
Pacific Metrics

Prepared by:
Gregory J. Cizek, PhD

July 25, 2016

Observation and Report on English Language Proficiency Assessment for the 21st Century (ELPA21)

Executive Summary

The English Language Proficiency Assessment for the 21st Century (ELPA21) consortium is a group of states that has developed an assessment system for English learners (ELs) to measure performance in four domains, Reading, Writing, Speaking, and Listening, and to produce an overall Proficiency Determination classification of EL student proficiency. Performance in each domain is reporting according to five performance levels: Level 1 - Beginner, Level 2 - Early Intermediate, Level 3 - Intermediate, Level 4 - Early Advanced, and Level 5 - Advanced. The overall Proficiency Determination is based on profiles of performance across the four domains and is reported as *Proficient*, *Progressing*, or *Emerging*.

An in-person standard setting workshop was conducted to derive cut scores to define the five levels using the Bookmark standard setting procedure (Lewis, Mitzel, Mercado, & Schulz, 2012). The workshop was conducted over four days on July 19-22, 2016 and involved 52 participants, split into two tables of four to five participants at each of six grade bands. There were four purposes for the workshop: 1) to derive cut scores for the thresholds of the Level 3 and Level 4 performance categories; 2) to articulate the system of cut scores across grades and domains; 3) to recommend profiles and decision rules for combining EL student performance across the four domains to yield an overall Proficiency Determination classification; and 4) to draft descriptors for each achievement levels (Achievement Level Descriptors, ALDs).

It was observed that the standard setting activities for the ELPA21 assessments were conducted professionally, efficiently and, with few exceptions, as described in the standard setting design document/plan. There were many strengths noted; few issues arose during the standard setting workshop. Issues that did arise were minor; they were handled deftly by the meeting facilitators and project staff; and they are unlikely to have a discernible negative effect on the results.

The procedures and processes used to derive recommended performance standards for the ELPA21 followed sound psychometric practices and no threats to the validity of the resulting performance standards were observed. One important source of validity information was not available at the time this report was written: the results of participants' evaluations.

With the assumption that the participants' evaluations do not suggest cause for concern, the available information and observations suggest that cut score recommendations produced at the workshop can be considered to be valid and reliable estimates of the cut scores for the ELPA21. Policy makers should have confidence that the recommendations from the standard setting activity were based on sound procedures, and produced trustworthy, valid, and defensible results.

Observation and Report on English Language Proficiency Assessment for the 21st Century (ELPA21)

I. Overview

The English Language Proficiency Assessment for the 21st Century (ELPA21) consortium is a group of states that has developed an assessment system for English learners (ELs). The ELPA21 assessments were administered operationally for the first time in spring 2016 and were designed to assess students across six grades or grade bands: K, 1, 2-3, 4-5, 6-8, and 9-12. Four domains are assessed at each grade band: Reading, Writing, Speaking, and Listening. Performance in each domain of the ELPA21 assessment is reported according to five performance levels: Level 1 - Beginner, Level 2 - Early Intermediate, Level 3 - Intermediate, Level 4 - Early Advanced, and Level 5 - Advanced. An overall proficiency determination is also reported for each student according to the following performance categories: Not Proficient, Nearly Proficient, and Proficient.

An in-person standard setting workshop was conducted over four days on July 19-22, 2016 at the downtown Westin Hotel in St. Louis, Missouri. A total of 49 participants (and three alternates) were selected to participate (eight or nine participants in each of the six grade bands, split into two tables of four to five participants at each of the grade bands). Participants were selected to meet general goals for representation, qualifications, and experience criteria established by member states of the ELPA21 consortium. Specific information on targets for these characteristics and how successfully they were met was not available to the auditor; however,

details concerning the realized characteristics of the sample are provided in the “ELPA21 Standard Setting Design” document (Pacific Metrics, 2016a). Specific information comparing intended/realized targets for gender, ethnicity, primary grade level expertise, years of experience, practice setting and role, state, special needs student experience, and so on would be desirable to support the validity and generalizability of the eventual cut score recommendations.

There were four purposes for the workshop:

- 1) to derive cut scores for the thresholds of the Level 3 and Level 4 performance categories;
- 2) to articulate the system of cut scores across grades and domains;
- 3) to recommend profiles and decision rules for combining EL student performance across the four domains to yield an overall proficiency determination; and
- 4) to draft descriptors for each achievement level (Achievement Level Descriptors, ALDs).

As it turned out, all of the 49 identified panelists were able to attend the workshop; a decision was made that, rather than dismiss the alternate panelists, they would be included in the appropriate grade-band groups and their judgments would be included in the data analysis. Thus, the total panel size was $n = 52$. The workshops were implemented using the Bookmark standard setting procedure (Lewis, Mitzel, Mercado, & Schulz, 2012).

The author of this report was contacted by Pacific Metrics to perform an independent, external observation/audit for the sessions and to submit a report of observations. The author has expertise and extensive experience in the area of setting performance standards (see, e.g., Cizek, 2001, 2012; Cizek & Bunch, 2007). In preparation for the workshop observations, the auditor reviewed the following materials:

* *ELPA21 Standard Setting Design* document (Pacific Metrics, 2016a);

* *ELPA21 Table Leader Orientation* [video] (Pacific Metrics, 2016b);

* *Standard Setting Table Facilitator Agenda* (Pacific Metrics, 2016c); and

* *Standard Setting High-Level Agenda* (Pacific Metrics, 2016d).

The remainder of this report provides a description of the standard setting activities, some recommendations, and a summary evaluation. The report is organized into four sections: 1) Overview; 2) ELPA21 workshop observations; 3) Summary and Recommendations; and 4) References.

II. ELPA21 Standard Setting Observations

Day 1: Morning Activities – Orientations and Reading Domain

The morning sessions on Tuesday, July 19, 2016 began at 7:30am with an orientation for table leaders. Dr. Daniel Lewis, Principal Research Scientist at Pacific Metrics, and co-facilitator of the workshop, led the orientation. The purpose of the table leader ordination was to ensure that the individual participants identified in advance to facilitate the work of each grade band table were familiar with their tasks and with use of the laptop computers at each table. Prior to the orientation, the table leaders participated in a two-hour webinar that provided them with specific training related to their role in the standard setting workshop.

The whole group of participants were then greeted by Mike Middleton (Director, Business & Select Assessments, Assessment and Student Information, Washington Office of Superintendent of Public Instruction) who welcomed participants, thanked them for their participation, and provided a general context for how the standard setting workshop activities fit into the ELPA21 development process. After the greeting by Dr. Middleton, participants were provided a detailed overview of ELPA21 domains, scores, policy achievement level definitions, and other aspects of the ELPA21 assessments by Dr. Mary Seburn, an independent consultant, serving as the ELPA21 Performance Standard Setting, Data and Score Reporting Task Management Team Leader.

At approximately 8:15am, the whole-group orientation continued with a welcome and informational session conducted by Dr. Lewis. The session addressed housekeeping issues, guidelines for cell phone use, reimbursements, non-disclosure/security agreements, and other logistics. Dr. Lewis then provided specific information on the following elements:

- * the purpose of the session;

- * nature of standard setting;
- * overview of the Bookmark standard setting process;
- * overview of the ELPA21 performance levels, with a focus on the Level 3 and Level 4 cut scores;
- * information on how the Level 2 and Level 5 cut scores will be set using the Contrasting Groups study data;
- * overview of the specific activities that participants would complete; and
- * introductions to the observers and other participants (e.g., state representatives, personnel from the Council of Chief State School Offices, content specialists) and their roles.

At approximately 8:40am, training in the Bookmark standard setting materials and process was provided by Dr. Karla Egan, Workshop Co-Facilitator, EdMetric, LLC. The materials included the item map, the Ordered Item Booklet (OIB), the ELPA21 assessment (participants were informed that they would be taking a form of the ELPA21 assessment as part of the workshop activities), and a review of the agenda and the specific activities planned for each day of the workshop. Dr. Egan ended the initial orientation with some time allocated to address participants' questions about EL assessment policy and the standard setting process. When the session was finished, Dr. Egan directed participants to complete a 10-item Readiness Survey, an instrument designed to gather information on participants' understanding of the goals and activities of the workshop and their evaluation of the initial orientation session. Dr. Egan invited all participants to raise their hands if they had any questions about the activities; there were no questions, and the Readiness Surveys were collected. This portion of the Day 1 morning activities was concluded by approximately 9:10am.

At approximately 9:15am, Dr. Egan led participants through the secure materials sign-out procedures. When participants completed this activity, Dr. Egan invited participants at each table to take several minutes to introduce themselves to each other.

At approximately 9:25am, Dr. Egan introduced Dr. Seburn to review the English Language Proficiency standards (ELPs), and the Achievement Level Indicators (ALIs) which were available to participants on the laptop computers at each table. Leslie Mugan (Vice President of Customer Experience and Program Management, Pacific Metrics) then greeted participants and provided an orientation to the secure browser and individual usernames and passwords necessary to access an online operational form of the ELPA21 assessment. Information on these procedures was also provided for participants in a document titled “Standard Setting Reminders” available at their tables.

As is typical of these kinds of activities involving large-scale coordinated group use of technology, participants experienced some difficulties signing in to the system, changing between browsers, applications, etc. Technical support personnel and workshop staff spent approximately 10-15 minutes to help participants who needed assistance to access all needed utilities.

Dr. Egan then projected a list of “Available Tests” that presented participants with the assessments available to them by grade band and, within grade band, domains. Participants accessed their respective Listening and Reading tests to interact with the items, gain familiarity with the interface and equipment (e.g., microphones, headsets), and how to enter data into the system. This activity lasted approximately 30 minutes, ending with review of the Reading test so that participants could move efficiently to the Reading-related activities.

A “floating” morning break occurred; when individual group members completed the introductory activities, they took an informal break. Table groups completed the activities at

different times; when a table was finished experiencing the test and had taken a break, table leaders then directed their tables to being a table-level review of their Reading Ordered Item Booklets (OIBs), item maps, and stimulus booklets. Table members then discussed each item in their Reading OIBs, considering two questions for each item: 1) “What does this item measure? That is, what do you know about a student who responds successfully to this item?” and 2) “Why is this item more difficult than the preceding items?” During this activity, Dr. Egan reminded the whole group about the primary purpose of the Reading OIB review; she pointed them to 3x5 index cards provided at each table on which participants could note issues they perceive with the construction, scoring, or other aspects of items, passages, rubrics, etc. and that their comments would be forwarded to the appropriate test development personnel.

The morning session adjourned at approximately 12:10pm when participants were dismissed for the lunch break. Before adjournment, Dr. Lewis polled the group as to their progress through the OIB in order to assess pacing for the afternoon activities, and to remind them regarding the rationale for review of the OIBs to note the information yielded by each item with respect to the knowledge and skills that an English learner would likely possess if he/she answered the item correctly.

Day 1: Afternoon Activities – Reading Domain

At approximately 1:05pm, the whole group was reconvened. The afternoon session began with some housekeeping announcements by Dr. Egan. Dr. Lewis then addressed pacing and process for the afternoon. He presented a plan to break out the whole group for the Bookmark training so that the training could be provided efficiently to the groups following their reviews of the OIBs; groups that finished their OIB reviews earlier were also instructed in a process for

analyzing the items in anticipation of creating the Achievement Level Descriptors (ALDs). This process helped to equalize the time spent in review of the test content and helped ensure equivalent depth of exposure across groups. The tables then resumed their reviews of their OIBs and item maps, focusing particularly on the knowledge and skills that distinguish performance on items from the knowledge and skill demands of the items that preceded them.

In response to the observation by the workshop facilitators that some grade band groups were completing the OIB/item map task earlier than others, at 2:30pm Dr. Lewis made an announcement that represented a slight modification of the prior schedule. It was announced that Dr. Egan would conduct Round 1 training, in a separate break-out room, in two separate sessions—the first session beginning at 2:45pm for groups that had completed the OIB review by that point, and the second beginning at 3:30pm for groups that took longer to complete the OIB review. More than once, Dr. Lewis emphasized that the scheduling was not intended to imply that any group(s) were working too quickly or too slowly, or to induce any participant to rush through the task at hand. The adjustment was intended only to ensure that panelists' time was used as efficiently as possible and to minimize unproductive waiting or “down time” for participants.

Bookmark procedure training for making Round 1 judgments was presented by Dr. Egan in a separate room from the whole group activities. The training began with an orientation to the *target student*—that is, the EL who is *just into* a specific Achievement Level—and target student descriptors. Participants were presented with existing Target Student Descriptors (TSDs) and a brief discussion ensued to help participants understand how the TSDs relate to the ALDs, EL student scaffolding/support needs at entry to an Achievement Level, how the TLDs relate to the content standards, ALIs, and achievement levels, and how they will use the TSDs in the Bookmark standard setting procedure.

At the end of the discussion at approximately 3:15pm, Dr. Egan gave the first group specific instruction on how to independently place their bookmarks in their OIBs, and information on the meaning of a specific bookmark placement. Participants were reminded that they would be placing only two bookmarks (i.e., setting two cut scores)—the Level 3 and Level 4 performance levels. Dr. Egan also presented information to illustrate the relationship between the bookmark placements and the cut score(s) on the ELPA21 test score scale. A number of participants asked follow up questions about the relationship between bookmark placements and cut scores; all questions were answered by Dr. Egan before moving on. She then proceeded to provide more specific information on the meaning of “mastery” of item content as a way of introducing the concept of response probability (RP) and the RP criterion used (.67) for the ELPA21 standard setting workshop. Because the time was running out with the second group awaiting their training, Dr. Egan wrapped up the training session at approximately 3:40pm and directed participants to ask any further or follow-up questions to Dr. Lewis. The training session for the first group concluded with the administration of a second process evaluation. Dr. Egan repeated the training for the second group beginning at 2:45pm (The auditor did not observe the Bookmark training for the second group, but followed the first group to observe the Round 1 bookmark placement activity.)

One observation relevant to the Bookmark training session was that the breakout room in which the training took place was fairly large, long, and narrow. At all times and without a microphone, Dr. Egan appeared to speak loudly and clearly enough for all participants to hear her presentation. However, at times, the room configuration and ambient noise appeared to make it difficult for some participants to hear the questions or comments of other participants. This limitation did not appear to significantly affect the group activity; for the future it might be advisable for the facilitator to repeat participants’ comments/questions for the group, or to have a

microphone available.

When a group's Bookmark training was completed, the group moved back to the whole-group room to begin their Round 1 judgments (i.e., placing Level 3 and Level 4 bookmarks) for the Reading domain. To begin this portion of the workshop, Dr. Lewis asked if there were any remaining questions from the Bookmark training session; there were none. Led by their respective table leaders, participants then began review of their OIBs to place Round 1 bookmarks. The workshop co-facilitator, Dr. Lewis, and other support staff circulated around the room during this portion of the session to ensure that all participants understood the task, to answer any questions, and to provide any assistance or clarifications requested. At approximately 4:15pm, the second group completed their Bookmark training and joined the first group in the whole-group meeting room where their table leaders also guided them to begin their Round 1 judgments.

When participants had completed their Round 1 ratings for the Reading domain, participants were able to review the Round 1 results within the automated system used for the standard setting workshop. A whole-group presentation or discussion of the results did not take place, but individual group conversations were noted among individual table groups. Instead, groups moved at their own pace to complete the review of their Round 1 ratings, then move directly to generating their Round 2 judgments.

As part of this portion of the workshop, the workshop design document and facilitator agenda (Pacific Metrics, 2016a, 2016c) described the provision of information related to external validity data from a Contrasting Groups Study (CGS), conducted prior to the standard setting workshop. However, due to a third-party delay in gathering and analyzing the data from the CGS, it was decided to omit inserting the CGS bookmarks into the OIBs as a source of information for the panelists. The change in procedures from the original design did not appear to cause any

concern or confusion on the part of the participants; its effect on their cut score recommendations is not known. It was decided that the CGS data would be reviewed, when available, by policymakers who would consider how the data would be used in making final performance standard decisions.

Some groups appeared to take longer to complete their ratings than was anticipated in the agenda for the meeting. Accordingly, the meeting facilitators decided not to rush participants to finish at the same time, but decided to slightly adjust the agenda to require only the completion of Round 2 ratings by the end of the first day, with final judgments (i.e., Round 3 ratings) completed at the beginning of the second day. Participants that finished early were dismissed for the day; participants that required additional time were encouraged to persist and complete their Round 2 ratings before leaving so that all participants could begin the Day 2 tasks as a whole group. A list of secure materials to be collected at the end of Day 1 was projected, and meeting support staff ensured that all secure materials were accounted for by the end of the Day 1 session. The first participants completed all tasks and began leaving at approximately 5:00pm; the last participant left the whole-group meeting room at approximately 5:35pm.

At approximately 5:20pm, while some participants were finishing their Round 2 ratings, the first of two debriefing sessions took place among contractor staff, table leaders, state representatives, other personnel associated with the ELPA21 project, and the auditor. The purpose of the debriefing was to evaluate the process, identify highlights of Day 1, issues encountered, and to consider potential changes to agenda, pace, or other aspects of the intended plan. No significant issues arose that were cause for concern; the first debriefing session ended at approximately 5:50pm. Following the first debriefing session, a second session was held beginning at approximately 5:55pm; this session was limited to state representatives and project personnel (and

the auditor). Participants in this session also noted several positive aspects of the day's activities; like the prior session, not substantial issues needing attention arose; the majority of the discussion focused on policy implications of the ELPA21 cut scores, gaps in the OIBs, suggestions for keeping participants on the same schedule, anticipated reactions to the impact data, and on potential changes in seating/table arrangements for Round 3.

Day 2: Morning Activities – Reading and Listening Domains

The second day of the workshop, Wednesday, July 20, 2016 began at 8:00am with an introduction to the day's activities by Dr. Lewis, who indicated that the grade band table groups would be working together (as planned) to finish their Reading judgments instead of remaining in separate table groups. Next, Ms. Cat Still (program manager, Council of Chief State School Officers) addressed some policy questions that had arisen among the panelists regarding the meaning of “proficiency” and how the groups’ judgments would be used, eventually, to help inform that term and resulting student decisions. Finally, Dr. Mary Seburn presented a review of the definitions of and relationships among the ELP Standards, the Achievement Level Indicators, the Policy Definitions of proficiency with particular emphasis on the “anchoring” role of the Level 4 Policy Definition, the draft Target Student Descriptors, the draft Achievement Level Descriptors used for standard setting, and the Achievement Level Descriptors used for reporting. Her presentation included a projected figure that illustrate the relationships. Using the figure, she describe how the panelists’ judgments fit into the process and she answered questions from panelists. Her presentation also included a review of ELPA21 scores—the different kinds of scores, the characteristics of each score, and relationship between the panelists’ standard setting activities and the subset of scores that are related to those activities. Her presentation also included

a review of the three proficiency classifications (that would be based on student profiles): *Proficient, Progressing, and Emerging*.

Following Dr. Seburn's presentation, Dr. Lewis then provided an overview of how the Reading domain activities were completed, the next domain to be addressed (Listening). He provided some suggestions for how tables could "synch up" to work at a more common pace and facilitate the groups finishing their tasks at approximately the same time (and promote keeping the entire workshop on schedule). He also reminded the whole group regarding the requirement to generate their first and second round ratings independently, not discussing their bookmark locations across tables until Round 3, and he provided instructions about generating the Round 3 Reading Domain ratings, collection of Reading materials, and distribution of Listening materials.

The final portion of Dr. Lewis' presentation focused on display of the groups' Round 2 results for the Reading domain. He projected a blank data summary sheet, showing participants how the summary descriptive information (table and grade band median Bookmark placements) and impact information (percentages of students that would be classified into Levels 1 & 2, Level 3, and Levels 4 & 5). Dr. Lewis then provided general information on how participants could use the impact information to adjust their bookmark placements if they perceive the impact as revealing the percentages of students classified at a given level as being too high or too low. He emphasized that, although prompted by their observations about the impact data, the participants should ground any considered bookmark adjustments in the content (i.e., Reading knowledge and skills reflected in the items) measured by the ELPA21 assessment.

Dr. Lewis then directed participants to open their grade band Excel files and review the impact information for grade bands. The first portion of this activity involved a brief QC task to verify that the data entry appeared to reflect the groups' bookmark placements; at approximately

8:50am, the groups then engaged in the impact information review and discussion activity. Using this information, the groups then proceeded to make their Round 3 bookmark placements for the Reading domain. It was unclear precisely how much time the groups spent on this activity; per instructions from the workshop facilitators, participants were instructed to complete their Round 3 judgments, complete a process evaluation, assemble all secure Reading domain materials for collection by workshop staff, sign out secure materials for the Listening domain, and begin their reviews of the Listening ALIs, OIBs, and item maps. As a helpful reminder to participants, these tasks were displayed on the main projection screen and visible to all tables. In general, the first groups appeared to complete their Round 3 ratings beginning at approximately 9:20am, at which point they turned to the standard setting tasks for the Listening domain.

The agenda for the meeting called for groups to complete all three rounds of their Listening domain judgments prior to the lunch break (i.e., noon) on Day 2. It appeared that none of the groups finished before the scheduled lunch break. An announcement was made by Dr. Lewis at 12:00pm, informing participants that lunch was ready and that participants could choose to adjourn at that time for lunch, or continue to work for a short time, taking lunch as was most convenient for them, and coming back from lunch on schedule to resume their Listening domain activities. Most table groups remained working past noon, adjourning for lunch when their tables finished one of the scheduled activities.

While participants were adjourned for lunch, the facilitators and other project personnel conferred to discuss how the remaining agenda might be reconfigured to ensure that all workshop activities were completed in the allocated time. Also discussed was a minor technological issue that did not allow participants to “pause” following completion of domain judgments; the application required participants working, for example, on the Writing domain judgments to re-

start the secure browser and click through Reading and Listening to get to the Writing.

Day 2: Afternoon Activities – Listening and Writing Domains

By 1:00pm, all participants returned from the lunch break and continued work on the Listening domain activities. All groups appeared to have completed their reviews of the OIB, ALIs, and item maps and were ready to begin Round 1 of standard setting judgments for Listening. By 1:35pm, some groups had completed two rounds of Listening domain judgments and were beginning to review impact information in preparation for making their final round of Listening judgments. By approximately 2:15pm, some groups had completed all of the Listening domain activities, and project staff began collection of secure Listening-related materials and distribution of secure Writing-related materials. The last group finished their Listening judgments by 4:00pm.

Upon receipt of the Writing domain materials, and moving back into smaller table arrangements, the grade band groups repeated the procedures that they had followed for the Listening and Reading domains for the Writing and Speaking domains; that is, they began by experiencing the Writing and Speaking assessments for their grade bands, then began working on review of the Writing domain ALIs, OIBs, and item maps. By 4:35pm, groups who had completed their review of the Writing OIBs and item maps concluded their work for the day. Their secure materials were logged in and they adjourned for the evening. All groups had left the meeting room by 4:55pm.

As was done at the end of Day 1, a debriefing session took place among contractor staff, table leaders, state representatives, other personnel associated with the ELPA21 project, and the auditor. The debriefing session began at approximately 5:00pm, by going around the assembled table leaders who were asked to provide brief comments on what they believed went well, and

suggestions, if any, for aspects that could be improved. Table leaders generally reported very favorable comments, including describing their groups as well-engaged, enjoying the process, and working at a pace that was generally improved on the second day (in part attributed to the pacing suggestions made by facilitators at the beginning of Day 2). Issues that arose that were identified for possible attention the next day included clarifications related to the writing rubrics, reminding participants that consensus in bookmark locations is not required, and reminding them that they should not spend time editing/critiquing items, but to focus on the bookmarking judgments.

At the conclusion of the meeting with table leaders (approximately 5:45pm), the follow-up debriefing meeting began with only project personnel, state representatives, and the auditor in attendance. Overall, those who summarized the day's event reported that the workshop had progressed well, participants were engaged and functioning appropriately, and that the procedures were being implemented as intended. The remaining and majority of the time in the debriefing session was spent on an update regarding data availability for the planned vertical moderation activity (a clean data file had not yet been confirmed), and substantial discussion occurred regarding contingency planning. The group considered how to proceed if the data were/were not available by a deadline of 2:00pm CDT on Thursday, July 21 and, if available, how to frame the vertical moderation session. The debriefing session concluded at approximately 6:45pm.

Day 3: Morning Activities - Writing and Speaking Domains

Day 3 of the standard setting workshop (Thursday, July 21, 2016) began at approximately 8:05am with an introductory presentation by Ms. Cat Still. Ms. Still provided information regarding "What happens next?" following the workshop and the participants' final cut score recommendation. The process she outlined (illustrated with powerpoint slides) included a review

by the ELPA21 Technical Advisory Committee (TAC) that will review the standard setting process and results, followed by a teleconference with member states and, eventually, a final set of cut scores that will be adopted across all ELPA21 member states.

Two other points were addressed by Ms. Still. First, she reminded participants that they were not required to come to consensus in their table groups. She emphasized that diversity of viewpoints and judgments was valued and uniformity of judgment was not expected or encouraged. Second, she reminded the whole group that they should not focus on issues they perceived with individual items. The primary focus is to make bookmark judgments related to items; comments or concerns about items should be recorded and passed along to project staff, but should not slow down their standard setting work.

Following Ms. Still's remarks, Dr. Karla Egan led a (re)introduction of participants by home state. She then began a presentation to aid participants in understanding how to place their bookmarks in the Writing OIB, with emphasis on the writing rubric, the constructed-response scoring, and the relationship of the Writing bookmarking activity to an activity that participants would engage in later in the workshop: producing Achievement Level Descriptors. In particular, she reminded participants about the appropriate level of detail of note taking when reviewing the OIB and item maps; the level of detail should be sufficient to describe the skills of an EL student at a given performance level, but not so much detail as to essentially represent a specific test item. Finally, Dr. Egan described a "cluster book"—a new resource for participants in making their judgments in the Speaking domain at grades K, 1, 3, and 5. Ms. Still provided additional elaboration on the format, structure, and scoring of cluster items. The opening presentations on Day 3 concluded at approximately 8:30am and, in their small grade band groups, participants resumed working on their Writing domain tasks. One group completed their Writing domain judgments substantially faster (by approximately 9:40am) than any of the other groups; at that

time, project support personnel began collecting the secure Writing materials and began distribution of the Speaking domain materials. All other groups continued working on Writing, with some finishing closer to the scheduled lunch break at noon and others staying somewhat past the beginning of the lunch break in order to finish their Writing domain judgments.

Day 3: Afternoon Activities - Speaking Domain

Following the lunch break, all participants returned to begin (or continue) their work on the Speaking domain tasks. No facilitator introduction or remarks were given; the table leaders appeared to help each group move seamlessly into the work and project staff ensured that all secure materials from the Writing domain were collected, evaluations were completed, and secure materials for the Speaking domain were distributed.

Because the table groups finished their Speaking domain tasks at different rates, grade band groups were dismissed for the day as they completed their work. The first group finished their work at approximately 4:35pm. They completed the required evaluation; their secure materials were collected, and they were dismissed for the day. The same procedure was followed as each subsequent group completed their work. Instead of the usual end-of-day debriefings with table leaders, as groups finished their work, Dr. Egan conducted “exit interviews” with table leaders to obtain their input on the strengths of the standard setting workshop and suggestions for improvement.

As had occurred at the end of each of the previous days, Day 3 ended with a debriefing session with project staff and state personnel which began at approximately 5:30pm. Because of limitations with the data files needed for vertical moderation, it was decided to forego the planned vertical moderation activity planned for the following day. Instead, the schedule was adjusted so that there would be brief opening remarks in the morning. Then groups that had finished their

Speaking domain activities would meet in a break-out room to receive training in ALD writing while other participants would continue finishing their Speaking judgments. The whole group would then be convened to participate in the Proficiency determination recommendation, followed by all groups completing their ALD writing tasks.

Day 4: Morning Activities – Speaking Domain, ALD Drafting, Proficiency Determination

The final day of the ELPA21 standard setting workshop (Friday, July 22, 2016) began at approximately 8:05am. Ms. Holly Garner, Director of Sales at Pacific Metrics, provided information in the whole-group setting on housekeeping matters, such as completion of expense reports, hotel checkout, airport transportation, and professional development certificates. This was followed by an update by Dr. Egan on the revised plan for Day 4. She announced that: 1) the Grade 5 group (who had finished their Speaking domain tasks the previous day) would meet in a separate room to receive training from Dr. Egan on writing their ALDs; 2) “rolling” training in ALD writing would occur for other groups as they finished their Speaking tasks; 3) groups would continue working on their Speaking domain activities until 10:00am; 4) at 10:00am, participants would meet as a whole group to participate in the overall Proficiency determination activities. The morning informational session ended at approximately 8:15am, and groups began their work.

An adjustment was made to the revised schedule because, at the planned 10:00am stop time some groups (notably, the groups working with the cluster items) needed additional time to work on their tasks. Additional time was given for two grade bands to complete their work; the remaining grade band groups (all of whom had received the ALD-writing training) used the time to begin working on their grade-level ALDs. At approximately 11:30am, the whole group was reconvened by Dr. Lewis, who gave a presentation to outline the next task: establishing the Proficiency Determination profile rules.

The planned process for making the Proficiency determination recommendation involves evaluating profiles of EL student performance across the four domains. Dr. Lewis' presentation included projected slides to illustrate profiles; he highlighted that Proficiency determination could involve differing performance profiles at different grade levels; and he reviewed the EL Expert Advisory Panel recommendations and the Contrasting Groups Study process and the role of that information in the Proficiency determination process. He explained that, following the consideration of the sources of information by the standard setting participants and their Proficiency Determination recommendations, the recommendations would then go to the CCSSO policy body for review and approval. His presentation also included a projected data chart showing the most common profile patterns found in the CGS study with their frequency of occurrence.

The final portion of the morning session was led by Dr. Seburn, who displayed an Excel worksheet that participants could refer to as they considered their Proficiency determination recommendations. The sheet include the policy definitions for each level, the most common profiles and the percentage of students with each profile by grade, ordered by overall level as identified by the respondents in the Contrasting Groups Study. Analysis of the profiles also suggested some "decision rules" for classifying profiles as Proficient based on specified performance patterns across the domains; the decision rules were provided in a separate Comments column for consideration by participants. Finally, some profiles in the data sheet were color coded red or blue. The color coding was used to highlight profile differences across grade band and where there was uncertainty identified in the Contrasting Groups Study about the overall proficiency classification of the particular profiles. Some whole-group discussion, questions and answers related to state-level policy decisions and use of the results, and additional directions were provided to help participants access the data sheets in the table groups and to guide their work that would begin after the lunch break. The group was dismissed at Noon.

Day 4: Afternoon Activities – Proficiency Determination, ALDs

Participants returned from the lunch break on Day 4, and the afternoon session began at approximately 1:15pm. Ms. Cat Still welcomed participants to the final session and thanked participants for their participation in the standard setting workshop. She encouraged the participants to be ambassadors for the ELPA21 and indicated that there would be additional information provided at the end of the session regarding what information about the process that could be disclosed following the standard setting meeting. She then oriented participants to the unique nature of the new content standards and assessments, and how those relate to EL students, the critical skills they need to be Proficient, and how all of those factors relate to their upcoming task of determining a Proficiency recommendation. She emphasized again that the groups were not required to come to consensus in those recommendations. Dr. Seburn also addressed the group to address any remaining questions about the next steps in the process.

Dr. Lewis then continued his presentation to provide additional background information on the Contrasting Groups Study results and the general rules that could be discerned regarding profiles of performance across domains associated with Proficient. He also reviewed the EL Expert Advisory Panel, who reviewed the Contrasting Groups Study results, discussed the various profile “rules”, and responded to a survey about specific profiles and their overall Proficiency determination. He initiated and facilitated a discussion among the whole group of participants regarding whether the profiles associated with *Proficient*, *Progressing*, and *Emerging* should be the same across all grade bands. The discussion lasted approximately 20 minutes and comprised detailed rationales and perspectives; it served as additional background for the whole group to break into their grade-band panels to further discuss the issue, and to develop grade-band recommendations that would be reported out to the whole group at the end of the discussion period.

Participants were also asked to consider development of profile rules for Proficient (and Emerging) determinations that would be clear and easy to communicate. He also describe conjunctive, compensatory, and mixed decision models, and he provided an explanation of the effect of conjunctive decision rules on classifications based on multiple sources of information—in this case, performances in the four domains. At the end of his presentation, at approximately 1:55pm, participants gathered in their grade-band groups to begin consideration of profile Proficient (and Emerging) determinations. Each table group had the questions to be answered, Excel files and other resource available on their table computers, as well as group recommendation recording forms.

At approximately 2:45pm, Dr. Lewis announced that an afternoon break had been readied. He encouraged participants to take a break as desired, then to assemble as grade-level groups to compare recommendations and prepare to report out to the whole group. He announced that the group reporting activity would begin at 3:00pm. When the whole group assembled, Dr. Lewis projected some of the grade-band suggestions for rules to define Proficient and Emerging. After projecting and discussing some of the grade group suggestions, he directed the groups to continue their discussions, enter their profile decision rules into the spreadsheets, save them to the Dropbox site, and prepare for reporting out to the entire group at the end of the activity.

At approximately 3:50pm, Dr. Lewis facilitated a discussion and led the grade level groups as they reported out to the whole group with their grade-band profile decision-rule recommendation. He emphasized again that it was not the purpose of the discussion to come to consensus on any individual grade-group decision rules, but to identify points of agreement and divergence. Drs. Lewis and Seburn independently captured the narrative rules described by panelists that would result in students being categorized into one of the three Proficiency Determination levels. Panelists' rules were also saved to Dropbox using Excel file worksheet

provided to panelists. The discussions concluded when each grade had an opportunity to communicate their recommendations defining the three Proficiency Determination profiles. At approximately 4:45pm, participants completed the Proficiency Determination task and overall workshop evaluations. The standard setting concluded at this time; participants were again thanked for their work over the course of the four-day workshop, secure materials were collected and audited, and participants were adjourned.

IV. Summary and Recommendations

Based on my observations of the procedures and processes used to obtain recommended performance standards, it is my opinion that the standard setting activities implemented for the ELPA21 assessments in Reading, Listening, Writing, and Speaking were, overall, conducted in accordance with the standard setting design plan intended, and in a professional manner consistent with sound psychometric practices. There was nothing observed that would lead to any other conclusion than the participants' cut score recommendations can be characterized as valid estimates of appropriate cut scores for the ELPA21 assessments.

Overall, the process was characterized by a number of strengths; few concerns also arose during the course of the standard setting. In the following sections some key strengths and considerations for the future are described.

Strengths

- 1) The contractor for setting ELPA21 performance standards developed appropriate and reasonably specific plans for implementing a professionally accepted standard setting method (i.e., the Bookmark method). The plan was reviewed in advance, approved by the relevant entities, and provided in advance to the auditor.

2) The implementation of the standard setting plan appeared to be a well-organized and faithful implementation of a modified Bookmark procedure. The contractor provided adequate resources and experienced personnel to ensure that the standard setting was conducted professionally and paced appropriately. The facilitators were skillful, prepared, non-intrusive, helpful, and provided clear guidance to the participants.

3) It appeared that all participants had strong qualifications for participation; they appeared to be knowledgeable about the relevant student populations, motivated to complete their work conscientiously, and they worked attentively and thoughtfully. No significant issues regarding personal agendas or domination of discussion in groups/tables were apparent. All participants appeared to be able to contribute to small and large group discussions as they desired and groups appeared to work collegially. Participants generally appeared to understand the nature of the feedback provided to them (i.e., the impact information) and likely used it appropriately. Participants who were identified as table leaders appeared to be well-trained and functioned well during the standard setting sessions.

4) Innovative and helpful technological utilities were incorporated for use in the workshop. These included, for example, specially-developed excel files and the use of a Dropbox application to link the excel files from the two tables at each grade band. This innovation allowed the table leaders to assist their groups to move seamlessly from round to round. The use of the Dropbox application allowed workshop facilitators to monitor data submission efficiently and without the need to collect paper rating forms. Hidden frequency distributions and item maps in the excel file allowed cut scores and impact data to be

estimated on the fly and also helped avoid data entry error.

5) Participants generally appeared to have sufficient time to complete their standard setting tasks; they did not appear to be rushed, nor was there excessive wait time/down time.

6) Technology used in the sessions (e.g., laptop computers, PowerPoint projections, data capture and data summary utilities, etc.) functioned well.

7) The meeting arrangements, food service, and other logistics were well-coordinated, appeared to be satisfactory for all participants, and generally appeared to facilitate the successful engagement in the standard setting tasks.

8) The materials, forms, and other items used appeared to be well-designed and easy for participants to use.

9) There was appropriate, conscientious concern for and attention to confidentiality and security of materials and results on the part of participants and project personnel.

10) At several junctures, Readiness or Process Evaluations were administered to participants and validity evidence based on the participants' responses was gathered.

Considerations for the Future

A few areas may warrant attention as information for policy makers, as caution for interpreting the results of the standard setting, or as information that states and/or contractors may

wish to consider when planning or implementing future standard setting activities.

1) According to the “ELPA21 Standard Setting Design” (Pacific Metrics, 2016a) document, participants were selected to meet general goals for representation, qualifications, and experience criteria established by member states of the ELPA21 consortium; that document also provides information on the characteristics of the group of panelists who were selected for participation in the workshop. Overall, it appears that a diverse and qualified group was empaneled, representing each of the ELPA21 consortium states. However, specific information on demographic targets and a comparison of intended/realized representation was not available to the auditor. Such information would be desirable to support the validity and generalizability of the eventual cut score recommendations.

2) Smaller table groups (n=4 to 5) versus larger table groups (e.g., n = 8 to 10) were used to foster effective table discussions. Most grade-band table groups were seated in close proximity to each other. At times, the proximity of the table groups, the overall size of the meeting room, and the enthusiastic engagement of the participants created an environment that was distracting for some participants. A helpful room modification was made at one point on Day 3 when a room divider was partially deployed in order to serve as a noise buffer during panelists’ work on their Speaking domain activities.

For the future, it may be advisable to physically separate grade band tables in the room (or use breakout rooms) to help minimize the potential for distractions and, primarily, to ensure the independence of the cross-table judgments. This recommendation assumes that, in addition to the goal of fostering discussion, an additional purpose of smaller,

separate table groups is to obtain estimates of standard error across tables/replicability of results across independent, equivalent panels.

3) Two fairly substantial deviations from the standard setting design document occurred:

a) external data from the Contrasting Groups Study was not provided as a source of information for participants, and b) the planned cross-grade vertical articulation activity was not conducted. Apparently, the omission of the Contrasting Groups Study data was due to a delay in obtaining, analyzing, and preparing the information for use in the standard setting workshop; likewise, data issues and timing most likely prompted the decision to defer the planned vertical articulation activity. Neither the Contrasting Groups Study information nor the vertical articulation activity is essential to the conduct of a successful Bookmark standard setting procedure; these are noted only because they represent departures from the intended procedures. It is not possible to determine the effect of this omission on panelists' judgments or the eventual cut score recommendations; the revised plan—i.e., that policy makers would review the relevant data when available and consider how the data would be used in making final performance standard decisions—is clearly within the defensible standard setting practices.

4) In my judgment, the Readiness and Process Evaluation forms used by participants could be improved to provide stronger procedural validity evidence. One improvement would be to phrase all questions so that they are not perceived (potentially) as participant self-evaluations. For example, consider questions phrased as the following: “I understand how to study the items in the ordered item booklet” or “I understand how to place my bookmarks” or “I understand how to interpret the impact data.” As phrased, the questions

essentially ask participants to self-evaluate their own understanding. In my experience, panelists—indeed, survey respondents in most contexts—are not apt to evaluate their own understanding negatively. The solution to this concern is easily accomplished: All questions can be phrased as an evaluation of the process or the facilitators. For example, the above questions might be rephrased as: “Information on how to study the items in the ordered item booklet was presented clearly” or “The facilitators did a good job of providing direction on how to place my bookmarks” or “The presentation on how to interpret the impact data gave me the understandings needed to complete that task.”

Also apropos to the evaluations, there did not appear to be any open-ended or other items where participants could indicate, for example, any aspect of the process about which they would like additional information before proceeding with their task. For example, there was an additional question that stated “I would like additional training on studying the ordered item booklet” followed by “No” and “Yes” options. However, because the forms were submitted anonymously it does not seem possible that individualized additional training could be provided should a “Yes” response be observed. And, the specific concern(s) that motivated the “Yes” response would not be known. In order to most efficiently provide whatever additional training might be needed, the above question might be rephrased as an open-ended format along the lines of “If you believe that additional clarification would be helpful related to any of the topics covered in the orientation session, please list the topic(s) below for which additional information would best prepare you to complete your standard setting tasks.”

Such a rephrasing—and analysis of these data—would be helpful in two ways. First, it would give facilitators specific information on the extent of clarifications required, and the specific topics that could be addressed. Second, by simply a quick visual review of

the responses, facilitators could determine that no additional review was required or, if appropriately documented in a report such as this one, that noted areas requiring clarification were addressed before proceeding with the next standard setting activity.

5) The pacing of the ELPA21 activities appeared to be somewhat challenging throughout the four-day process. To some extent, this challenge was unavoidable given the diversity of the groups (i.e., spanning kindergarten to high school levels). Workshop facilitators made some adjustments during the workshop that helped the grade-band tables to pace themselves and complete their tasks at approximately the same times. However, that adjustment did not address the pacing across *different* grade-band groups, and moderate variation in pacing, start/end of activities, etc., was still observed. For the future, it might be advisable to provide enhanced guidance to groups on the time allocated for each activity, and to implement some mechanism to aid groups in pacing their work (e.g., projected “Time Remaining” notices; announcements related to progress milestones, etc.).

6) The final activity of the standard setting workshop involved participants identifying proficiency profiles and decision rules for combining EL student performance in four areas to arrive at an overall proficiency determination. Although this activity appeared to be well facilitated and appropriately engaged in by the participants, the plan for the activity was an innovative approach for which there did not exist a professional “best practice” for conducting such an activity. For the future, if the procedure were repeated, it would seem essential to add one data source to the process: impact data. That is, it seems essential that participants, as they are contemplating decision rules for combining EL student performance in the four domains to be aware of the impact of the decision rules they are

recommending. For example, suppose a group were to recommend that, to be Proficient, an EL student must receive a score of 5 in three domains and at least a 4 in either Writing or Speaking. It is conceivable that the group would just such a rule to be more or less appropriate if it were shown to classify 95% or 2% of students, respectively, as Proficient.

It appears that there were two reasons why such impact data were not provided: 1) final results for producing accurate impact data were not yet available; and 2) a goal of the activity was for panelists to produce multiple, acceptable decision rules for subsequent consideration by the relevant policy bodies. Nonetheless, more explicit consideration by the participants of the impact of the recommendations would be likely not only to influence their eventual recommendations, but would also provide an additional source of validity evidence.

Conclusion

Overall, the standard setting procedures for the ELPA21 assessments were conducted professionally, efficiently and, with only minor modifications, as described in the standard setting design document/plan. There were many strengths observed (described previously) and only minor issues arose during the standard setting workshop. Issues that did arise were handled deftly by the meeting facilitators and project staff who provided additional explanations, clarifications, slight adjustments to the agenda, or appropriate modifications to the intended plan as deemed necessary. All such issues were not substantial and were unlikely to have any discernible negative effect on the procedures or results.

The procedures and processes used to derive recommended performance standards for the ELPA21 followed sound psychometric practices and no threats to the validity of the resulting performance standards were observed. One important source of validity information was not

available at the time this report was written: the results of panelists' evaluations. With the assumption that those data do not suggest cause for concern, the available information and observations suggest that it is warranted to conclude that the panelists' cut score recommendations should be considered to be valid and reliable estimates of the cut scores for the ELPA21. Policy makers should have confidence that the standard setting workshop was implemented as intended, that recommendations from the standard setting activity were based on accepted procedures, and that the procedures produced trustworthy, valid, and defensible results.

V. References

- Cizek, G. J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Cizek, G. J., & Bunch, M. (2007). *Standard setting: A practitioner's guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: SAGE.
- Cizek, G. J. (Ed.) (2012). *Setting performance standards: Foundations, methods, and innovations*. New York: Taylor and Francis.
- Lewis, D. M., Mitzel, H. C., Mercado, R., Schulz, E. M. (2012) The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations* (pp. 225-254). New York: Taylor and Francis.
- Pacific Metrics. (2016a). *ELAP21 standard setting design*. Monterey, CA: Author.
- Pacific Metrics. (2016b). *ELPA21 table leader orientation* [video]. Monterey, CA: Author.
- Pacific Metrics. (2016c). *Standard setting table facilitator agenda*. Monterey, CA: Author.
- Pacific Metrics. (2016d). *Standard setting high-level agenda*. Monterey, CA: Author.