



2021 No. 093

Summative English Language Proficiency Assessments for California (ELPAC) Alignment Study Report

Prepared for: California Department of Education
Assessment Development and
Administration Division
1430 N Street, Suite 4401
Sacramento, CA 95814-5901

Authors: Arthur Thacker
Emily Dickinson
Justin Paulsen
Wade Buckland

Prepared under: CN180100.2

Editors: Michele Hardoin
Christa Watters

Date: November 19, 2021

This page is intentionally blank.

Summative English Language Proficiency Assessments for California (ELPAC) Alignment Study Report

Table of Contents

Executive Summary	ES-1
Context and Overview of the Study	ES-1
Research Questions	ES-2
Review of Summative ELPAC Documentation	ES-2
Summative ELPAC Alignment Workshop and Outcomes	ES-5
Conclusions	ES-8
Recommendations.....	ES-10
Chapter 1: Introduction.....	1-1
Background.....	1-1
Research Questions	1-3
Organization and Contents of the Alignment Study Report	1-3
Chapter 2: Review of Summative ELPAC Documentation	2-5
Introduction	2-5
Method	2-5
Results	2-6
Summary and Discussion	2-22
Chapter 3: Summative ELPAC Alignment Workshop and Outcomes.....	3-23
Introduction	3-23
Summative ELPAC Alignment Criteria	3-23
Methods	3-27
Results	3-33
Summary and Discussion	3-46
Chapter 4: Conclusions and Recommendations	4-51
References.....	55
Glossary of Acronyms.....	57
Appendix A: Summative ELPAC Documentation Reviewed by HumRRO	A-1
Appendix B: Materials for Virtual Alignment Workshop	B-1
Appendix C: Summative ELPAC Item-person Maps	C-1

List of Tables

Table ES.1. Summative ELPAC Alignment Results	ES-8
Table 2.1 Rating Scale for Evaluating Strength of Evidence for Testing Standards	2-6
Table 2.2 Ratings on the Testing Standards for Summative ELPAC Alignment	2-7
Table 3.1 ELPAC to ELD Standards Alignment Criteria	3-23
Table 3.2 Summative ELPAC Alignment Panelists’ Demographics	3-28
Table 3.3 Summative ELPAC Alignment Evaluation Survey Results	3-32
Table 3.4 Results for Criterion 1: Link to Standards	3-33
Table 3.5 Link to ELD Proficiency Levels Kindergarten	3-35
Table 3.6 Link to ELD Proficiency Levels Grade One	3-35
Table 3.7 Link to ELD Proficiency Levels Grade Two	3-36
Table 3.8 Link to ELD Proficiency Levels Grades Three through Five	3-36
Table 3.9 Link to ELD Proficiency Levels Grades Six through Eight	3-37
Table 3.10 Link to ELD Proficiency Levels Grades Nine and Ten	3-37
Table 3.11 Link to ELD Proficiency Levels Grades Eleven and Twelve	3-38
Table 3.12 Grade Two Item-Person Map, Oral	3-40
Table 3.13. Range Adequacy Criterion: Mode of Communication: K and Grades 1, 2, and 3 through 5	3-43
Table 3.14. Range Adequacy Criterion: Ways of Using Language: K and Grades 1, 2, and 3 through 5	3-43
Table 3.15. Range Adequacy Criterion: Mode of Communication: Grades Six through Eight, Nine and Ten, and Eleven and Twelve	3-43
Table 3.16. Range Adequacy Criterion: Ways of Using Language: Grades Six through Eight, Nine and Ten, and Eleven and Twelve	3-44
Table 3.17. Range Adequacy Criterion: Standards Representation	3-44
Table 3.18. Balance of Knowledge Correspondence: Kindergarten	3-45
Table 3.19. Balance of Knowledge Correspondence: Grade One	3-45
Table 3.20. Balance of Knowledge Correspondence: Grade Two	3-45
Table 3.21. Balance of Knowledge Correspondence: Grades Three through Five	3-45
Table 3.22. Balance of Knowledge Correspondence: Grades Six through Eight	3-46
Table 3.23. Balance of Knowledge Correspondence: Grades Nine and Ten	3-46
Table 3.24. Balance of Knowledge Correspondence: Grades Eleven and Twelve	3-46
Table 3.25. Summative ELPAC Alignment Results	3-47
Table 3.26. Correlation Between Panelist ELD Proficiency Level Ratings and Metadata-based ELPAC Performance Levels	3-49
Table 3.27. Correlations Between Panelist ELD Proficiency Level Ratings, Metadata-based ELPAC Performance Levels, and Targeted ELPAC Performance Levels: Dichotomously Scored Items Only	3-50

List of Figures

Figure 3.1 ELD Proficiency Levels and Reporting Levels for ELPAC	3-26
--	------

Executive Summary

This report summarizes a study of the alignment between the Summative English Language Proficiency Assessments for California (ELPAC) and the 2012 California English Language Development (ELD) Standards. Alignment studies are required as part of the federal assessment peer review process, provide validity evidence that the assessment is measuring the intended content, and inform future assessment item development.

Context and Overview of the Study

An increasing number of non-native English speakers are enrolling in U.S. public elementary and secondary schools. Nationwide, the “English language learner (EL)” student group increased from 9.2 percent of enrolled students in fall 2000 to 10.2 percent in fall 2018. While enrollment of EL students in California has dropped in recent years (National Center for Education Statistics [NCES], 2020), the number and percentage (19.4 percent) of students who met the qualifications for this classification in fall 2018 make procedures related to assessing their English language proficiency (ELP) a continued priority.

Public education policy makers have taken an increasingly comprehensive approach to serving ELs. The Elementary and Secondary Education Act of 1965 (ESEA) offered grants to improve the quality of education and committed to equal opportunity for all students. With the reauthorization of ESEA in 2015 under the Every Student Succeeds Act (ESSA), the federal government provides funding to states and districts to identify ELs and implement specialized elementary and secondary programs and services to increase their English skills, which allows access to the instructional curriculum.

California identifies students in need of English language services and supports via the Initial ELPAC, and monitors English language proficiency across Listening, Speaking, Reading, and Writing via the Summative ELPAC. The Summative ELPAC is California’s state ELP assessment that EL students take every year until they are reclassified as fluent English proficient.

ESSA requires that annual ELP tests be available for ELs at all levels, from kindergarten through grade twelve (ESSA, 2015). In California, the Summative ELPAC is administered in seven grade spans—Kindergarten, one, two, three through five, six through eight, nine and ten, and eleven and twelve. In grades kindergarten and one, all domains are administered individually. In grade two, all domains are administered one-on-one except Writing, which is administered in small groups. In grades three through twelve, all domains are administered in a group, except for Speaking. The Summative ELPAC administration window is open from February 1 through May 31. Administration is computer-based (CDE, 2020).

The ELPAC is designed to measure performance on California’s ELD Standards. The standards were developed by California educators and published in 2012 after approval by the California State Board of Education. To evaluate the Summative ELPAC’s

alignment to the 2012 ELD Standards, we first investigated the nature of the assessment itself: how the standards guided the development of the test items (and how the standards and items should therefore relate to one another) and the interpretations to be made from ELPAC scores. This component of the study is described in *Chapter 2: Review of ELPAC Documentation*. Secondly, we modified traditional alignment methods to account for the test structure and design, a process in keeping with best practices in test validation that facilitates using alignment study results in an overall validity argument. This component of the study is described in *Chapter 3: ELPAC Alignment Workshop and Outcomes*.

Research Questions

Evidence of the alignment between assessments and standards is a requirement under the U.S. Department of Education’s assessment peer review process (Primarily addresses Peer Review Critical Element 3.1—Overall Validity, Including Validity Based on Content, but touches on other elements as well). Alignment evidence supports that students’ test scores can be used to make valid inferences about student performance on the content being tested. The CDE identified several research questions to guide the alignment evidence collected. Activities conducted for the ELPAC Alignment Study were designed to provide information to answer the following research questions:

1. To what extent do the test design and intended distributions for ELP domains (Listening, Speaking, Reading, Writing) for the Summative ELPAC support the claims to be made about student performance on the assessment?
2. To what extent do the 2021 Summative ELPAC test forms and test items reflect the test design and intended distributions for ELP domains?
3. To what extent do 2021 Summative ELPAC test forms show balance across the domains?
4. Does the Summative ELPAC include items that cover an appropriate range of cognitive complexity (or linguistic difficulty) to address the ELD Standards?

Review of Summative ELPAC Documentation

HumRRO researchers collected and reviewed Summative ELPAC design and test development materials provided by California Department of Education (CDE) and Educational Testing Service (ETS, the testing contractor) staff, as well as information about the ELPAC shared with the public on the CDE website. HumRRO researchers evaluated the alignment of the Summative ELPAC test design and development documentation to the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014; hereafter referred to as the *Testing Standards*).

First, HumRRO researchers identified fifteen specific standards from the *Testing Standards* that are directly relevant to how alignment is considered during test development. Next, researchers identified and collected the types of documentation

needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed the documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard. For each consensus rating, the researchers wrote a rationale to provide qualitative feedback related to the standard.

HumRRO developed and applied the following five-point rating scale to evaluate the degree to which the evidence for the assessment supports alignment to each standard:

1. No evidence of the Standard found in the materials.
2. Little evidence of the Standard found in the materials; less than half of the Standard was covered in the materials and/or evidence of key aspects of the Standard could not be found.
3. Some evidence of the Standard found in the materials; approximately half of the Standard was covered in the materials, including some key aspects of the Standard.
4. Evidence in the materials mostly covered the Standard.
5. Evidence in the materials fully covered all aspects of the Standard.

From the *Testing Standards*, we identified the following fifteen standards for review:

- Standard 1.9. When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.
- 1.11. When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.
- Standard 1.12. If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or

scorers are part of the argument for validity, similar information should be provided.

- Standard 1.14. When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.
- Standard 2.3. For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.
- Standard 2.16. When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure.
- Standard 3.2. Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.
- Standard 3.9. Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.
- Standard 4.0. Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.
- Standard 4.1. Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).
- Standard 4.6. When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.

- Standard 4.12. Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.
- Standard 4.16. The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.
- Standard 4.23. When a test score is derived from the differential weighting of items or subscores, the test developer should document the rationale and process used to develop, review, and assign item weights. When the item weights are obtained based on empirical data, the sample used for obtaining item weights should be representative of the population for which the test is intended and large enough to provide accurate estimates of optimal weights. When the item weights are obtained based on expert judgment, the qualifications of the judges should be documented.
- Standard 12.4. When a test is used as an indicator of achievement in an instructional domain or with respect to specified content standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes reflected in the target domain should be provided. Both the tested and the target domains should be described in sufficient detail for their relationship to be evaluated. The analyses should make explicit those aspects of the target domain that the test represents, as well as those aspects that the test fails to represent.

Fourteen of the identified standards were rated as fully covered based on the available evidence and one identified standard was rated as mostly covered. These results indicate that the ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to English language development standards (see *Chapter 2: Review of Summative ELPAC Documentation*).

Summative ELPAC Alignment Workshop and Outcomes

The alignment workshop was designed to collect evidence of whether the Summative ELPAC test forms effectively measure the content and linguistic rigor reflected in the targeted content domain and the test blueprint. During the workshop, educators with content expertise evaluated how well the 2021 test items represent the ELD Standards.

Alignment Criteria Evaluated

HumRRO developed alignment criteria based on documentation provided by CDE and ETS. These criteria represent several aspects of the overall alignment of the Summative ELPAC to the California ELD Standards. Failure to meet any single criterion does not

indicate that the test is invalid or flawed in some way, only that that aspect of the assessment may need to be addressed through future item development or by other means (e.g., blueprint adjustments, revisions to standards).

We drew on the concepts outlined in the Webb alignment method (1997, 1999, 2006), but tailored Webb's alignment criteria to be appropriate given the design of the Summative ELPAC. We also considered the growing literature on evaluating linguistic difficulty rather than depth-of-knowledge in English language proficiency assessments (Cook, 2005; 2007).

For a full discussion of how and why the alignment criteria were created, see chapter 3. HumRRO developed the following modified criteria for evaluating the Summative ELPAC: Link to Standards, Link to ELD Proficiency Levels, Range Adequacy, and Balance of Knowledge Correspondence.

Alignment Workshop Methods

HumRRO conducted a four-day ELPAC Alignment Study Workshop virtually via Microsoft Teams on March 22–25, 2021; each day included a four-hour session. HumRRO worked collaboratively with the CDE to recruit and select a group of 42 educators experienced with the ELD Standards to serve on seven ELPAC alignment review panels (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, grades 9–10, and grades 11–12).

HumRRO developed data collection tools and adapted several other materials to support the data collection process. Data collection tools included electronic spreadsheets for panelists and workshop facilitators to enter test item ratings (See Appendix B). Support materials included both paper and electronic copies of the (a) ELD Standards, (b) ELD Proficiency Level Descriptors (PLDs), and (c) detailed workshop outline and instructions for both panelists and facilitators. ETS created seven online test forms for the alignment workshop (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, grades 9–10, and grades 11–12) consisting of all the operational 2021 Summative ELPAC items. ETS also created accounts for HumRRO researchers and participants to securely access the items online using the IBIS™ Content Review Tool (CRT).

Alignment panelists received two rounds of training at the outset of the virtual alignment workshop. First, the full group of panelists received general training that provided some background on alignment and a high-level description of the alignment process. Following the general training session, panelists moved into grade-level/span panel groups and received more detailed training on the data collection processes and procedures.

After the panel-specific training presentation by the HumRRO facilitator, each panel engaged in a calibration activity using the first 1–3 items. Panelists accessed the items electronically and made their independent ratings. Panelists discussed their independent ratings and engaged in consensus discussion to come to agreement on

the final item ratings of record. Once panelists had a clear understanding of the rating process and a common understanding of the rating categories applied, they moved on to rating the remaining operational items. Items were reviewed and rated by domain, in the following order: Listening, Speaking, Reading, and Writing.

Item ratings were generated via the following steps:

1. Panelists reviewed test items and relevant scoring information independently and assigned ratings of:
 - a. Primary ELD standard measured by item
 - b. Secondary ELD standard measured by item (Speaking and Writing items only)
 - c. ELD proficiency level (using a six-point scale) for each item score point (multi-point items can assess differing levels of ELD proficiency). The rating scale was based on the ELD Proficiency level continuum defined in the ELD Standards: (1) Early Emerging, (2) Exiting Emerging, (3) Early Expanding, (4) Exiting Expanding, (5) Early Bridging, and (6) Exiting Bridging.
2. Panelists discussed their independent ratings and came to initial consensus
3. Panelists came to consensus (or majority) ratings
4. HumRRO facilitator recorded consensus/majority ratings
5. HumRRO facilitator shared item metadata and finalized the ratings

The HumRRO facilitator recorded the final consensus (or majority) item ratings in a spreadsheet. Once all consensus ratings were recorded, panelists completed three online surveys: a demographic questionnaire, a debriefing form, and a process evaluation survey. The debriefing form was designed to give panelists the opportunity to describe their overall view of the quality of alignment. The anonymous evaluation survey (optional) elicited feedback about the quality of the workshop, including panel facilitation, materials, and processes (see chapter 3 for more detail on workshop processes and procedures).

Alignment Workshop Results

Table ES.1 summarizes the alignment criteria results for the Summative ELPAC for all grades/grade-spans. These results show that the Summative ELPAC items are linked to standards across all assessments, although grade one panelists did not indicate secondary standards for several items. HumRRO aggregated ELD proficiency level ratings to three levels (Emerging, Expanding, and Bridging) for reporting purposes. The Link to ELD Proficiency Levels criterion showed that most Summative ELPAC items reflected knowledge, skills, and abilities associated with the Emerging or Expanding proficiency levels. Few items reflected the knowledge, skills, and abilities associated with the Bridging level, and no test met the requirement of 25 percent of items at the

Bridging level. At the higher grades, most of the items were rated as Expanding, while lower grades had more Emerging items. The Summative ELPAC tended to do a good job addressing most of their associated standards, especially those linked to modes of communication (categorized in Part I of the ELD Standards as Collaborative, Interpretive, and Productive). Fewer items were associated with ways of using language (categorized in Part II of the ELD Standards as Structuring Cohesive Texts, Expanding and Enriching Ideas, and Connecting and Condensing Ideas), especially in grades one, two, and the three through five grade span. The Summative ELPAC is not evenly balanced by number of test items for either modes of communication or ways of using language.

Table ES.1. Summative ELPAC Alignment Results

Criteria	K	1	2	3–5	6–8	9–10	11–12
Link to Standards	Met	Partially Met	Met	Met	Met	Met	Met
Link to ELD Proficiency Levels	Partially Met	Partially Met	Not Met	Not Met	Not Met	Not Met	Not Met
Range Adequacy	Met	Partially Met	Partially Met	Partially Met	Met	Met	Met
Balance-of-Knowledge Correspondence	Partially Met	Not Met	Partially Met	Partially Met	Met	Partially Met	Partially Met

While the results include several instances where the Summative ELPAC partially or did not meet the alignment criteria established before the workshop, they do provide a great deal of information that could be used to improve the ELPAC. See chapter 3 for a discussion of each criterion, descriptions of possible follow-up analyses, and potential changes to the alignment criteria for future alignment investigations.

Conclusions

This study combined documentation review and item ratings by content experts to evaluate the alignment between the Summative ELPAC and the California ELD Standards. Here we present the conclusions reached for each of the four research questions posed at the beginning of the study:

Research Question 1: To what extent do the test design and intended distributions for ELP domains for the Summative ELPAC (Listening, Speaking, Reading, and Writing) support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprints for the Summative ELPAC support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment (see table 2.2). Review of 2021

Summative ELPAC test forms found that all test items are linked to the California ELD Standards (see table 3.4). Test forms include substantive sections related to each of the ELP domains, but the sections are not evenly distributed in terms of numbers of items or points. Writing and Listening domains require more time per item than Reading and Speaking Domains and have fewer items. The ELPAC is designed to produce interpretable scores at the domain level in addition to an overall score.

Research Question 2: To what extent do the 2021 Summative ELPAC test forms and test items reflect the test design and intended distributions for ELP domains?

Data from the alignment workshop component of the study provide mixed support for the overall alignment of the Summative ELPAC to the ELD Standards. Some items intended to measure multiple standards in grade one were rated as measuring only a single standard. Items were well-aligned by domain based on panelists' ratings (e.g., Listening items were linked to Listening standards). The ELD Standards are also identified as either addressing "modes of communication" or "ways of using language." The Summative ELPAC test forms tended to do a good job addressing the standards linked to modes of communication. Fewer items were associated with ways of using language. The Summative ELPAC is not evenly balanced by number of test items for modes of communication compared to ways of using language.

Research Question 3: To what extent do 2021 Summative ELPAC test forms show balance across the domains?

The number of items per language domain varies greatly for all grade levels. Listening and Reading tend to have many more items, all machine scored, than Speaking and Writing, which include items that are hand scored using rubrics. Speaking and Writing items are the only ones that routinely include secondary standards. This leads to tests that are not balanced by strict numbers or percentages of items. If the imbalance is intentional, it may be prudent to capture that imbalance in test blueprints so that it can be addressed in future alignment analyses. This would not require a change to the blueprints, but an indication of the number of points and/or items associated with each domain. Alternatively, the test blueprint could more directly indicate that balance is achieved at the score reporting level, which combined reading and writing into a written language score and speaking and listening into an oral language score.

Research Question 4: Does the Summative ELPAC include items that cover an appropriate range of cognitive complexity (or linguistic difficulty) to address the ELD Standards?

Most Summative ELPAC items reflected knowledge, skills, and abilities associated with the Emerging or Expanding proficiency levels, and few items reflected the knowledge, skills, and abilities associated with the Bridging proficiency level. Lower grades had more items that assessed the Emerging level.

Recommendations

We offer four recommendations based on the results from this alignment study. In several instances, the recommendations are presented as options to either revise the assessment or the guiding documents to make alignment priorities clear.

Recommendation 1. Review grade one Summative ELPAC items that are intended to measure multiple standards to verify that students must demonstrate language abilities related to the intended secondary standard to correctly respond to the item.

The grade one panel found six items that were intended to measure multiple standards measured only one standard. This led to weaker alignment statistics for grade one than for other grades. We recommend reviewing those items prior to making more substantive revisions to the assessment.

Recommendation 2. Develop additional items at the Bridging ELD proficiency level for all grades.

The Link to ELD Proficiency Level is a new criterion for alignment studies. It was created for this study to examine how the Summative ELPAC items assessed the knowledge, skills, and abilities of students as described in the ELD proficiency levels. The ELD proficiency levels include much of the same information as would be found in a linguistic difficulty scale, plus other information to describe student performance more fully at each proficiency level. The ELD proficiency levels are also specific to California, whereas no linguistic difficulty scale is specific to California.

The ELD Standards, however, are not designed to address a single proficiency level. The standards documents describe student performance at each proficiency level by modes of communication or ways of using language, and those constructs are represented by small groups of standards. It would not be appropriate to assign a “level” to each standard for matching since each standard can be demonstrated at multiple proficiency levels. This issue is often overlooked when a linguistic difficulty scale is used because each standard must be assigned a linguistic difficulty level that can be matched to item ratings to compute an index. However, California’s ELD proficiency levels demonstrate just how problematic that practice can be if the standards are not discrete by level (and they rarely are). If the standards may be addressed at multiple levels, then it is more sensible for an alignment study to verify that the items address the full range of linguistic difficulty, or ELD proficiency levels, intended by the assessment. The Link to ELD Proficiency Level criterion takes this approach.

It should be noted that there were no test blueprints that indicated how many items should be expected to target each ELD proficiency level, nor was there information related to ELD proficiency level in the item metadata. The ranges for how many items were expected to fall in each level (per the alignment criteria) were established logically, rather than empirically. We reasoned that an assessment that included three proficiency levels (each split into “Early” and “Exiting” sublevels) should include a substantial

number of items addressing each level. Alignment criterion thresholds were set at 25 percent as the minimum percent of items at any level (Emerging, Bridging, Expanding) and 50 percent as the maximum. This distribution would allow for similarly accurate score estimation across the full range of performance described by the ELD proficiency levels. Accurate scoring at all levels is one key for answering Research Question 1 and supporting the claims associated with students' ELPAC performance level. The accuracy of scores toward the higher range is particularly important since students typically need to score toward the upper half of the Bridging level in all language domains to be considered proficient overall.

Results show that the items tend to be clustered toward the lower (Emerging) and middle (Expanding) ELD proficiency levels, especially at the higher grades. Lower grades had more items that assessed the lower level. No Summative ELPAC met the threshold of 25 percent of items at the higher (Bridging) level. The Summative ELPAC has three important threshold scores and generates four performance categories, one of which is toward the upper end of the ELD proficiency levels.

Panelists' results were compared with item-person (Wright) maps provided by ETS. Those data support the panelists' findings. There are few items on the ELPAC with difficulties in the part of the scale associated with the top score category. This means that students who score in that category must get nearly all of the test items correct and that the accuracy of those scores may not be as high as for other parts of the ELPAC.

Recommendation 3. Revise test items and/or metadata to better reflect the specific “ways of using language” standards that are intended to be measured.

The Summative ELPAC assessments were not evenly balanced with respect to modes of communication or ways of using language. The differences were sufficiently large that the test did not seem designed to address each of these concepts with an equal number of items, as is suggested by the Webb alignment method. If the assessments are designed to emphasize certain content over other content, or if there are items that should be given more weight, those decisions should be reflected in a test blueprint or similar document. This alignment study set a criterion for balance that assumed equal emphasis, but those analyses may be inaccurate if that assumption was incorrect. It is also concerning that the metadata indicated as many as eight secondary standards associated with a single item. Inclusion of more specific targets in the blueprints and/or item metadata could help ensure that these standards are a priority during item development.

The Range Adequacy criterion was met for the Summative ELPAC in kindergarten and grade spans six through eight, nine and ten, and eleven and twelve. Tests in the other grades/grade spans did not meet this criterion because they did not include enough items aligned to the standards reflecting “ways of using language.” These standards might be considered more difficult to measure than those associated with “modes of communication.” These standards were often assigned to items as secondary ratings. The grade one panel indicated that several items that were intended to measure secondary standards did not do so clearly enough to reach a consensus rating.

Recommendation 4. Refine the test blueprint to specify (a) the number and/or percentage of items associated with each of the modes of communication and ways of using language and (b) the number and/or percentage of items at each ELD proficiency level.

If adjustments to the test blueprint are adopted, it would also be helpful to include proportions of items at each ELD proficiency level. A test blueprint can be a valuable tool for ensuring that the test designers and developers create an assessment that adequately reflects the intentions of educators in measuring the ELD Standards. It can also provide a roadmap for future investigations of test-to-standards alignment.

Chapter 1: Introduction

Background

English learners (ELs) are an important and diverse student population, and their development of English proficiency is vital to success in school and society. Over the past several years, English language proficiency (ELP) for ELs has grown in focus, and rightly so, given the growing presence of ELs in U.S. classrooms. As a result, current versions of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014), hereafter referred to as the *Standards*, as well as the U.S. Department of Education’s (USDOE) *Assessment Peer Review Process* guidance (USDOE, 2018) give specific attention to ELs and ELP assessments. Per the *Standards*, “ELP tests are based on ELP standards and are held to the same standards for precision of scores and validity and fairness of score interpretations for intended uses as are other large-scale tests (AERA, APA, NCME, 2014, p. 191).” To this end, *Standard 4.8* recommends that experts, independent of the test developers, judge the degree to which content matches content categories in the test specifications and whether test forms provide balanced coverage of the targeted content. Similarly, *Standard 4.12* references “an independent study of the alignment of test questions to the content specifications” to validate the developer’s internal processes for ensuring appropriate content coverage (AERA, APA, NCME, 2014, p. 89). Federal Peer Review guidance also notes specific critical elements pertaining to EL alignment studies to meet statutory and regulatory requirements under the Every Student Succeeds Act (ESSA). Per Federal peer review guidelines, ELP assessments must be aligned to the ELP standards and measure ELs proficiency levels in domains of Listening, Speaking, Reading, and Writing. Furthermore, federal peer review guidance requires evidence of alignment through such means as an external independent alignment study.

State and federal laws require that all students whose primary language is other than English be assessed for English language proficiency (ELP). The legal basis for requiring ELP testing, as stated in the California Code of Regulations, Title 5 (5 CCR), Section 11518, is that all students have the right to an equal and appropriate education, and any English language limitations left unidentified and/or unaddressed could preclude a student from accessing that right. The English Language Proficiency Assessments for California (ELPAC) is the state’s designated test of ELP. It is administered (1) as an initial assessment to newly enrolled students whose primary language is not English, as indicated on a home language survey (HLS); and (2) annually as a summative assessment to students who have been previously identified as EL students (CDE, 2020).

In November 2012, the California State Board of Education adopted the English Language Development Standards (2012 ELD Standards). The 2012 ELD Standards are aligned with key knowledge, skills, and abilities described in the California Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects. The ELPAC is aligned with the 2012 ELD Standards (CDE, 2012). This report details a study conducted by the Human Resources Research

Organization (HumRRO) to evaluate the quality of the alignment between the Summative ELPAC and the ELD Standards.

The Summative ELPAC is one component of a system that also includes the Initial ELPAC. While the Initial ELPAC identifies students who are English learners (ELs), the Summative ELPAC assesses the progress of ELs in the domains of Listening, Speaking, Reading, and Writing. The Summative ELPAC is administered to seven grades or grade spans: kindergarten, grade one, grade two, grades three through five, grades six through eight, grades nine and ten, and grades eleven and twelve. The grade-span assessments have the same items for all grades within the span. In 2019, the ELPAC was transitioned from a paper-based assessment to a computer-based assessment. The first operational, computer-based form was finalized for use in spring 2020. The ELPAC is currently a fixed-form (non-adaptive) assessment.

HumRRO approaches alignment studies as one means to gather validity evidence to demonstrate the quality of intended interpretations and uses of assessment scores. That is, alignment studies indicate whether a test effectively measures what it is intended to measure.

The ELPAC is designed to measure students' language proficiency across four domains: Listening, Speaking, Reading, and Writing. Individual student scores for all grade levels Kindergarten through twelve, include:

- an overall score based on a continuous scale, four English performance levels (Levels 1–4 named Beginning to Develop, 2 Somewhat Developed, 3 Moderately Developed, and 4 Well Developed)
- an oral language subscore that reflects performance on the Listening and Speaking domains based on a continuous scale, four English performance levels (Levels 1–4)
- a written language subscore that reflects performance on the Reading and Writing domains based on a continuous scale, four English performance levels (Levels 1–4), and
- the student's proficiency within each domain (i.e., Listening, Speaking, Reading, and Writing), three English performance levels (Beginning to Develop, Somewhat/Moderately, Well Developed).

For the ELPAC, evaluating alignment involves examining the test items in terms of their representation of the ELD Standards, as well as the test's capacity to indicate students' readiness for content taught in English in academic courses. The California ELP standards provide a framework for English language development, but we also consider the language demands required to access the academic content represented by California standards for English language arts, mathematics, and science.

The first step in conducting the investigation for the ELPAC alignment was to investigate the nature of the assessment itself, how the standards guided the development of the

test items (and how the standards and items should therefore relate to one another), and the interpretations to be made from ELPAC scores. HumRRO then modified traditional alignment methods to account for the test structure and design, a process in keeping with best practices in test validation that facilitates using alignment study results in an overall validity argument. This process also supports federal peer review goals.

Research Questions

Activities conducted for the Summative ELPAC Alignment Study were designed to provide information to answer four research questions. HumRRO developed these research questions to ensure that the study focused on the key aspects of test-to-standards alignment. Research questions were informed by prior alignment studies and published methodologies and federal peer review guidance related to alignment. The research questions include:

1. To what extent do the test design and intended distributions across ELP domains for the Summative ELPAC (Listening, Speaking, Reading, Writing) support the claims to be made about student performance on the assessment?
2. To what extent do the 2021 Summative ELPAC test forms and test items reflect the test design and intended distributions across ELP domains?
3. To what extent do 2021 Summative ELPAC test forms show balance across the domains?
4. Does the Summative ELPAC include items that cover an appropriate range of cognitive complexity (or linguistic difficulty) to address the ELD Standards?

Organization and Contents of the Alignment Study Report

The remaining chapters and appendices of this report describe the Summative ELPAC Alignment Study activities, findings, and conclusions.

- Chapter 2, *Review of Summative ELPAC Documentation*, presents the methods, rating scale, and data analysis activities HumRRO conducted to evaluate the alignment of development documentation of the Summative ELPAC to relevant *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014), hereafter referred to as *Testing Standards*. The chapter identifies the list of Summative ELPAC documents reviewed for each test standard and describes the rationale for HumRRO's alignment rating. The chapter concludes with a summary of HumRRO's evaluation of Summative ELPAC documentation.
- Chapter 3, *Summative ELPAC Alignment Workshop and Outcomes*, presents HumRRO's method for evaluating the alignment of the 2021 Summative ELPAC test forms to the ELD Standards and Summative ELPAC blueprint. The chapter presents HumRRO's four alignment criteria; describes the alignment workshop

data collection activities, including panelist training and item rating procedures; and presents results of data analysis. The results section provides outcomes by grade level or grade span (i.e., kindergarten, grade 1, grade 2, grades 3-5, grades 6-8, grades 9-10, grades 11-12) for each alignment criterion. The chapter concludes with an overall summary of HumRRO's evaluation of the alignment of Summative ELPAC grade-level/grade span test forms, by alignment criterion.

- Chapter 4, *Conclusions and Recommendations*, presents HumRRO's overall alignment study conclusions as informed by results of the Summative ELPAC documentation review and the Summative ELPAC item ratings by content experts. The chapter offers six recommendations based on HumRRO's evaluation of the alignment of Summative ELPAC grade-level/grade span test forms.
- Appendix A, *Summative ELPAC Documentation Reviewed by HumRRO*, lists the file names of all documents reviewed for the study. Documents are grouped by these topics of focus: (a) ELD Standards, (b) test design, (c) item development and information, (d) test administration, (e) item scoring, (f) score reporting, (g) accessibility, (h) field test, and (i) standard setting.
- Appendix B, *Alignment Workshop Materials*, includes documents provided to content experts participating in the workshop. Materials include the workshop agenda, panelist item rating instructions, sample panelist rating form, and questions from three surveys administered online at the conclusion of the workshop (overall debrief, evaluation of alignment workshop training and procedures, demographic information).
- Appendix C, *Summative ELPAC Item-Person Maps*, presents ETS's item-person maps, which display for each grade level or grade span, the comparison between Summative ELPAC item difficulty and student performance (thetas).

Chapter 2: Review of Summative ELPAC Documentation

Introduction

To begin the alignment study and build knowledge of the English Language Proficiency Assessments for California (ELPAC), HumRRO researchers collected and reviewed ELPAC design and test development materials provided by California Department of Education (CDE) and Educational Testing Service (ETS) staff, as well as publicly available information about the ELPAC shared on the CDE website.

HumRRO researchers completed the first major task of the alignment study by conducting an evaluation of the alignment of ELPAC test design and development documentation to the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014; hereafter referred to as the Testing Standards). This chapter presents the methods and outcome of the evaluation of ELPAC documentation. The review of ELPAC materials also informed HumRRO's plans and preparation for the second major task of the study, the alignment workshop.

Method

HumRRO's evaluation of the test design and development documentation was informed by industry best practices as outlined in the Testing Standards. First, HumRRO researchers identified specific standards from the Testing Standards that are directly relevant to how alignment is considered during test development. We identified standards from Chapter 1 (Validity), Chapter 2 (Reliability/Precision and Errors in Measurement), Chapter 3 (Fairness in Testing), Chapter 4 (Test Design and Development), and Chapter 12 (Educational Testing and Assessment). Next, researchers identified and collected the types of documentation needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed the documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard.

Documents Collected

HumRRO worked in cooperation with the CDE and ETS to obtain documentation related to the design and development of the ELPAC. We also searched ELPAC website pages to identify additional relevant information. Appendix A lists the full complement of documents HumRRO collected and reviewed. The documents generally focus on the following areas: California ELD Standards; test design; item development information; test modality; test fairness, accessibility, and accommodations; and scoring and administration.

Rating Scale

HumRRO developed a rating scale to evaluate the degree to which the evidence for the assessment supports adherence to these testing standards. The rating scale ranged from 1 to 5, with higher scores indicating stronger evidence of compliance with the standard (See table 2.1).

Table 2.1 Rating Scale for Evaluating Strength of Evidence for Testing Standards

Rating Level	Description
1	No evidence of the Standard found in the materials ^a
2	Little evidence of the Standard found in the materials ^a ; less than half of the Standard was covered in the materials and/or evidence of key aspects of the Standard could not be found.
3	Some evidence of the Standard found in the materials ^a ; approximately half of the Standard covered in the materials, including some key aspects of the Standard.
4	Evidence in the materials ^a mostly covered the Standard.
5	Evidence in the materials ^a fully covered all aspects of the Standard.

^a Materials include all documents and data provided, any emails or phone calls with CDE and/or ETS staff, as well as information available on the CDE website.

Results

Ratings for Testing Standards

The results in table 2.2 represent the outcome of HumRRO's review of assessment planning and item development processes. The leftmost column in table 2.2 presents the evaluated testing standards.¹ Standards are numbered to reflect the chapter in which they appear and their order of presentation in the chapter. The center column lists the names of the files considered as supporting documentation for the processes and procedures related to each evaluated testing standard. Finally, the rightmost column provides an overall rating for each testing standard based on our review of this supporting documentation.

¹ To address *Standard 4.8*, HumRRO conducted a workshop with subject matter expert panelists, as reported in chapter 3, rather than independently evaluating the testing contractor's documentation as evidence.

Table 2.2 Ratings on the Testing Standards for Summative ELPAC Alignment

Standard	Supporting Documentation	Standard Rating
<p>Standard 1.9. When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.</p>	<ul style="list-style-type: none"> • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 706-2020A_v3_FOR ARCHIVE ELPAC 2020 New Item Review PPT_123019 • Participants List Spreadsheet 092820 • ELPAC Moodle Training Site 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 1.11. When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.</p>	<ul style="list-style-type: none"> • B-25 REFERENCE-ELPAC Report on Domain Specific PLDS_v4_APP • Definitions of Task Types for the ELPAC 3-2020 • ELPAC_Stimulus_WordCounts • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 685-2021 v2 FOR ARCHIVE ELPAC Summative Test Development Specifications_062220 • 690-2020A V2 FOR ARCHIVE Revised_2019-20 ELPAC Educator IWT Presentation 090919 • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 	5
<p>Standard 1.12. If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.</p>	<ul style="list-style-type: none"> • Mode Comparability Study memo (aug20adad02.docx) • C-211 Summative ELPAC Standard Setting Technical Report v6_APP • DFAs for all grades/grade spans • California ELD Standards (eldstndpublication14.pdf) • ELPAC Design_Standards and Practices_Final with Audio • ELPAC_Writing-Rubrics • Integrating CA ELD Standards in K-12 Math and Science Teaching and Learning 12-16-2015 • sep17item18 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
Standard 1.14. When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. Where composite scores are developed the basis and rationale for arriving at the composites should be given.	<ul style="list-style-type: none"> • sep17item18 • C-211 Summative ELPAC Standard Setting Technical Report v6_APP • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean 	5
Standard 2.3. For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.	<ul style="list-style-type: none"> • C-211 Summative ELPAC Standard Setting Technical Report v6_APP • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean 	5
Standard 2.16. When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure.	<ul style="list-style-type: none"> • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 3.2. Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.</p>	<ul style="list-style-type: none"> • Elpacbareporttagged • elpaccognitiverpt • ELPAC_Stimulus_WordCounts • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 676-2020_v8_FOR ARCHIVE ELPAC CBA Conversion Specifications D-305 • 689-2021D_v2_FOR ARCHIVE_ELPAC 2021_Summative gR3-5 Form 1_082020_FINAL • 690-2020A V2 FOR ARCHIVE ELPAC Educator IWT_Fairness Guidelines and Item Review Checklist • 690-2020A V2 FOR ARCHIVE Revised_2019-20 ELPAC Educator IWT Presentation 090919 • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 • 706-2020A_v3_FOR ARCHIVE ELPAC 2020 New Item Review PPT_123019 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 3.9. Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.</p>	<ul style="list-style-type: none"> • caaccessibilitymtrx2021 • elpaccognitiverpt • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 685-2021 v2 FOR ARCHIVE ELPAC Summative Test Development Specifications_062220 • 735-2021A_v2_FOR REVIEW_2021-Summative-DFA_LSRW 	5
<p>Standard 4.0. Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.</p>	<ul style="list-style-type: none"> • aug20adad02 • B-25 REFERENCE-ELPAC Report on Domain Specific PLDS_v4_APP • elpacbareporttagged • sep17item18 • 456-2019 v5 FOR ARCHIVE ELPAC CBA High-Level Test Design_052119 (002) • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 690-2020A V2 FOR ARCHIVE ELPAC Educator IWT_Fairness Guidelines and Item Review Checklist 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 4.1. Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).</p>	<ul style="list-style-type: none"> • sumelpacfactsheet • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 685-2021 v2 FOR ARCHIVE ELPAC Summative Test Development Specifications_062220 • 1456-2019 v5 FOR ARCHIVE ELPAC CBA High-Level Test Design_052119 	4
<p>Standard 4.6. When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.</p>	<ul style="list-style-type: none"> • may19item01 • sep17item18 • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • Specifications for 2016-17 Content Review and Bias and Sensitivity Review Mtg v6 • ELPAC Specifications for 2017-18 CRP and BSRP Mtg v3 • ELPAC Specifications for 2018-19 CRP and BSRP Meeting v3 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 4.12. Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.</p>	<ul style="list-style-type: none"> • sep17item18 • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 • 706-2020A_v3_FOR ARCHIVE ELPAC 2020 New Item Review PPT_123019 • Specifications for 2016-17 Content Review and Bias and Sensitivity Review Mtg v6 • ELPAC Specifications for 2017-18 CRP and BSRP Mtg v3 • ELPAC Specifications for 2018-19 CRP and BSRP Meeting v3 • ELPAC Grade 3-8 Content and Bias Panel Review v4 • ELPAC Grade 9-12 Content and Bias Panel Review v4 • ELPAC Grade K-2 Content and Bias Panel Review v4 • ELPAC_CandB Panel Review SpreadsheetLRSW_Dec16 (separate file got each grade level) • Reference Document_D-12_ELPAC Grade 3-8 Content and Bias Panel Review 2-26-2018 • Reference Document_D-12_ELPAC Grade 9-12 Content and Bias Panel Review 2-26-2018 • Reference Document_D-12_ELPAC Grade K-2 Content and Bias Panel Review 2-26-2018 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 4.16. The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.</p>	<ul style="list-style-type: none"> • ELPAC--How-to-Use-the-Technology-Readiness-Checker-for-Students • ELPAC--Listening-Speaking-Reading-and-Writing-Training-Test-DFA-Grade-K.2020-21 • ELPAC Moodle • ELPAC--Writing-Training-Test-Grade-K.2020-21 • 303-2020B_v4 FOR ARCHIVE CAASPP ELPAC Post-Test Survey 052920 • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean.090420 	5
<p>Standard 4.23. When a test score is derived from the differential weighting of items or subscores, the test developer should document the rationale and process used to develop, review, and assign item weights. When the item weights are obtained based on empirical data, the sample used for obtaining item weights should be representative of the population for which the test is intended and large enough to provide accurate estimates of optimal weights. When the item weights are obtained based on expert judgment, the qualifications of the judges should be documented.</p>	<ul style="list-style-type: none"> • nov17item08 • nov17item08addendum • C-211 Summative ELPAC Standard Setting Technical Report v6_APP • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 735-2021A_v2_FOR REVIEW_2021-Summative-DFA_LSRW 	5

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<p>Standard 12.4. When a test is used as an indicator of achievement in an instructional domain or with respect to specified content standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes reflected in the target domain should be provided. Both the tested and the target domains should be described in sufficient detail for their relationship to be evaluated. The analyses should make explicit those aspects of the target domain that the test represents, as well as those aspects that the test fails to represent.</p>	<ul style="list-style-type: none"> • Anchor Charts created SRF 2020 • ELPAC CBA - Evidence of Complexity 8-28-2020 • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 • 706-2020A_v3_FOR ARCHIVE ELPAC 2020 New Item Review PPT_123019 • ELPAC Grade 3-8 Content and Bias Panel Review v4 • ELPAC Grade 9-12 Content and Bias Panel Review v4 • ELPAC Grade K-2 Content and Bias Panel Review v4 • ELPAC_CandB Panel Review SpreadsheetLRSW_Dec16 (separate file got each grade level) • Reference Document_D-12_ELPAC Grade 3-8 Content and Bias Panel Review 2-26-2018 • Reference Document_D-12_ELPAC Grade 9-12 Content and Bias Panel Review 2-26-2018 • Reference Document_D-12_ELPAC Grade K-2 Content and Bias Panel Review 2-26-2018 	5

Rationales for Ratings for Testing Standards

This section presents the rationales for HumRRO's ratings in table 2.2 and explains to what extent each relevant testing standard was met based on evidence from the test development documentation. HumRRO also provides suggestions for further strengthening compliance with the testing standards.

Standard 1.9. When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 1.9.

The ELPAC relies on opinions of experts at several stages of its development and operation, including activities such as the standards correspondence review, item review, data review, transition review, range finding, standard setting, and item writing. The vendor has documented, both in technical reports and participant lists, the expert selection process and qualifications, as well as the training raters receive and processes the raters undertake. Because the Speaking portion of the assessment is scored by test administrators, the training of these stakeholders is highly critical to the production of results that can lead to valid score interpretations. The assessment vendor has provided general test administration training online via structured learning modules and, as of the 2020–2021 school year, provides scoring training and calibration online. Ample examples, practice, and necessary resources are afforded to scorers. The qualifications of participants and procedures to make decisions based on human judgment appear appropriate and well-thought out.

1.11. When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 1.11.

The ELPAC is based on the 2012 California English Language Development Standards, which were created by a large number of stakeholders to ensure they are appropriate for kindergarten through grade twelve ELs. ETS assessment specialists assembled the summative ELPAC tests, which were approved by the CDE. The test vendor's technical

report (p. 9) states that "this process began with the creation of test development specifications, which described the content characteristics, psychometric characteristics, and quantity of items to be used in the Summative ELPAC." The blueprint and the item writing guidelines document the relationship between ELPAC task types and the ELD Standards and associated claims, as well as the performance level descriptors that should be targeted by the task type. The item writing guidelines also document possible stem types and what they tend to test, with associated examples. Item writing training provides guidance on how to write tasks to elicit information targeting the different levels of claims. As part of a study examining the transition from paper and pencil to computer-based ELPAC, EL classroom educators indicated that summative ELPAC task types measured EL proficiency consistently with how they taught standards in the classroom.

Standard 1.12. If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 1.12.

The ELPAC attempts to measure the knowledge, skills, and abilities necessary for ELs to succeed in school and subsequent life. Results from a dimensionality analysis, which documents the extent to which an assessment measures multiple dimensions or factors within the larger English proficiency construct, suggest that items do measure different domains. So, at the highest level there is evidence that items measure listening, speaking, reading, and writing processes. The standard setting technical report indicates that panelists discussed how sets of items differed by knowledge and skills. An educator review panel meeting had educators evaluate the degree to which educators observed the same skills elicited across modes. While technical reports do not present the raw results of the standard setting panel or educator review detailing the knowledge and skills differentiating items, they do suggest the reviewers were satisfied with the differentiation of the processes required by the items. Additionally, the speaking and writing rubrics indicate that the differences in scores are related to student language production.

Standard 1.14. When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 1.14.

Like many ELP tests, ELPAC reports several composite scores. Performance levels are reported for each domain (Listening, Speaking, Reading, and Writing), which feed into Oral Language (Listening and Speaking) and Written Language (Reading and Writing)

composite (scale) scores, which then feed into an overall scale score. Technical documents report on the distinctiveness, reliability, and interrelationships among the different scores at each level of the score. The basis for the composite score comes from the dimensionality study, which also provides evidence that the four domains are distinct.

The test vendor reported field test correlations between the different domains, reliability coefficients, and classification consistency and accuracy for each domain for the summative assessment and initial assessment. The vendor also conducted investigations regarding use of a productive and receptive scale score. Weighting of composite scores appropriately varies by grade level with lower grades prioritizing oral language skills compared to the higher grades, where an even written/oral language split is documented. Technical reports are especially clear and comprehensive regarding scores.

Standard 2.3. For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 2.3.

Estimates of reliabilities and standard errors of measurement for total scores and subscores are reported in the testing vendor's technical report. Reliability coefficients for each domain range from good to excellent. The standard error at each score point on the distribution is typically smaller between levels 1 and 2 than between levels 3 and 4, but this appears within reason. Interrater reliability for writing items appears excellent. Very few items were scored discrepantly between two raters.

Standard 2.16. When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 2.16.

The technical reports provide estimates of classification accuracy and classification consistency at each threshold score for the Overall, Written, and Oral classifications. The test vendor uses a proprietary formula to report classification accuracy and consistency for the composite scores. The vendor took a test-retest approach, and made decision classification analyses by grade, domain, and composite.

Standard 3.2. Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 3.2.

ELPAC's test vendor has made efforts to minimize construct-irrelevant variance across the test development cycle. Items are developed according to specifications (e.g., word count, universal design, accessibility), reviewed for bias and sensitivity, analyzed for differential item functioning by group, and tried out using cognitive labs. The vendor developed several protocols for the transition from a paper-based assessment to a computer-based assessment to ensure fairness and construct relevance. The vendor has implemented usability pilot programs to understand how users with limited computer skills and physical disabilities interact with the assessment application.

Standard 3.9. Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 3.9.

ELPAC's transition to a computer-based assessment has allowed the test vendor to integrate universal design for learning principles into the test interface. In addition, various designated supports (e.g., amplification) and accommodations (e.g., scribe) are provided. ELPAC's test vendor conducted usability studies with students with limited computer skills and physical disabilities to determine how they interact with the assessment application and to improve the interface and directions.

Standard 4.0. Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 4.0.

The documents show a close attention to how test specifications should adjust as the test evolves, as demonstrated after conducting the field test and when transitioning the test to computer-based administration. The vendor conducted a variety of small-scale studies to ensure fairness across populations while also building fairness training into their item development processes. Technical reports include sections describing validity evidence and future studies of validity to consider.

Standard 4.1. Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).

Rating: 4, Evidence in the materials mostly covered Standard 4.1.

The technical report states the purpose of the ELPAC and its goals. The ELPAC is a standards-based assessment and the Computer Based Assessment (CBA) Summative Blueprint links task types with ELD Standards and indicates the number of these tasks to be administered by grade/grade span. Raw scores on the four domains are combined into three composite scores (Written, Oral, and Overall), following with current federal testing requirements (ESSA, 2015). The test uses a vertical scale to support the interpretation of growth. The vendor conducted standard setting processes to ensure the level of proficiency necessary for reclassification is appropriate for exiting the EL classification. The testing vendor should include a rationale supporting the uses of test results for the intended purpose(s) of the ELPAC. A high-level test design proposed by ETS during the transition of ELPAC from paper-based to computer-based delivery states that the Summative ELPAC is designed to inform student reclassification as fluent English proficient. It does not, however, specify what unique information test scores contribute to the body of evidence that is used to make reclassification decisions. The system would benefit from a Theory of Action that makes more explicit the logic behind the intended uses of Summative ELPAC scores.

Standard 4.6. When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 4.6.

The ELPAC has had several external reviews related to its test specifications. Prior to operational use, the assessment's Technical Advisory Group, the Regional Assessment Network, and various California educators reviewed the test specifications to ensure they reflect the depth, breadth, and rigor of the 2012 California English Language Development Standards. External reviews of items were conducted by panels and the California Department of Education (CDE) prior to the field test. CDE also reviewed the test form assembly and item statistical properties, conducted user acceptance testing and reviewed directions for administration. The standards were developed to be in conjunction with many external standards, and a content review meeting was held. In general, the qualifications of experts and the process of the reviews are weakly documented, educator qualifications would be reviewed and approved by CDE. Specifications documents detail the process for gathering data from experts and how the data generated during the workshop will be used.

Standard 4.12. Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 4.12.

The blueprint and item guidelines identify which tasks target which of the ELD standards that are intended to be measured by the assessment. In addition to this internal evidence, the technical reports describe a content review panel that reviewed items to ensure they aligned with 2012 ELD Standards. Specifications documents further detail the processes employed during the alignment review. The New Item Review presentation describes that item reviewers rated how well items aligned to standards but reports no evidence on this review; however, the test vendor maintains item ratings and recommendations generated by the review panels. The dimensionality analysis suggests that items do measure the four sub-domains (Listening, Speaking, Reading, and Writing).

Standard 4.16. The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 4.16.

The vendor provides practice tests and practice Directions for Administration (DFAs) at every grade and grade span. The DFAs include rubrics that describe how constructed response items are scored, providing administrators and potentially students with a clear sense of the criteria for scoring. The vendor also provides a tool for users to determine how familiar students are with computers and the functions necessary to conduct online testing. In addition, the technical reports describe usability pilots used to determine how ELs engaged with the computer-based test relative to their non-EL peers and made changes to the process and interface to improve engagement.

Standard 4.23. When a test score is derived from the differential weighting of items or subscores, the test developer should document the rationale and process used to develop, review, and assign item weights. When the item weights are obtained based on empirical data, the sample used for obtaining item weights should be representative of the population for which the test is intended and large enough to provide accurate estimates of optimal weights. When the item weights are obtained based on expert judgment, the qualifications of the judges should be documented.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 4.23.

In the Standard Setting document, two different weighting schemes are described: one where kindergarten and first grade are differentially weighted by subscore (90/10 in K, 70/30 in first), one where only Kindergarten is differentially weighted (70/30). Eventually the K-only option was selected. The rationale for weighting the Kindergarten overall score is provided (i.e., Kindergarten students' written literacy skills are less developed and oral skills should be weighted more). Additional reporting to the SBE provides more detail on the options considered leading to the selected weighting scheme. Data to

determine the composite weighting comes from the 2017 field test, which was intended to be representative of the target population. Qualifications of experts involved in the weighting decision process is documented in the standard setting document.

Standard 12.4. When a test is used as an indicator of achievement in an instructional domain or with respect to specified content standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes reflected in the target domain should be provided. Both the tested and the target domains should be described in sufficient detail for their relationship to be evaluated. The analyses should make explicit those aspects of the target domain that the test represents, as well as those aspects that the test fails to represent.

Rating: 5, Evidence in the materials fully covered all aspects of Standard 12.4.

The ELPAC is used as an indicator of progress toward achievement on ELD Standards and has an item blueprint which aligns items to standards and Proficiency Level Descriptors (PLDs) with each task type. Students require language skills in the context of grade-level academic subject area courses. The ELD Standards incorporate an awareness of the cognitive development of students at various ages to make them appropriate at different grade levels. They note the purpose of language for different ELD proficiency levels (e.g., analyzing, evaluating). As students gain more English proficiency, they can use English to demonstrate more complex cognitive processes. Test specifications do not delineate the processes to be accessed, but these processes are appropriately presented and noted in the standards and rubric to the extent one would expect for an ELP assessment with this design.

The test vendor conducted a "content review meeting" and assisted with a "correspondence study" with California educators as raters. The "content review meeting" considered whether each item would appropriately measure the aligned standards (as designated by item writers) and was grade level appropriate. Based on several ratings, items were classified as "approve the item as is," "approve the item with revisions," and "reject." CDE reviewed these ratings and made final decisions about how to treat each item. The test vendor maintains item ratings and recommendations generated by the review panels. The correspondence study (from WestEd) examined the linkage between items and content area skills.

Summary and Discussion

Fourteen of the identified standards were rated as fully covered based on the available evidence and one identified standard was rated as mostly covered. These results indicate that the ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to English language development standards. Chapter 3 of this report describes the alignment workshop convened to document the extent to which test forms are adequately aligned to the ELD standards.

Chapter 3: Summative ELPAC Alignment Workshop and Outcomes

Introduction

Human Resources Research Organization (HumRRO) conducted an alignment study of the English Language Proficiency Assessments for California (ELPAC). This study provides evidence of whether the ELPAC system produces test forms that effectively measure the intended construct, as described in the English Language Development (ELD) standards. It does so by evaluating how well the test items fully sample the construct represented by the ELD Standards. This report describes the methods used and summarizes the study results, including the evaluation of four major alignment criteria.

Summative ELPAC Alignment Criteria

Alignment studies can provide evidence to support the claim that assessments measure the content they are intended to measure. In this case, the content, or the measurement construct, is described for the ELPAC by the ELD Standards. The alignment workshop is designed to evaluate how well the test items represent (align with) the ELD Standards. For ELPAC, four main criteria were evaluated. Table 3.1 provides a brief description of the criteria addressed in this study.

Table 3.1 ELPAC to ELD Standards Alignment Criteria

Criteria	Description
Link to Standards	<ul style="list-style-type: none">At least 90% of items within each domain are matched to a primary standard.For Speaking and Writing domains only, at least 90% of items are also matched to a secondary standard.
Link to ELD Proficiency Levels	At least 25% and no more than 50% of items across the assessment are rated at each of the three ELD proficiency levels (Emerging, Expanding, and Bridging).
Range Adequacy	<ul style="list-style-type: none">Across the four domains, each of the three modes of communication (collaborative, interpretive, and productive) is measured by at least one item.Across the four domains, each of the three ways of using language (structuring cohesive texts, expanding and enriching ideas, and connecting and condensing ideas) is measured by at least one item.At least 50% of the ELD Standards are represented by at least one item.
Balance-of-Knowledge Correspondence	Webb's balance-of-knowledge correspondence criteria is used, computed for modes of communication and ways of using language, across domains. Both must meet Webb's threshold of 0.70.*

* The index ranges from 0–1, with 1 representing perfect balance (the same number of items per content objective/standard within a topic). The more unevenly the items are distributed, the lower the index.

Link to Standards

All test items should reflect content described by the standards. An item that does not directly relate to the standards would be considered “construct irrelevant.” Construct irrelevant items measure something other than what the test is intended to measure and can potentially introduce error into the ability estimate. For that reason, this criterion requires that nearly all items on the tests measure content as described in the standards. The criterion is less than 100 percent only to prevent a single poorly written item from triggering an inappropriate alignment review. However, any item that is not directly linked to content from the standards should be revised to address the intended content or eliminated from the assessment. Speaking items and Writing items are constructed-response items developed to measure a primary and a secondary standard, as outlined in the ELPAC Blueprint. Secondary standards are implicitly accounted for in the constructed-response rubrics.

Webb’s *Categorical Concurrence* criterion is a basic measure of alignment between content standards and test items, and it is the most similar to our Link to Standards criterion. The term *categorical concurrence* refers to the proportion of overlap between the content stated in the standards document and that assessed by items on the test. Webb’s criterion is based on the minimum number of items required to achieve acceptable reliability for reporting. We prefer to directly examine the reliability of the ELPAC, which is available in the ELPAC technical report for the 2018–2019 administration (CDE, 2020). Reliability of scores should be evaluated for overall ELPAC scores and subscores.

Webb’s *Categorical Concurrence* criterion is derived by determining if there are at least six items per reporting category on the assessment. Six was determined to be the minimum number of items required to establish a reasonable reliability (Subkoviak, 1988). California is currently reporting an overall score and composite scores for the oral (Listening and Speaking) and written (Reading and Writing) language skills. So, at the most basic level, California could meet Webb’s criteria if at least three items per domain were included on the assessments. This would be a weak criterion for determining the sufficiency of items for generating reliable student scores.

In addition to the criterion, we will also report the full item-level data to CDE in a separate electronic file. This file includes a side-by-side comparison of the panelists’ final consensus data with the testing contractor’s metadata. Items that are not linked to a standard will be flagged for additional scrutiny.

Link to ELD Standards Proficiency Levels

English language proficiency assessment items are often categorized by linguistic difficulty level (Cook, 2005) in alignment studies. The standards can also be categorized by the linguistic difficulty level required to demonstrate the content described in each standard content statement. Then, it is possible to match the two to create an analog to Webb’s *Depth-of-Knowledge (DOK) Consistency* statistic. Webb’s DOK Consistency criterion measures the type of cognitive processing required by items compared to the cognitive processing required by the matched content standards. For example, is a

student expected to simply identify or recall basic facts, to use reason to manipulate information, or to strategize how to best solve a complex problem? In English language proficiency assessments, items and standards can be categorized by the difficulty of the language students are expected to interpret, interact with, or produce and can be matched in the same way.

The test blueprints for the Summative ELPAC do not require a particular distribution of linguistic difficulty levels for the items making up the assessment, nor does the Summative ELPAC testing contractor include a linguistic difficulty level indicator in the item metadata. Instead, California built linguistic difficulty directly into the ELD Standards proficiency level descriptors. This allows California to more rigorously address linguistic difficulty in terms of both item difficulty and students' scores, by categorizing items in a way that is connected to performance levels and student scores. The Summative ELPAC performance level descriptors are used to set standards (determine threshold scores to assign students to performance categories), and the ELD proficiency level descriptors reference linguistic difficulty. So, students in a particular performance category could be expected to have facility with language at the difficulty level described by the corresponding ELD proficiency level. This allows California's educators to reference the standards and ELD proficiency levels to students' Summative ELPAC performance categories, and better understand the linguistic difficulty of language with which the students have demonstrated proficiency, rather than rely on external linguistic difficulty definitions that may not directly relate to student scores.

While the advantages to California's approach to linguistic difficulty are clear from a test design standpoint, it changes the nature of how we approach alignment to standards. We considered introducing a linguistic difficulty scale into the alignment study so we could match items to standards in the same manner that DOK is typically used, but the scale did not correspond to the ELD proficiency levels and created a potential translation challenge for interpreting the alignment results. The ELD Standards were not written referencing a particular linguistic difficulty scale, and most of the standards address a potentially wide range of linguistic difficulty. Assigning a single linguistic difficulty level to each standard would not have been informative, and therefore not useful for the purposes of matching item-level linguistic difficulty. We chose instead to link the items directly to the ELD proficiency levels. The ELD proficiency level descriptors incorporate linguistic difficulty and matching items to them make the alignment results more meaningful and useful.

California uses three ELD proficiency levels in the ELD Standards; Emerging, Expanding, and Bridging. Each of those levels is further divided into early and exiting, resulting in six proficiency levels total. These proficiency levels relate to the four ELPAC reporting levels applied to the overall score, the oral language score, and the written language score, as depicted in figure 3.1. The ELD proficiency levels are on the top row of the figure, the second row shows the reporting levels for the ELPAC, and the final row provides a brief policy definition of what each reporting level means.

ELD – EMERGING	ELD – EXPANDING		ELD – BRIDGING
Summative ELPAC Level 1	Summative ELPAC Level 2	Summative ELPAC Level 3	Summative ELPAC Level 4
Minimally developed English skills	Somewhat developed English skills	Moderately developed English skills	Well developed English skills

Figure 3.1 ELD Proficiency Levels and Reporting Levels for ELPAC

Note that the ELD proficiency levels and the Summative ELPAC reporting levels do not correspond one-to-one. Level 1 equates to ELD—Emerging, ELD—Expanding encompasses reporting Level 2 and part of Level 3. ELD—Bridging includes part of reporting Level 3 as well as Level 4. Because the reporting categories for overall, oral, and written scores fall across the range of ELD proficiency levels, we reasoned that items should substantially address all ELD proficiency levels. The criterion for Link to ELD Proficiency Levels was set such that each category should include at least 25 percent of the items, and no category should include more than 50 percent of the items. This criterion ensures that all ELDs are represented, supporting the reporting described in figure 3.1.

In addition to evaluating the item-level ELD proficiency levels for meeting this criterion, we will provide panelists’ final ELD proficiency level ratings for each item to CDE in the electronic file. We will also provide descriptive statistics (number and proportion of items at each level).

Range Adequacy

Webb’s *Range-of-Knowledge Correspondence* criterion examines the extent to which the test items reflect the full range of knowledge, skills, and abilities contained in the standards document. Where *Categorical Concurrence* notes whether a sufficient number of items on the test covers each general content topic (reporting category), the *Range-of-Knowledge Correspondence* measure indicates the number of specific content objectives within each broader topic that are assessed by the test items.

Webb’s *Range-of-Knowledge Correspondence* criterion requires that at least 50 percent of the standards from each reporting category are addressed on the assessment. For the ELPAC, we use a Range Adequacy criterion. It incorporates Webb’s requirement that at least 50 percent of the ELD standards are represented by test items but adds a requirement that the ELPAC also reflect the modes of communication and ways of using language. Modes of communication include collaborative, interpretive, and productive modes. Ways of using language include structuring cohesive texts, expanding and enriching ideas, and connecting and condensing ideas. To fully meet the Range Adequacy criterion the ELPAC test forms must (a) include items representing at least 50 percent of the ELD Standards, (b) include at least one item representing each mode of communication, and (c) include at least one item representing each identified way of using language. Items can meet these requirements based on either the primary identified standard or via a secondary identified standard. The standards documents are arranged by mode of communication (Part I: Interacting in Meaningful Ways) and way of

using language (Part II: Learning About How English Works). Identifying the standard to which an item is aligned categorizes the item by mode of communication or way of using language. No additional rating task was required for panelists to evaluate this criterion.

In addition to the criterion, we will provide descriptive statistics (number and proportion of items) matched to each mode of communication and way of using language.

Balance of Knowledge Correspondence

Webb's *Balance of Knowledge Representation* focuses on content coverage in yet more detail. In this case, the number of items matched to the content objective does matter. The *Balance of Knowledge Representation* criterion determines whether the assessment measures the content objectives equitably within each content topic using only those content objectives identified by panelists as measured by the test items. Based on Webb's (1997) method, items should be distributed evenly across the objectives per content topic for good balance. The *Balance of Knowledge Representation* is determined by calculating an index, or score, for each content topic. Each topic should meet or surpass a minimum index level to demonstrate adequate balance. Webb's index ranges from 0–1, with 1 representing perfect balance (the same number of items per content objective/standard within a topic). The more unevenly the items are distributed, the lower the index.

For ELPAC, we computed Webb's *Balance of Knowledge Representation* index for mode of communication and for ways of using language. To meet the Balance of Knowledge Correspondence criterion, the ELPAC test forms were required to have balance indexes of greater than 0.70 for both. This criterion extends Webb's work to address the major topics addressed by an English language proficiency assessment.

Methods

The evaluation of the alignment criteria is based on item ratings and professional judgments collected during an alignment workshop. This section describes the workshop participants (henceforth referred to as "alignment panelists" or "panelists"), workshop materials, training, and workshop processes and procedures.

Alignment Panelists

HumRRO worked collaboratively with the CDE to recruit and select a group of 42 educators to serve on seven ELPAC alignment review panels (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, grades 9–10, and grades 11–12). Recruitment for the virtual workshop during the unprecedented circumstances of the global pandemic provided unique challenges. A specific challenge to recruitment was potential panelist lack of comfort with or access to the technology necessary to participate effectively. There were 10 last-minute cancellations. Of these, 5 were due to the unpredictable return to in-person instruction or health issues related to vaccinations. Due to these last-minute cancellations and a limited pool of alternates, 36 panelists participated in the workshop. Instead of the planned 6 members per panel, only two had 6 members; four

panels (kindergarten, grades 3–5, grades 9–10, and grades 11–12) each had five participants; and one panel (grades 6–8) had only four. Across the seven panels, 24 California school districts were represented.

Approximately 47 percent of panelists reported currently working in the role of teacher (including lead teacher), and the remaining 53 percent reported working in roles such as coordinator, specialist, instructional coach, program director, assistant principal, or superintendent. In addition to their current professional roles, all panelists reported having some level of experience with the ELD Standards. The types of experience reported ranged from teaching the ELD Standards to providing CA EL-related training to other educators to proctoring the Summative ELPAC. Across the seven panel groups, 89 percent of panelists reported experience teaching students from diverse socioeconomic and cultural backgrounds. Additionally, 75 percent of the panelists reported experience working with students with mild-to-moderate and/or significant disabilities.

Geographically, 28 percent of panelists were from Northern or Central California and 72 percent were from Southern California. Based on self-reports of race and ethnicity, 58 percent of panelists were White, 11 percent Asian, 11 percent multi-race, 6 percent Mexican or Mexican American, 6 percent Black or African American, 3 percent Central American, and 3 percent American Indian or Alaskan Native. Table 3.2 summarizes key demographics of each alignment panel.

Table 3.2 Summative ELPAC Alignment Panelists’ Demographics

Panel	# of Panelists	# of Districts	% Female	% Hispanic	Years of Experience Mean (SD)
Kindergarten	5	4	100%	40%	29 (8)
Grade 1	6	6	83%	50%	17 (4)
Grade 2	6	4	100%	50%	19 (8)
Grades 3–5	5	5	100%	40%	19 (11)
Grades 6–8	4	3	100%	25%	15 (14)
Grades 9–10	5	5	80%	40%	18 (7)
Grades 11–12	5	5	80%	20%	17 (4)

Workshop Logistics

HumRRO conducted a four-day ELPAC Alignment Study Workshop virtually via Microsoft TEAMS on March 22–25, 2021. Panelists received five calendar invitations which linked them to one whole group training session and four grade level panel meetings where further training and calibration, and item rating occurred. During the workshop, panels of educators evaluated how well each ELPAC item assessed the ELD standard and how ELPAC Level 4 requirements correspond to specific academic

expectations related to language use. Prior to entering the workshop, panelists were required to sign nondisclosure agreements as a condition of participation.

Workshop Materials

CDE and ETS provided HumRRO with documents and data to facilitate the development of materials for the alignment workshop. These included test design documentation and item metadata. ETS created seven online test forms for the alignment workshop (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, grades 9–10, and grades 11–12) consisting of all the operational 2021 Summative ELPAC items. ETS also created accounts for HumRRO researchers and participants to securely access the items using the IBIS™ Content Review Tool (CRT).

HumRRO developed several data collection tools and adapted other materials to support the data collection process. Data collection tools included electronic spreadsheets for panelists and workshop facilitators to enter test item ratings. Support materials included both paper and electronic copies of the (a) ELD Standards, (b) ELD Proficiency Level Descriptors (PLDs), and (c) detailed workshop outline and instructions for both panelists and facilitators. For security purposes, item scoring information for all grade levels and Directions for Administration (DFA) for Speaking (grades K–12) and Writing (grades K–2) were shared via a secure online portal only. Debriefing and evaluation surveys were administered online and completed at the end of the workshop. Example workshop materials are presented in Appendix B.

Training

At the outset of the virtual alignment workshop, we spent several minutes ensuring panelists had adequate audio and video access, had all materials on hand, and were comfortable using the various functionalities of Microsoft Teams (e.g., the chat window, raising a "virtual" hand). Alignment panelists then received two rounds of training. First, the full group of panelists received general training that provided some background on alignment and a high-level description of the alignment process. Following the general training session, panelists moved into grade-level/span panel groups and received more detailed training on the data collection processes and procedures. Those processes and procedures are described in more detail in the following section.

Workshop Processes and Procedures

Prior to the workshop, panelists verified they had access to a desktop or laptop computer with a microphone and camera installed, a quiet secure place to work, and availability for the full duration of the alignment study. Non-secure printed materials had been mailed to the panelists in advance of the workshop to limit the number of electronic files that were required to be open on their computers simultaneously. These included copies of the ELD Standards, the ELD PLDs, an annotated sample rating form with instructions, and the agenda. Operational test items were accessed via an online secure platform set up by ETS. Electronic rating forms were provided to panelists via an emailed link or by placing the link in their chat windows during the workshop. HumRRO

provided panelists access to electronic versions of secure documents (e.g., grades K through 2 Writing PDFs, DFAs) via access-restricted Google Drive folders.

After the panel-specific training presentation by the HumRRO facilitator, each panel engaged in a calibration activity using the first 1–3 items. Panelists accessed the items electronically and made their independent ratings. Rating forms were designed to allow only prescribed rating options. All ratings were automatically saved and used to populate a data monitoring sheet used by the panel facilitator to check for completion of ratings and to facilitate discussion among the panelists. Panelists discussed their independent ratings and engaged in consensus discussion to come to agreement on the final item ratings of record. Once panelists had a clear understanding of the rating process and a common understanding of the rating categories applied, they moved on to rating the remaining operational items.

The panelists rated a small group of operational items at a time. Items were presented by domain, beginning with Listening, then followed by Speaking, Reading, and finally Writing. Items were rated in small groups to facilitate discussion and consensus building and to keep panelists on roughly the same schedule. For each group of items, panelists first made their ratings for each item independently. Then panelists discussed their ratings for an item, reached initial consensus, examined the item metadata, engaged in more discussion, then reached their final consensus/majority rating for that item before moving on to the next.² Once consensus/majority ratings were recorded for that group of items, the panel moved on to the next group and repeated this process. Item ratings were generated via the following steps:

1. Panelists reviewed test items and relevant scoring information independently and assigned ratings of:
 - a. Primary ELD standard measured by item
 - b. Secondary ELD standard measured by item (Speaking and Writing items only)
 - c. ELD proficiency level (using a six-point scale) for each item score point (multi-point items can assess differing levels of ELD proficiency). The rating scale was based on the ELD proficiency level continuum defined in the ELD Standards: Early Emerging (1), Exiting Emerging (2), Early Expanding (3), Exiting Expanding (4), Early Bridging (5), and Exiting Bridging (6).
2. Panelists discussed their independent ratings and come to initial consensus
3. Panelists came to consensus (or majority) ratings
4. HumRRO facilitator recorded consensus/majority ratings
5. HumRRO facilitator shared item metadata

²When consensus could not be reached, we recorded the majority rating.

During the workshop, we modified step 5 in the rating process. Our original process entailed showing item metadata prior to consensus ratings. However, we observed during the workshop that this impeded the progress of several panels' discussions. We ultimately revised the process, with CDE approval, to show item metadata after consensus/majority ratings were reached.

Once all panelists had completed their independent ratings, the HumRRO facilitator managed the group discussion and encouraged all panelists to share their ratings. Typically, the facilitator polled the group about each rating, and asked for panelists to provide a rationale when independent ratings differed among them. Panelists were trained to retain their independent ratings unless they realized they had made a coding error, or if group discussion revealed to them an error in their thinking about an item and/or the ELD Standards. The facilitator then polled the group to determine consensus on the ratings that had been discussed. If the group could not reach true consensus, the facilitator recorded the rating of the majority of panelists.

The HumRRO facilitator recorded the final consensus (or majority) item ratings in a spreadsheet. Once all consensus ratings were recorded, panelists completed three online Microsoft Forms surveys: a debriefing form, a process evaluation survey, and a demographic questionnaire (See Appendix B for survey questions). The debriefing form was designed to give panelists the opportunity to describe their overall view of the quality of alignment. The anonymous evaluation survey (optional) elicited feedback about the quality of the workshop, including panel facilitation, materials, and processes. Panelists were then released from the workshop. Table 3.3 summarizes the process evaluation survey results.

Table 3.3 Summative ELPAC Alignment Evaluation Survey Results

Evaluative Statement	% Strongly Disagree	% Disagree	% Somewhat Disagree	% Somewhat Agree	% Agree	% Strongly Agree
The training presentation in the large group provided useful information about the Summative ELPAC and HumRRO's alignment method.	6.5	3.2	0.0	6.5	41.9	41.9
After the additional training in my small group, I felt prepared to review and rate test items.	3.2	0.0	0.0	0.0	51.6	45.2
HumRRO staff seemed knowledgeable of the Summative ELPAC and alignment steps.	3.2	0.0	0.0	9.7	35.5	51.6
The Panelist Instruction document was clear, understandable, and useful in performing the alignment steps.	3.2	0.0	0.0	6.5	32.3	58.1
The Google Sheets file was understandable and relatively easy to use to enter item ratings.	3.2	0.0	0.0	6.5	32.3	58.1
The process for reaching consensus ratings was conducted fairly.	3.2	0.0	0.0	0.0	35.5	61.3

Results

This section summarizes the data/information collected during the Summative ELPAC alignment workshop.

Criterion 1: Link to Standards

At the most basic level, an assessment must address its intended measurement construct. In simple terms, when we establish an assessment's link to standards, we are responding to the question, "Does the test measure what it's supposed to measure?" In the case of the Summative ELPAC, we ask if the test items relate directly to the content standards on which the test is based. To meet this criterion, at least 90 percent of items within each domain must be matched to a primary standard and, for Speaking and Writing domains only, 90 percent of items must also be matched to a secondary standard. Table 3.4 presents the findings for all grade levels/spans for the Link to Standards criterion for the Summative ELPAC.

Table 3.4 Results for Criterion 1: Link to Standards

Grade Level/Span	Items Matched to Primary Standard	Speaking and Writing Items Matched to Secondary Standard	Acceptable?
K	51/51 (100%)	9/9* (100%)	Yes
1	59/59 (100%)	7/13* (54%)	No
2	65/65 (100%)	16/16* (100%)	Yes
3–5	66/66 (100%)	18/18 (100%)	Yes
6–8	66/66 (100%)	18/18 (100%)	Yes
9–10	66/66 (100%)	18/18 (100%)	Yes
11–12	66/66 (100%)	18/18 (100%)	Yes

*Two task types for Writing in grades K, 1, and 2 were not intended to have a Secondary Standard, per the ELPAC Summative Blueprint.

All Summative ELPAC items across all grades were aligned to a primary standard, meeting the criterion overall and by domain. Except for grade one, all Speaking and Writing ELPAC items that were intended to be aligned to a secondary standard were matched to a secondary standard. Only 7 of 13 (54%) Speaking and Writing items on the grade one Summative ELPAC were matched to a secondary standard, 6 of 9 for Speaking (67%) and 1 of 4 for Writing (25%), causing the grade one assessment not to meet the criterion.

Criterion 2: Link to ELD Proficiency Levels

This criterion is based on panelists' ratings of the knowledge, skills, and abilities students must access to correctly respond to an item. The ELD proficiency levels describe the knowledge, skills and abilities of students at each of three proficiency levels (Emerging, Expanding, and Bridging). The table of ELD proficiency level descriptors further splits each level into two subcategories (Entering and Exiting). Panelists used a rating scale with six levels and selected the ELD proficiency level for each item (or each item score point for multi-point items) at the subcategory level (i.e., Early or Exiting), but HumRRO aggregated results for evaluation of the criterion to the category levels (i.e., Emerging, Expanding, Bridging). To be considered acceptable, each Summative ELPAC test form must include at least 25 percent and no more than 50 percent of items across the assessment rated at each of the three ELD proficiency levels. This criterion is similar to Webb's DOK consistency criterion, but more appropriate for an English language proficiency assessment because it references the specific knowledge, skills, and abilities associated with language development, rather than more generic cognitive processing depth.

Other English language alignment studies have used linguistic difficulty level (LDL) as an analog to Webb's DOK. It is possible to assign a linguistic difficulty rating to both English language development standards and test items. Then the ratings can be matched and a consistency statistic can be computed in the same way as Webb's DOK (Cook, 2005). California did not reference a particular linguistic difficulty scale when writing the ELD Standards, and many are written to address a broad range of linguistic difficulty. The standards are intended to be used in conjunction with the ELD proficiency levels, since students could demonstrate facility with the content described in the standard at multiple levels. Linking test items to the ELD proficiency levels provides California with more and better information than a more traditional approach.

Tables 3.5 through 3.11 present the findings by grade/span for the Link to ELD Proficiency Levels criterion for the Summative ELPAC. Each table presents the number of items rated at each ELD proficiency level for each possible score point and the aggregated ratings by category (Emerging, Expanding, and Bridging). For example, in table 3.5, the 1 Point column indicates a score of one point on 8 items (all one-point items) was rated at Level 1; a score of one point on 19 items (including both one- and two-point items) was rated at Level 2, a score of one point on 9 items was rated at Level 3, a score of one point on 12 items was rated at Level 4, a score of one point on 3 items was rated at Level 5, and a score of one point on 0 items was rated at Level 6. The Total column indicates the total number of kindergarten items (all score points) rated at each ELD proficiency level and the combined total for each ELD category. The Combined Total (%) by Category column indicates 46 percent of all kindergarten items (all score points) were rated Emerging. The "Yes" in the final column indicates the Link to ELD Proficiency Levels criterion was met for the Emerging level (not less than 25% or more than 50%).

Table 3.5 Link to ELD Proficiency Levels Kindergarten

ELD Proficiency Rating	1 Point	2- Points	3- Points	4- Points	Total	Combined Total (%) by Category	Category Criterion Met?
Emerging	27	6	0	0	33	46%	Yes
Level 1- Early	8	0	0	0	8	n/a	n/a
Level 2- Exiting	19	6	0	0	25	n/a	n/a
Expanding	21	8	3	3	35	49%	Yes
Level 3- Early	9	8	0	0	17	n/a	n/a
Level 4- Exiting	12	0	3	3	18	n/a	n/a
Bridging	3	0	0	0	3	4%*	No*
Level 5- Early	3	0	0	0	3	n/a	n/a
Level 6- Exiting	0	0	n/a	n/a	n/a	n/a	n/a

Note: * denotes that criterion was not met.

Rounding leads to a Combined Total % by Category of 99%.

Table 3.6 Link to ELD Proficiency Levels Grade One

ELD Proficiency Rating	1 Point	2- Points	3- Points	4- Points	Total	Combined % by Category	Category Criterion Met?
Emerging	30	9	0	0	39	48%	Yes
Level 1- Early	13	0	0	0	13	n/a	n/a
Level 2- Exiting	17	9	0	0	26	n/a	n/a
Expanding	24	4	7	3	38	46%	Yes
Level 3- Early	11	4	4	1	20	n/a	n/a
Level 4- Exiting	13	0	3	2	18	n/a	n/a
Bridging	5	0	0	0	5	6%*	No*
Level 5- Early	5	0	0	0	5	n/a	n/a
Level 6- Exiting	0	0	0	0	0	n/a	n/a

Note: * denotes that criterion was not met.

Table 3.7 Link to ELD Proficiency Levels Grade Two

ELD Proficiency Rating	1 Point	2-Points	3-Points	4-Points	Total	Combined % by Category	Category Criterion Met?
Emerging	19	2	0	0	21	23%*	No*
Level 1- Early	2	0	0	0	2	n/a	n/a
Level 2- Exiting	17	2	0	0	19	n/a	n/a
Expanding					57	61%*	No*
Level 3- Early	14	12	0	0	26	n/a	n/a
Level 4- Exiting	22	2	7	0	31	n/a	n/a
Bridging	11	0	0	4	15	16%*	No*
Level 5- Early	10	0	0	4	14	n/a	n/a
Level 6- Exiting	1	0	0	0	1	n/a	n/a

Note: * denotes that criterion was not met.

Table 3.8 Link to ELD Proficiency Levels Grades Three through Five

ELD Proficiency Rating	1 Point	2-Points	3-Points	4-Points	Total	Combined % by Category	Category Criterion Met?
Emerging	20	2	0	0	22	23%*	No*
Level 1- Early	6	0	0	0	6	n/a	n/a
Level 2- Exiting	14	2	0	0	16	n/a	n/a
Expanding	37	14	4	1	56	59%*	No*
Level 3- Early	13	11	1	0	25	n/a	n/a
Level 4- Exiting	24	3	3	1	31	n/a	n/a
Bridging	9	0	4	4	17	18%*	No*
Level 5- Early	9	0	4	1	14	n/a	n/a
Level 6- Exiting	0	0	0	3	3	n/a	n/a

Note: * denotes that criterion was not met.

Table 3.9 Link to ELD Proficiency Levels Grades Six through Eight

ELD Proficiency Rating	1 Point	2- Points	3- Points	4- Points	Total	Combined % by Category	Category Criterion Met?
Emerging	19	2	0	0	21	22%*	No*
Level 1- Early	6	0	0	0	6	n/a	n/a
Level 2- Exiting	13	2	0	0	15	n/a	n/a
Expanding	43	14	7	0	64	67%*	No*
Level 3- Early	16	12	0	0	28	n/a	n/a
Level 4- Exiting	27	2	7	0	36	n/a	n/a
Bridging	4	0	2	4	10	11%*	No*
Level 5- Early	4	0	2	3	9	n/a	n/a
Level 6- Exiting	0	0	0	1	1	n/a	n/a

Note: * denotes that criterion was not met.

Table 3.10 Link to ELD Proficiency Levels Grades Nine and Ten

ELD Proficiency Rating	1 Point	2- Points	3- Points	4- Points	Total	Combined % by Category	Category Criterion Met?
Emerging	11	0	0	0	11	12%*	No*
Level 1- Early	0	0	0	0	0	n/a	n/a
Level 2- Exiting	11	0	0	0	11	n/a	n/a
Expanding	54	16	3	0	73	77%*	No*
Level 3- Early	34	6	0	0	40	n/a	n/a
Level 4- Exiting	20	10	3	0	33	n/a	n/a
Bridging	1	0	6	4	11	12%*	No*
Level 5- Early	1	0	6	1	8	n/a	n/a
Level 6- Exiting	0	0	0	3	3	n/a	n/a

Note: * denotes that criterion was not met.

Table 3.11 Link to ELD Proficiency Levels Grades Eleven and Twelve

ELD Proficiency Rating	1 Point	2- Points	3- Points	4- Points	Total	Combined % by Category	Category Criterion Met?
Emerging	15	1	0	0	16	17% *	No*
Level 1- Early	1	0	0	0	1	n/a	n/a
Level 2- Exiting	14	1	0	0	15	n/a	n/a
Expanding	46	15	6	0	67	71%*	No*
Level 3- Early	17	9	1	0	27	n/a	n/a
Level 4- Exiting	29	6	5	0	40	n/a	n/a
Bridging	5	0	3	4	12	13%*	No*
Level 5- Early	5	0	3	1	9	n/a	n/a
Level 6- Exiting	0	0	0	3	3	n/a	n/a

Note: * denotes that criterion was not met.

The current Summative ELPAC items tend to be concentrated in the middle range of the ELD proficiency levels. Most items were rated at the Expanding level. Grades kindergarten and one had acceptable proportions of items in the Emerging and Expanding levels, but both levels were approaching 50 percent. All other grade/grade-span assessments had more than 50 percent Expanding level items, and fewer than 25 percent of items in either the Emerging or Bridging levels.

Analysis of Item-Person Maps

To further investigate the Link to ELD Proficiency level results, HumRRO requested and obtained item-person (Wright) maps from ETS to demonstrate where items and students perform on the Summative ELPAC on a common scale.

Item-person maps, or Wright maps, illustrate the correspondence between test takers' ability and the difficulty of the test items. Ideally, test items will be at an appropriate level of difficulty to measure the test takers' ability level, ensuring that the test provides information about test performance that is meaningful and useful. For example, test scores on a test in which most items are too difficult for most test takers could result in an underestimation of true achievement levels. Test score information depends on item information from items with difficulties near the ability level of the students taking the tests. Item-person maps produced by ETS for each grade level, aggregated by Oral and Written test sections, are presented in Appendix C.

We compared panelists ELD proficiency level ratings for items to those items' difficulty parameters to help add context to the outcomes of our analysis of item classifications by ELD proficiency level.

The item-person map for grade 2 oral language (weighted combination of Speaking and Listening domains) is replicated in table 3.12 for illustrative purposes. The number of students scoring at each level is presented on the left side of the figure. The number of items at each scoring level is presented on the right side of the figure. There are horizontal lines representing each threshold score, the score where a student would change categories (i.e., 1 to 2, 2 to 3, and 3 to 4). Therefore, easier items are located toward the bottom of the figure, while more difficult items are located toward the top.

The Summative ELPAC was developed such that students scoring at Level 4 in oral language are judged Proficient in the assessed English language domains. When we examine the distribution of students, we see a mostly normal curve (depicted vertically on the left side of the figure below). The curve's highest point, representing the most students, is in Level 3, but there are a substantial number of students in Levels 2 and 4 as well. There are fewer students in Level 1.

If we examine the item locations, we see that most item difficulties are in Levels 1 and 2. There are very few items in Levels 3 or 4. The Os on the figure represent dichotomous items. There is only one dichotomous item in Level 3 and one in Level 4. The Ps on the figure represent polytomous items, with a P indicated on the figure for each scoring level included for an item. For example, an item with 4 possible score points would occur on the figure 4 times, once for each possible score point. We see that most of the items in Levels 3 and 4 are the higher score points from polytomous items.

The other item-person maps for other grade levels and for written language follow the same general pattern. Items difficulties tend to be clustered toward the bottom of the scale and toward the lower scoring levels. This corresponds to the panelists' results for item-level ELD Proficiency ratings. Panelists rated most of the items in the lower part of the ELD Proficiency scale, and very few items in the upper levels. Note that the ELD Proficiency scale ranges from Level 1 (Early Emerging) to 6 (Exiting Bridging), while the performance categories range from 1 to 4, so we cannot directly compare panelists' ratings with the categories depicted on the item-person map, but we can see that the general patterns are very similar. Both data sources indicate an abundance of lower-difficulty items and very few higher difficulty items.

Table 3.12 Grade Two Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
223	.XX	4.0	O	1	
0		3.8	-	0	
0		3.6	-	0	
25	.	3.4	-	0	
0		3.2	-	0	
21	.	3.0	-	0	
33	.	2.8	-	0	
41	.	2.6	-	0	
36	.	2.4	-	0	
123	.X	2.2	-	0	
74	.	2.0	-	0	
117	.X	1.8	-	0	
410	.XXXX	1.6	-	0	
288	.XX	1.4	-	0	
470	.XXXX	1.2	-	0	
613	.XXXXXX	1.0	P	1	
942	.XXXXXXXXXX	0.8	P	1	
1171	.XXXXXXXXXXXX	0.6	-	0	
1478	.XXXXXXXXXXXXXXXX	0.4	-	0	
1700	XXXXXXXXXXXXXXXXXXXX	0.2	-	0	Level 4 (0.03)
2044	.XXXXXXXXXXXXXXXXXXXXXXXX	0.0	-	0	
2338	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.2	P	1	
2452	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.4	PP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table 3.12 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
2572	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.6	O	1	
2594	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.8	-	0	
2327	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.0	-	0	Level 3 (-1.18)
2082	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.2	-	0	
1789	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.4	-	0	
1442	.XXXXXXXXXXXXXXXXXXXX	-1.6	OOPP	4	
1123	.XXXXXXXXXXXX	-1.8	OOPPPP	6	
819	.XXXXXXXX	-2.0	-P	1	
571	.XXXXX	-2.2	OPP	3	Level 2 (-2.28)
412	.XXXX	-2.4	OOP	4	
275	.XX	-2.6	OOPPPP	6	
204	.XX	-2.8	PPPP	4	
157	.X	-3.0	OPP	3	
117	.X	-3.2	OOO	3	
102	.X	-3.4	OOP	3	
96	.	-3.6	OOP	4	
67	.	-3.8	OO	2	
267	.XX	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Criterion 3: Range Adequacy

This criterion has three components. Each component helps evaluate how well the ELPAC addresses the full range of content described by the ELD Standards. The standards are arranged according to modes of communication, which include collaborative, interpretive, and productive, and by ways of using language, which include structuring cohesive texts, expanding and enriching ideas, and connecting and condensing ideas. For the assessments to fully represent the standards, items should address each of the modes of communication and each of the ways of using language.

In addition, the test items should fully address the measurement construct represented by the standards. Webb's method requires that test items should represent at least half of the standards on which they are based to produce a score. This requirement is also reasonable for an English language proficiency assessment and is included as the third component of the Range Adequacy criterion. The three components of the Range Adequacy criterion are summarized below.

- Across the four domains, each of the three modes of communication (collaborative, interpretive, and productive) is measured by at least one item.
- Across the four domains, each of the three ways of using language (structuring cohesive texts, expanding and enriching ideas, and connecting and condensing ideas) is measured by at least one item.
- At least 50 percent of the ELD Standards are represented by at least one item.

Tables 3.13 through 3.16 summarize the results for the first two components of the Range Adequacy criterion by grade level/span. Each assessment is represented in the tables by two columns, the first representing the primary standard indicated by the panelists and the second representing the secondary standard (labeled Prim and Sec, respectively). All tests met the first component and included items measuring each of the modes of communication. Grades K, six through eight, nine and ten, and eleven and twelve also met the second component and included items measuring each of the ways of using language. Grade one panelists did not rate any items as measuring ways of using language. Grade two panelists indicated the test included items measuring structuring cohesive texts and connecting and condensing texts but did not include items measuring expanding and enriching ideas. Grade three through five panelists indicated the test measured structuring cohesive texts and expanding and enriching ideas but did not measure connecting and condensing ideas. It should be noted that ways of using language were often measured through secondary standards and typically on more complex Writing and Listening items, and there were fewer of those than Reading and Speaking items.

Table 3.13. Range Adequacy Criterion: Mode of Communication: K and Grades 1, 2, and 3 through 5

Mode of Communication	K Prim	K Sec	1 Prim	1 Sec	2 Prim	2 Sec	3–5 Prim	3–5 Sec
Collaborative	6	9	6	0	14	7	9	2
Interpretive	26	3	43	4	36	5	46	5
Productive	10	3	10	3	9	4	8	6

Table 3.14. Range Adequacy Criterion: Ways of Using Language: K and Grades 1, 2, and 3 through 5

Ways of Using Language	K Prim	K Sec	1 Prim	1 Sec	2 Prim	2 Sec	3–5 Prim	3–5 Sec
Structuring cohesive texts	6	1	0	0	6	1	1	4
Expanding and enriching ideas	1	1	0	0	0	0	2	1
Connecting and condensing ideas	2	0	0	0	0	1	0	0

Table 3.15. Range Adequacy Criterion: Mode of Communication: Grades Six through Eight, Nine and Ten, and Eleven and Twelve

Mode of Communication	6–8 Prim	6–8 Sec	9–10 Prim	9–10 Sec	11–12 Prim	11–12 Sec
Collaborative	10	3	6	0	8	2
Interpretive	41	2	39	2	45	1
Productive	9	5	11	6	8	3

Table 3.16. Range Adequacy Criterion: Ways of Using Language: Grades Six through Eight, Nine and Ten, and Eleven and Twelve

Ways of Using Language	6–8 Prim	6–8 Sec	9–10 Prim	9–10 Sec	11–12 Prim	11–12 Sec
Structuring cohesive texts	1	3	5	3	2	3
Expanding and enriching ideas	3	5	1	7	2	5
Connecting and condensing ideas	2	0	4	0	1	4

Results for the final component of the Range Adequacy criterion is presented in table 3.17. The table presents the grade level, the number of standards represented by at least one item, and the total number of standards (presented as a fraction, e.g., 13/19 standards represented by at least one item in kindergarten), followed by the percentage of standards the fraction represents, and an indicator of acceptability (Yes if at least 50% of the standards are represented). This component indicates how completely the standards are represented on the assessments.

Table 3.17. Range Adequacy Criterion: Standards Representation

Grade/Grade Span	Standards Represented by Items on Test	Acceptable?
K	13/19 (68%)	Yes
1	8/19 (42%)	No
2	15/19 (79%)	Yes
3–5	15/19 (79%)	Yes
6–8	16/19 (84%)	Yes
9–10	17/19 (89%)	Yes
11–12	17/19 (89%)	Yes

ELPAC grade eleven and twelve met all components of the Range Adequacy criterion. Assessments for grade two and grade span three through five also met the standards representation component but did not meet the requirement to include items from all three ways of using language, as described in the standards. The grade one assessment did not meet any of the components of the Range Adequacy criterion.

Criterion 4: Balance-of-Knowledge Correspondence (Revised for English Language Proficiency)

This criterion is evaluated based on number of items that panelists rate as directly and clearly matched to a mode of communication (e.g., collaborative), or a way of using language (e.g., structuring cohesive texts). Webb’s balance-of-knowledge correspondence index is computed separately for mode of communication and ways of using language. It builds upon the Range Adequacy criterion by quantifying the extent to which balance is achieved within the modes of communication and ways of using language. The balance index is computed based on the total number of items that were matched to either dimension of language development. The index ranges from 0 to 100, with 100 indicating perfect balance (e.g., if there were exactly the same number of items for each of the three modes of communication). The criterion is considered Acceptable if the calculated balance index is 70 or higher (Webb, 2007). Tables 3.18 through 3.24 provide the balance index for modes of communication and for ways of using language for each grade/grade-span ELPAC test form.

Table 3.18. Balance of Knowledge Correspondence: Kindergarten

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	82	Yes
Ways of using language	64	No

Table 3.19. Balance of Knowledge Correspondence: Grade One

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	62	No
Ways of using language	NA	NA

Table 3.20. Balance of Knowledge Correspondence: Grade Two

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	78	Yes
Ways of using language	61	No

Table 3.21. Balance of Knowledge Correspondence: Grades Three through Five

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	66	No
Ways of using language	88	Yes

Table 3.22. Balance of Knowledge Correspondence: Grades Six through Eight

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	72	Yes
Ways of using language	76	Yes

Table 3.23. Balance of Knowledge Correspondence: Grades Nine and Ten

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	69	No
Ways of using language	87	Yes

Table 3.24. Balance of Knowledge Correspondence: Grades Eleven and Twelve

ELPAC Dimension	Balance Index	Acceptable?
Modes of communication	65	No
Ways of using language	87	Yes

The only grade-level Summative ELPAC test form that met the balance criterion for both modes of communication and ways of using language was the one for grades six through eight. No balance statistic is reported for grade one ways of using language since no items were rated as measuring those standards. The remainder of the results are mixed, indicating the Summative ELPAC is not typically balanced in terms of the number of items representing any grouping of standards by modes of communication or ways of using language.

It should be noted that these results treat primary and secondary standards matches as if items rated for multiple standards appeared twice on the assessment. This is consistent for judging the emphasis of the items by standards dimension on the assessment. The balance index represents how evenly these aspects of the standards are represented overall on the tests.

Summary and Discussion

Summary Results

Table 3.25 summarizes the alignment criteria results for the Summative ELPAC for all grades/grade spans. These results show that the Summative ELPAC items are linked to standards across all assessments, although grade one panelists did not indicate secondary standards for several items. The Link to ELD Proficiency Levels criterion showed that most Summative ELPAC items reflected knowledge, skills, and abilities associated with the Emerging or Expanding proficiency levels. Few items reflected the knowledge, skills, and abilities associated with the Bridging level, and no test met the requirement of 25 percent of items at the Bridging level. At the higher grades, most of

the items were rated as Expanding, while lower grades had more Emerging items. The Summative ELPAC tended to do a good job addressing most of their associated standards, especially those linked to modes of communication. Fewer items were associated with ways of using language, especially in grades one, two, and the three through five span. The Summative ELPAC is not evenly balanced by number of test items for either modes of communication or ways of using language.

Each criterion for each grade level/span reported in table 3.25 is labeled as “Met,” “Not Met,” or “Partially Met.” A criterion is labelled as partially met if there were multiple indicators for the criteria (see table 3.1, page 322 for full descriptions) and at least one of those indicators was met. For example, the criterion Link to Standards is partially met for grade one because all grade one items were linked to a primary standard, but fewer than 90 percent of items designed to measure a secondary standard were linked to a secondary standard by panelists. The first indicator for this criterion was met, but not the second. The definition of “partially met” was not included in the alignment criteria but was developed after the analysis.

Table 3.25. Summative ELPAC Alignment Results

Criterion	K	1	2	3–5	6–8	9–10	11–12
Link to Standards	Met	Partially Met	Met	Met	Met	Met	Met
Link to ELD Proficiency Levels	Partially Met	Partially Met	Not Met	Not Met	Not Met	Not Met	Not Met
Range Adequacy	Met	Partially Met	Partially Met	Partially Met	Met	Met	Met
Balance-of-Knowledge Correspondence	Partially Met	Not Met	Partially Met	Partially Met	Met	Partially Met	Partially Met

Discussion

While the results presented above include several instances where the Summative ELPAC partially met or did not meet the alignment criteria established before the workshop, they do provide a great deal of information that could be used to improve the ELPAC. In this section we will discuss each criterion, any follow-up analyses that might provide further context for the results, and potential changes to the alignment criteria for future investigations of alignment.

For Link to Standards, only grade one failed to meet the criterion. When we examined the individual data from the panelists, we found that half or more of the panelists provided a secondary rating where no consensus rating was recorded. The failure to reach consensus on which secondary standard the items assessed, or whether they assessed a secondary standard at all, led the facilitator to leave the secondary standard field in the spreadsheet blank. This was the only group that failed to reach consensus

on any “alignment to standard” field. Unfortunately, there was not sufficient agreement among the individual panelists on these secondary standards to generate a consistent “majority” rating either. It is also the case that the grade one panel did not select, as the consensus rating, any standards related to ways of using language. Some of these standards were selected by individual panelists as their initial ratings, but none of those ratings were clearly the majority rating, nor were any selected after group discussion. Ultimately, all Summative ELPAC items were rated as aligned to standards, but some of the items that were expected to align to multiple standards were judged not to be sufficiently related to secondary standards by the grade one panel.

For Link to ELD Proficiency Levels, the Summative ELPAC did not meet the criterion for any grade/grade span assessment. Most items were clustered in the middle (Expanding) ELD levels, especially in the higher grades. The lower grades tended to have a mix of lower (Emerging) and middle (Expanding) level items. There were few higher (Bridging) level items at any grade level. This could signal a need to expand the item pool to include more items that address both the Emerging and Bridging level knowledge, skills, and abilities described in the ELD proficiency levels. It may also indicate that there is some misinterpretation of the ELD proficiency levels by the panelists.

Item-person (Wright) maps provided by ETS demonstrated that there are relatively few items with difficulties located in the higher scoring ranges of the Summative ELPAC. Most of the items are clustered in scoring levels 1 and 2, although most students’ scores are clustered around level 3. While the panelists’ ELD Proficiency results cannot be directly compared to the scoring results (there are 3 ELD Proficiency levels, each with Early and Exiting sublevels, and 4 ELPAC scoring levels), both data sources indicate a lack of higher-level items. (See figure 3.1 for comparison of ELD Proficiency levels to ELPAC scoring levels.)

Although panelists’ ratings of item ELD proficiency levels cannot be directly compared to the scoring results, the two values can be correlated. Table 3.26 presents the correlation between the ELP proficiency level ratings made by panelists and the ELPAC performance levels based on the item difficulty parameters from the metadata and the performance level threshold scores. At the lower grade levels, there is no relation between the two. This is likely due to the restriction of range resulting from many items classified at the lower performance levels, as evidenced by the item-person maps. For the remaining grade spans starting at grade 3 through 5, there is a moderate, positive correlation, indicating some similarity among panelist-based ELD proficiency levels ratings and metadata-based ELPAC performance level classifications.

Table 3.26. Correlation Between Panelist ELD Proficiency Level Ratings and Metadata-based ELPAC Performance Levels

Grade Span	Correlation
Kindergarten	**
Grade 1	**
Grade 2	**
Grades 3–5	.47
Grades 6–8	.39
Grades 9–10	.51
Grades 11–12	.45

**Not statistically significantly different from 0 ($p < .05$)

The item metadata also contains targeted ELPAC performance level(s) for each item. This provides an indication of the performance level targeted by item writers during the item development process. Correlating panelists' ELD proficiency level ratings, metadata-based ELPAC performance level item classifications, and targeted ELPAC performance levels provides an indication of the quality of panelists' ELD proficiency level ratings. Table 3.27 presents these correlations for dichotomously scored items. Polytomously scored items were not included, as they were often written with multiple student performance levels in mind. Table 3.27 shows three columns of correlations. Panelists' ELD proficiency level ratings tended to have the strongest correlations with the targeted performance levels, and these correlations tended to be higher than the correlations between the metadata-based item ELPAC performance levels and the targeted performance levels. The second column of the table (Ratings-Metadata) shows, by grade span, the correlation between panelists' ratings and metadata (actual item performance). These correlations are higher than those in the fourth column (Metadata-Targets), which shows the correlation between the metadata and the performance targets (based on item type) provided by the testing contractor. This indicates that the panelists' ratings were closer to the actual data than the targets, supporting the accuracy of the panelists' ratings. The highest correlations were between the panelists' ratings and the performance targets (third column, Ratings-Targets). This shows that panelists tended to rate according to the expected performance levels associated with each item type as well.

The Range Adequacy criterion was fully met for grades kindergarten, six through eight, nine and ten, and eleven and twelve. Grades two and the three through five grade span missed meeting this criterion due to a single component, the one requiring that all ways of using language should be represented by items. Grade one missed two of the components for this criterion, largely because the panels did not reach consensus on several items' secondary standards and because the grade one panel did not indicate any of the ways of using language standards as a consensus rating for any item. In most cases, the ways of using language standards are met via secondary standards, which are only available for the Speaking and Writing domains of the Summative

ELPAC. These standards may be more difficult to measure than the modes of communication standards and may also be more difficult for panelists to rate.

Table 3.27. Correlations Between Panelist ELD Proficiency Level Ratings, Metadata-based ELPAC Performance Levels, and Targeted ELPAC Performance Levels: Dichotomously Scored Items Only

Grade Span	Ratings-Metadata	Ratings-Targets	Metadata-Targets
Kindergarten	**	.69	.35
Grade 1	.60	.70	.58
Grade 2	.43	.56	**
Grades 3–5	.51	.76	.57
Grades 6–8	**	.67	.32
Grades 9–10	**	.40	.52
Grades 11–12	.40	.54	.32

**Not statistically significantly different from 0 ($p < .05$)

It is also important to note that the metadata related to secondary standard often indicated multiple standards (as many as eight standards in Speaking and seven in Writing) for a single item. The alignment study limited panelists to selecting only two standards for any single item. It is difficult to imagine a single item that substantively addresses as many as nine standards (one primary plus eight secondary standards) but limiting panelists to a single secondary standard may have attenuated the results of the alignment study for the Range Adequacy criterion.

The Balance-of-Knowledge Correspondence criterion was met only for the grade six through eight test form. The number of items per domain (e.g., Listening, Speaking, Reading, and Writing) varies greatly for all grade levels. Listening and Reading tend to have many more items than Speaking and Writing. Speaking and Writing items are the only ones that routinely include secondary standards. This leads to tests that are not balanced by strict numbers or percentages of items. If the imbalance is intentional, it may be prudent to capture that imbalance in test blueprints so that it can be addressed in future alignment analyses. ELPAC is reported by combining Reading and Writing into a written language score, and by combining Speaking and Listening into an oral language score. The number of items is more balanced if we compare written language to oral language, but the analyses for this study were based on the more discrete domain scores.

Chapter 4: Conclusions and Recommendations

This study combined documentation review and item ratings by content experts to evaluate the alignment between the Summative ELPAC and the California ELD Standards. The documentation review was successful. Of the 15 identified *Testing Standards* we reviewed in respect to the alignment study, eleven were rated as fully covered based on the available evidence and four were rated as mostly covered. These results indicate that the ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to English language development standards.

Data from the alignment workshop component of the study provides mixed support for the alignment of the Summative ELPAC to California's ELD Standards. These mixed results may reflect the intentional priorities of the assessment program, but those intentions are not manifest in test blueprints or similar guiding documents. In this section, we will draw conclusions and make recommendations. In several instances, the recommendations are presented as options to either revise the assessment or the guiding documents to make alignment priorities clear.

First, all items on the Summative ELPAC, across all grades and language domains, were rated as aligned to ELD Standards by panelists. No items were flagged for poor quality or as outside the measurement construct. This represents strong evidence that the Summative ELPAC does reflect its intended construct.

Examining an item's link to standards is often accompanied by counting items associated with each score or subscore. Webb's categorical concurrence criterion requires at least six items per score to generate reasonable reliability for reporting. ETS provided a technical report for the 2018–2019 administration that includes both reliability estimates and classification accuracy results (CDE, 2020). These statistics are much more appropriate for judging the reliability of the assessment scores than simply counting items. The technical report indicates acceptable reliabilities and classification accuracy. Overall reliability coefficients (alpha) for the Summative ELPAC ranged from 0.89 to 0.94. Subtest reliability estimates ranged from 0.63 to 0.92, and most were higher than 0.80 (Appendix 6.B, p. 519). Overall classification accuracy ranged from 0.76 to 0.81 (p. 512).

Recommendation 1. Review grade one Summative ELPAC items that are intended to measure multiple standards to verify that students must demonstrate language abilities related to the intended secondary standard to correctly respond to the item.

The grade one panel found six items that were intended to measure multiple standards measured only one standard. This led to weaker alignment statistics for grade one than for other grades. Furthermore, the grade one panel did not align any test item with standards related to ways of using language. Some panelists did indicate these standards in their individual ratings, but none of those ratings were selected after consensus. This may indicate an issue with the panelists or an issue with the grade one

test items. We recommend reviewing those six items prior to making more substantive revisions to the assessment.

Recommendation 2. Develop additional items at the Bridging ELD proficiency level for all grades.

The Link to ELD Proficiency Level is a new criterion for alignment studies. It was created for this study to examine how the Summative ELPAC items assessed the knowledge, skills, and abilities of students as described in the ELD proficiency levels. The ELD proficiency levels include much of the same information as would be found in a linguistic difficulty scale, plus other information to describe student performance more fully at each proficiency level. The ELD proficiency levels are also specific to California, whereas no linguistic difficulty scale is specific to California.

The ELD Standards, however, are not designed to address a single proficiency level. The standards documents describe student performance at each proficiency level by modes of communication or ways of using language, and those constructs are represented by small groups of standards. It would not be appropriate to assign a “level” to each standard for matching since each standard can be demonstrated at multiple proficiency levels. This issue is often overlooked when a linguistic difficulty scale is used because each standard must be assigned a linguistic difficulty level that can be matched to item ratings to compute an index. However, California’s ELD proficiency levels demonstrate just how problematic that practice can be if the standards are not discrete by level (and they rarely are). If the standards may be addressed at multiple levels, then it is more sensible for an alignment study to verify that the items address the full range of linguistic difficulty, or ELD proficiency levels, intended by the assessment. The Link to ELD Proficiency Level criterion takes this approach.

It should be noted that there were no test blueprints that indicated how many items should be expected to target each ELD proficiency level, nor was there information related to ELD proficiency level in the item metadata. The ranges for how many items were expected to fall in each level (per the alignment criteria) were established logically, rather than empirically. We reasoned that an assessment that included three proficiency levels (each split into “Early” and “Exiting” sublevels) should include a substantial number of items addressing each level. Alignment criterion thresholds were set at 25 percent as the minimum percent of items at any level (Emerging, Bridging, Expanding) and 50 percent as the maximum. This distribution would allow for similarly accurate score estimation across the full range of performance described by the ELD proficiency levels. Accurate scoring at all levels is one key for answering Research Question 1 and supporting the claims associated with students’ ELPAC performance level. The accuracy of scores toward the higher range is particularly important since students must score toward the upper half of the Bridging level in all language domains to be considered proficient overall.

Results show that the items tend to be clustered toward the lower (Emerging) and middle (Expanding) ELD proficiency levels, especially at the higher grades. Lower grades had more items that assessed the lower level. No Summative ELPAC met the threshold of 25 percent of items at the higher (Bridging) level. The Summative ELPAC

has three important cut scores and generates four performance categories, one of which is toward the upper end of the ELD proficiency levels.

Panelists' results were compared with item-person (Wright) maps provided by ETS. Those data support the panelists' findings. There are few items on the ELPAC with difficulties in the part of the scale associated with the top score category. This means that students who score in that category must get nearly all of the test items correct and that the accuracy of those scores may not be as high as for other parts of the ELPAC.

Recommendation 3. Revise test items and/or metadata to better reflect the specific “ways of using language” standards that are intended to be measured.

The Summative ELPAC assessments were not evenly balanced with respect to modes of communication or ways of using language. The differences were sufficiently large that the test did not seem designed to address each of these concepts with an equal number of items, as is suggested by the Webb alignment method. If the assessments are designed to emphasize certain content over other content, or if there are items that should be given more weight, those decisions should be reflected in a test blueprint or similar document. This alignment study set a criterion for balance that assumed equal emphasis, but those analyses may be inaccurate if that assumption was incorrect. It is also concerning that the metadata indicated as many as eight secondary standards associated with a single item. Inclusion of more specific targets in the blueprints and/or item metadata could help ensure that these standards are a priority during item development.

The Range Adequacy criterion was met for the Summative ELPAC in kindergarten and grade spans six through eight, nine and ten, and eleven and twelve. Tests in the other grades/grade spans did not meet this criterion because they did not include enough items aligned to the standards reflecting “ways of using language.” These standards might be considered more difficult to measure than those associated with “modes of communication.” These standards were often assigned to items as secondary ratings. The grade one panel indicated that several items that were intended to measure secondary standards did not do so clearly enough to reach a consensus rating.

Recommendation 4. Refine the test blueprint to specify (a) the number and/or percentage of items associated with each of the modes of communication and ways of using language and (b) the number and/or percentage of items at each ELD proficiency level.

If adjustments to the test blueprint are adopted, it would also be helpful to include proportions of items at each ELD proficiency level. A test blueprint can be a valuable tool for ensuring that the test designers and developers create an assessment that adequately reflects the intentions of educators in measuring the ELD Standards. It can also provide a roadmap for future investigations of test-to-standards alignment.

This page is intentionally blank.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- California Department of Education. (2012). California English Language Development Standards (Electronic Edition): Kindergarten through Grade 12). Retrieved from: <https://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf>.
- California Department of Education. (2019). Updated Reclassification Guidance for 2018–19. Retrieved from: <https://www.cde.ca.gov/sp/el/rd/reclassguide1819.asp>
- California Department of Education (2020). 2020–21 English Language Proficiency Assessments for California Information Guide. Retrieved from <https://www.cde.ca.gov/ta/tg/ep/documents/elpacinfo20.pdf> California Department of Education (2020). Summative English Language Proficiency Assessments for California Technical Report: 2018–2019 Administration. Retrieved from: <https://www.cde.ca.gov/ta/tg/ep/documents/selpac19techrpt.pdf>
- Cook, H. G. (2005). Aligning English language proficiency tests to English language learning standards. In *Aligning assessment to guide the learning of all students* (pp. 135–153). Washington, DC: Council of Chief State School Officers.
- Cook, H. G. (2005). *Aligning English Language Proficiency Tests to English Language Learning Standards*. State Collaborative on Assessment and Student Standards: Council of Chief State School Officers (pp. 131–54).
- Cook, H. G. (2007). *Alignment study report: The WIDA Consortium's English language proficiency standards for English language learners in kindergarten through grade 12 to ACCESS for ELLs assessment*. Madison, WI: University of Wisconsin.
- Every Student Succeeds Act, 114–95 C.F.R. (2015).
- Gil, L., & Bardack, S. (2010). *Common assumptions vs. the evidence: English language learners in the U.S. A reference guide*. American Institutes for Research. Washington, DC. Retrieved from: <https://files.eric.ed.gov/fulltext/ED511353.pdf>
- National Center for Education Statistics. (2020). *Digest of Education Statistics, 2019*. Retrieved from https://nces.ed.gov/programs/digest/2019menu_tables.asp
- National Center for Educational Statistics. (n.d.) *Fast Facts: English language learners*. Retrieved from <https://nces.ed.gov/fastfacts/display.asp?id=96>
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25, 47–55.

- U.S. Department of Education (n.d.). ED Data Express. Retrieved from <https://eddataexpress.ed.gov/>
- Valdés, G. (2001). *Learning and not learning English: Latino students in American schools*. New York: Teachers College Press.
- Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.
- Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states (Research Monograph No. 18). Madison, WI: University of Wisconsin–Madison, National Institute for Science Education.
- Webb, N. (2006). Identifying content for student achievement tests. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 155–180). Mahwah, NJ: Erlbaum.
- Webb, N. L. (2007). Issues Related to Judging the Alignment of Curriculum Standards and Assessments. *Applied Measurement in Education*, 20(1), 7–25. Madison, WI: University of Wisconsin–Madison. Retrieved from <https://www.cehd.umn.edu/edpsych/C-BAS-R/Docs/Webb2007.pdf>

Glossary of Acronyms

Acronym	Glossary
CAASPP	California Assessment of Student Performance and Progress
CBA	Computer Based Assessment
CDE	California Department of Education
CRT	Content Review Tool
DFA	Directions for Administration
EL	English Learner
ELD	English Language Development
ELP	English Language Proficiency
ELPAC	English Language Proficiency Assessments for California
ESEA	Elementary and Secondary Education Act
ESSA	Every Student Succeeds Act
LDL	Linguistic Difficulty Level
NCES	National Center for Education Statistics
NCLB	No Child Left Behind Act
PLD	Proficiency Level Descriptor (for ELD Standards)

This page is intentionally blank.

Appendix A: Summative ELPAC Documentation Reviewed by HumRRO

Table A.1. Summative ELPAC Documents Reviewed

Document Focus	Document File Name
English Language Development Standards	<ul style="list-style-type: none"> • California ELD Standards (eldstndspublication14.pdf) • Correspondence Study Report by WestEd 6-3-2015.pdf • Integrating CA ELD Standards into K-12 Math and Science Teaching and Learning 12-16-2015.docx
Test Design	<ul style="list-style-type: none"> • 461-2019_ELPAC CBA Summative Blueprints 06-22-20 for HumRRO • Definitions of Task Types for the ELPAC 3-2020.pdf • 685-2021 v2 FOR ARCHIVE ELPAC Summative Test Development Specifications_062220.docx • 1456-2019 v5 FOR ARCHIVE ELPAC CBA High-Level Test Design_052119.docx • ELPAC SA 2021 PKG Configuration Visio_v6_FINAL.pdf • ELPAC SA 2021 Configuration Considerations_v2_070920.docx • 562-2019-V4 FOR ARCHIVE_ELPAC Summative Technical Report.072020-Word.zip • Mode Comparability Study memo (aug20adad02.docx) • ELPAC Design_Standards and Practices_Final with Audio • CBE September 2017 Agenda (sep17item18.docx) • CBE May 2019 Agenda (may19item01.docx) • Considerations in the transition of the English Language Proficiency Assessments for California (ELPAC) Paper-Pencil Tests to Computer-Based Assessments (elpacbareporttagged.pdf) • 676-2020_v8_FOR ARCHIVE ELPAC CBA Conversion Specifications D-305 • Summative ELPAC Assessment Fact Sheet (sumelpacfactsheet.pdf) • Anchor Charts created SRF 2020

Table A.1. (cont.)

Document Focus	Document File Name
Item Development and Information	<ul style="list-style-type: none"> • Definitions of Task Types for the ELPAC 3-2020 • ELPAC_Stimulus_WordCounts • 690-2020A V2 FOR ARCHIVE Revised_2019-20 ELPAC Educator IWT Presentation 090919 • 690-2020A V2 FOR ARCHIVE ELPAC Educator IWT_Fairness Guidelines and Item Review Checklist • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)_080720 • 706-2020A_v3_FOR ARCHIVE ELPAC 2020 New Item Review PPT_123019 • Participants List Spreadsheet 092820 • 696-2020 v4 FOR ARCHIVE ELPAC Item Writing Guidelines (RFR A-47)__080720.docx • 676-2020_v8_FOR ARCHIVE ELPAC CBA Conversion Specifications D-305.docx • 689-2021D_v2_FOR ARCHIVE_ELPAC 2021_Summative gR3-5 Form 1_082020_FINAL
Test Administration	<ul style="list-style-type: none"> • 2020B_v4 FOR ARCHIVE CAASPP ELPAC Post-Test Survey 052920.docx • K-2 Writing Answer Books: 2021 Summative ELPAC Writing AB K_Final.pdf • 2021 Summative ELPAC Writing AB G1_Final.pdf • 2021 Summative ELPAC Writing AB G2_Final.pdf • DFAs for all grades/grade spans • 735-2021A_v2_FOR REVIEW_2021-Summative-DFA_LSRW • ELPAC--How-to-Use-the-Technology-Readiness-Checker-for-Students • ELPAC--Listening-Speaking-Reading-and-Writing-Training-Test-DFA-Grade-K.2020-21 • ELPAC Moodle Training Site • ELPAC--Writing-Training-Test-Grade-K.2020-21
Item Scoring	<ul style="list-style-type: none"> • CDE Writing Anchor Sample reviews-2020 Writing RF.xlsx • 2020 CDE Approved Anchors.zip • Anchor Charts Created SRF 2020.xlsx • ELPAC_Writing-Rubrics
Score Reporting	<ul style="list-style-type: none"> • https://www.elpac.org/s/pdf/Summative-ELPAC-Grade-5-Sample-Student-Score-Report-English.2019-20.pdf

Table A.1. (cont.)

Document Focus	Document File Name
Accessibility	<ul style="list-style-type: none"> • Accessibility and Usability for the English Language Proficiency Assessments for California: A Cognitive Lab Study with Students Who Are Deaf or Hard of Hearing and Students Who Are Blind or Have Low Vision (elpaccognitiverpt.pdf) • California Assessment Accessibility Resources Matrix (caaccessibilitymtrx2021.docx)
Field Test	<ul style="list-style-type: none"> • 644-2020-v2_FOR ARCHIVE_Summative ELPAC CBA FT Technical Report-clean • 562-2019-v4-FOR ARCHIVE_ELPAC Summative Technical Report.072020-clean
Standard Setting	<ul style="list-style-type: none"> • ELPAC memo-oct17item01 Standard Setting Plan.doc • C-211 Summative ELPAC Standard Setting Technical Report v6 APP.doc • B-25 REFERENCE-ELPAC Report on Domain Specific PLDS_v4_APP • English Language Proficiency Assessments for California: Approve the Operational Summative Assessment Threshold Scores and Composite Weights for the English Language Proficiency Assessments for California and Approve the Local Educational Agency Apportionment Rates (nov17item08addendum) • ELPAC CBA - Evidence of Complexity 8-28-2020

Note: Documents reviewed may not have been cited for supporting an evaluated testing standard.

This page is intentionally blank.

Appendix B: Materials for Virtual Alignment Workshop

List of Materials

- Virtual Workshop Agenda (March 22–25, 2021)
- Panelist Instructions
- Sample Panelist Rating Form
- ELD Standards Proficiency Levels Rating Scale
- Summative ELPAC Alignment Overall Debrief Survey
- Evaluation: Virtual Alignment Workshop Training and Procedures
- Summative ELPAC Alignment Panelist Demographic Information Survey

Virtual Workshop Agenda (March 22–25, 2021)

Day 1 - Monday, March 22

- 1:00 – 1:50 p.m. Join Teams Meeting with All Panelists, HumRRO Facilitators, and California Department of Education Staff
Welcome, logistics, overview of ELPAC, general alignment training
- 1:50 – 2:00 p.m. BREAK
- 2:00 – 3:15 p.m. Join Teams Meeting for Assigned Grade Level Panel
Panelist introductions
Confirm access to online documents and Content Review Tool
Review Panelist Instructions for rating Summative ELPAC items
- Identify Primary ELD Standard (and Secondary ELD Standard for Speaking and Writing only)
 - Assign ELD proficiency level rating
- Begin iterative alignment rating process:
- Independent rating
 - Discussion and consensus building
 - Panel review of metadata
 - Final independent and consensus ratings
- 3:15 – 3:30 p.m. Break
- 3:30 – 5:00 p.m. Continue iterative alignment rating process
Adjourn for the day

Day 2 - Tuesday, March 23

- 1:00 – 2:00 p.m. If needed: Review and rerate items from Day 1
Continue iterative alignment rating process
- 2:00 – 2:15 p.m. Break
- 2:15 – 3:15 p.m. Continue iterative alignment rating process
- 3:15 – 3:30 p.m. Break
- 3:30 – 5:00 p.m. Continue iterative alignment rating process
Adjourn for the day

Day 3 - Wednesday, March 24

- 1:00 – 2:00 p.m. If needed: Review and rerate items from Day 2
Continue iterative alignment rating process
- 2:00 – 2:15 p.m. Break
- 2:15 – 3:15 p.m. Continue iterative alignment rating process
- 3:15 – 3:30 p.m. Break
- 3:30 – 5:00 p.m. Continue iterative alignment rating process
Adjourn for the day

Day 4 - Thursday, March 25

- 1:00 – 2:00 p.m. If needed: Review and rerate items from Day 3
Continue iterative alignment rating process
- 2:00 – 2:15 p.m. Break
- 2:15 – 3:15 p.m. Complete iterative alignment rating process
- 3:15 – 3:30 p.m. Break
- 3:30 – 4:45 p.m. Review documents from the panelist packet
- ELPAC performance level descriptors
 - Academic performance level descriptors
 - Standards for mathematical practices
 - Standards for the science and engineering practices
- As a panel, create summary statement about the quality of link between the English language proficiency expectations and academic content expectations
- 4:45 – 5:00 p.m. Complete two short online surveys:
- Debrief/ Workshop evaluation
 - Demographic information
- Adjourn*

Panelist Instructions

#	Panelist Materials	File Location on Google Drive
1	Summative ELPAC Panelist Instructions	Non-secure
2	Content Review Tool (CRT)	IBIS Online Tool
3	K, 1, and 2 Writing Item Rubrics (Gr 3-12 Writing Rubrics are in CRT)	Secure (Grades K, 1, 2 only)
4	Summative ELPAC Rating Form	Individual link in 3/16 email
7	Directions for Administration (DFA)	Secure
8	CA English Language Development (ELD) Standards	Non-Secure (hard copy)
9	ELD Proficiency Level Descriptors	Non-Secure (hard copy)
10	Summative ELPAC Performance Level Descriptors	Non-Secure (hard copy)
11	Academic Performance Level Descriptors	Non-Secure (hard copy)
12	CA Standards for Mathematical Practice	Non-Secure (hard copy)
13	Science and Engineering Practices in the NGSS	Non-Secure (hard copy)
14	English Language Arts Standards	Non-Secure
15	Debriefing: Analysis of Alignment Outcomes	Link to be sent via email
16	Panelist Demographic Information	Link to be sent via email
17	Workshop Evaluation	Link to be sent via email

Prior to the Workshop (by 3/21): Technology check to confirm access to MS Teams meeting.

Prior to alignment ratings:

1. Introductions
2. Review of panelist materials
3. Familiarization with Content Review Tool (CRT) for accessing items to be rated.

Rate Summative ELPAC Items

Orient to Rating Form:

1. You will review several Summative ELPAC items and will enter the ELD standard rating(s) and ELD proficiency level rating for each item.
2. Access Summative ELPAC Rating Form Google Sheet file:
 - a. Select the link in email sent on March 16 (Final Instructions)
 - b. Your Google Sheets will save automatically, no manual saving needed.
3. Review rating categories using your Google Form
 - a. You will only need to review items on the first tab. The other tab is for internal use only.

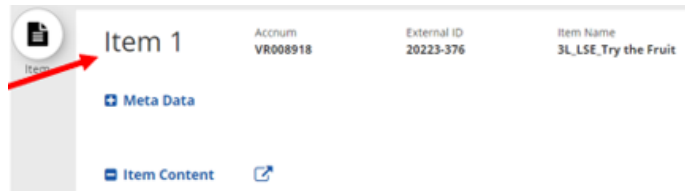
C	D	E	F	G	H	I	J	K	L	M
Index # (matches CRT)	Item Type	Max Points	Domain	Identify the Primary ELD Standard (Select from drop down menu)	Identify the Secondary ELD Standard (Select from drop down menu)	Assign an ELD proficiency level for scoring 1 point on the item (Select from drop down menu)	Assign an ELD proficiency level for scoring 2 points on the item (Select from drop down menu)	Assign an ELD proficiency level for scoring 3 points on the item (Select from drop down menu)	Assign an ELD proficiency level for scoring 4 points on the item (Select from drop down menu)	Comments (Provide any comments about the appropriateness of stimuli, item quality, etc.)
1	Set Leader		Listening							
2	MCSS - Member	1	Listening							
3	Set Leader		Listening							
4	MCSS - Member	1	Listening							
5	Set Leader		Listening							
6	MCSS - Member	1	Listening							
7	MCSS - Member	1	Listening							
8	Set Leader		Listening							
9	MCSS - Member	1	Listening							

- b. Columns C through F contain information about each Summative ELPAC item.
 - Column C provides the index number, which corresponds to the number of the item in the Content Review Tool (CRT) or the Writing item number for Grades K, 1, and 2 (in DFA Score Info PDF).
 - Column D provides the item type.
 - Column E provides the maximum number of points possible on the item.
 - Column F provides the content domain.
- c. Column G asks for the primary ELD standard measured by the item. An example standard code is **P1.A.1**. All standard codes are presented in a dropdown menu that is accessible by selecting the arrow that appears to the right of the cell where the rating is to be made.
- d. Column H asks for the secondary ELD standard measured by the item.
 - Only Speaking and Writing items will be rated for a secondary standard.
 - Listening and Reading are shaded gray because they will not be rated for a secondary ELD standard.
 - All standard codes are presented in a dropdown menu that is accessible by selecting the arrow that appears to the right of the cell where the rating is to be made (identical choices as for Primary standard).

- e. Column I asks for the ELD proficiency level associated with answering the item correctly or scoring 1 point on an item that is worth more than one point. The ELD proficiency levels are presented in a dropdown menu that is accessible by selecting the arrow that appears to the right of the cell where the rating is to be made.
- f. Column J asks for the ELD proficiency level associated with scoring 2 points on an item that is worth 2 or more points.
- g. Column K asks for the ELD proficiency level associated with scoring 3 points on an item that is worth 3 or more points.
- h. Column L asks for the ELD proficiency level associated with scoring 4 points on an item that is worth 4 points.
- i. Column M is available for entering any comments or notes to clarify or qualify any of the other ratings.

Make item ratings:

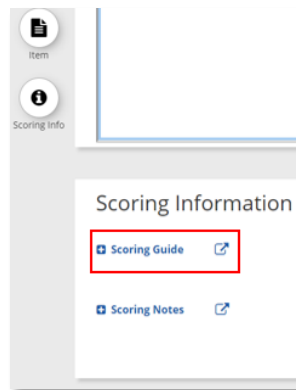
1. Rate the first item independently, all relevant columns.
 - a. Locate the large font “Item” number in the CRT and confirm that it matches the Item Index number on your Google Sheets rating form.



- i. When the item type is “Set Leader” there will be no ratings made. This is a stimulus that subsequent items refer to. You should review the content of the set leader and can make comments in the rating form if they have concerns about its quality. Note that the CRT will also display the first item adjacent to the Set Leader (which can be confusing).
 - ii. Writing items in Kindergarten, Grade 1 and Grade 2 are not presented in the CRT, but in the secure DFAs (pdf files accessible in Google Drive in the Secure Documents folder). For these items, confirm that the item sequence number in the pdf file (black number in a blue box) matches the item sequence number on the rating form. The pdf shows a screen shot of what the printed student version of the Writing item looks like.
 - iii. The grade K, 1, and 2 Writing Rubrics are in the Google Drive Secure Documents folder.
- b. Review the content of the item as well as any related directions presented in the Directions for Administration (DFAs), which are secure PDFs.
 - i. For grades K, 1, and 2, the DFA file includes Speaking items and Writing items. Writing items are only in the DFA, not the CRT. The DFAs (a) include text that test examiners say to the examinee and (b) provide additional scoring information. For Speaking, red

numbers in the DFA align with CRT Item numbers. For Writing, the item numbers match the rating form Item Index Number.

- ii. For grade K, the DFA also includes Reading items (with text test examiners say to the examinee) and provide additional scoring information.
 - iii. For grades 3 – 12, the DFA file is for Speaking items only. The DFAs include text that test examiners say to the examinee and provide additional scoring information. Red numbers in the DFA align with CRT Item numbers.
- c. Review the ELD Standards.
 - d. Use the dropdown menu to rate the primary ELD standard measured by the item.
 - e. For Speaking and Writing items, rate the secondary ELD standard measured by the item, using the dropdown menu.
 - f. Review the six ELD Proficiency Levels (e.g., Level 1 Early Emerging).
 - g. Use the dropdown menu to rate the ELD proficiency level(s) of the item. Multi-point items will be rated more than one time (once for each possible score).
 - i. Grades K-2: refer to the Writing Rubrics PDF.
 - ii. Grades 3-12: select the Scoring Info icon, then select the Scoring Guide to access the item’s Writing rubric in the CRT.



- h. Provide comments, as needed. **All items have been thoroughly reviewed.** Comments are not required, but you may choose to enter comments to:
 - i. provide context for your ratings
 - ii. indicate an issue related to the quality of the item
 - iii. explain if/why you are torn between two ratings
 - iv. explain if/why you do not agree with the final consensus/majority rating
 - v. provide an explanation for a blank ELD Standard field (either Primary or, if applicable, Secondary) if you believe that the item does NOT measure any of the ELD Standards

2. After all panelists have rated the first item, the facilitator will lead the group in a discussion of their independent ratings.
 - a. You should not change ratings during or after this discussion unless you are **certain** you made a data entry error (e.g., coding error or misunderstanding of the standards).
3. The HumRRO facilitator will next share the item metadata (test developer's assigned ELD standard(s)). The group will then discuss any discrepancies. Note that your expert judgments, not the metadata, are the "right answers."
 - a. Do NOT change any *independent* ratings after seeing the metadata.
4. The HumRRO facilitator will poll the group regarding each rating and will capture the final consensus rating. If true consensus cannot be reached, the rating of the majority of panelists will be recorded.
5. The HumRRO facilitator will repeat this process for at least 3 times, one item at a time, until all panelists are comfortable and "calibrated."
6. For the remaining Summative ELPAC items, your facilitator will assign you sets of 5-8 items to review and rate independently.
7. Repeat the process described above (discuss independent ratings, review and discuss metadata, and settle on consensus ratings) for each set of items.
8. Work independently; however, you may occasionally raise a discussion point with the group about any item(s) that are difficult to rate.

Post rating activity:

1. Following the completion of all rating tasks, complete the following three short surveys, using the links sent to you via email:
 - a. Summative ELPAC Alignment Debriefing
 - b. Demographic Questionnaire
 - c. Workshop Evaluation

Sample Panelist Rating Form

Panelists were given a link to their individual Google Sheet to enter their independent ratings and comments about the Summative ELPAC items they review. The screen shots below illustrate that panelists used drop down menus to make their ratings. A comments field was available for all items but is not shown below.

Index # (matches CRT)	Item Type	Max Points	Domain	Identify the Primary ELD Standard <small>(Select from drop down menu)</small>	Identify the Secondary ELD Standard <small>(Select from drop down menu)</small>	Assign an ELD proficiency level for scoring 1 point on the item <small>(Select from drop down menu)</small>
1	Set Leader		Listening			
2	MCSS - Member	1	Listening	PI.A.1		
3	Set Leader		Listening			
4	MCSS - Member	1	Listening			
5	Set Leader		Listening			
6	MCSS - Member	1	Listening			
7	Set Leader		Listening			
8	MCSS - Member	1	Listening			
9	MCSS - Member	1	Listening			

Click the ▾ button to reveal the standards. Rater selects the Primary ELD Standard that best matches the item.

Index # (matches CRT)	Item Type	Max Points	Domain	Identify the Primary ELD Standard <small>(Select from drop down menu)</small>	Identify the Secondary ELD Standard <small>(Select from drop down menu)</small>	Assign an ELD proficiency level for scoring 1 point on the item <small>(Select from drop down menu)</small>	Assign an ELD proficiency level for scoring 2 points on the item <small>(Select from drop down menu)</small>
1	Set Leader		Listening				
2	MCSS - Member	1	Listening			Level 1- Early Emerging	

Click the ▾ button to reveal the proficiency levels. Rater selects the ELD proficiency level that best matches the item (one rating for each possible score).

If a cell is solid grey, then no rating is needed. Set leaders will not be rated.

ELD Proficiency Levels Rating Scale

Mode of Communication	Level 1 - Early Emerging	Level 2 - Exiting Emerging	Level 3 - Early Expanding	Level 4 - Exiting Expanding	Level 5 - Early Bridging	Level 6 - Exiting Bridging
Collaborative	<ul style="list-style-type: none"> Express basic personal and safety needs and ideas and respond to questions on social and academic topics with gestures and words or short phrases. Use basic social conventions to participate in conversations. 	<ul style="list-style-type: none"> Express basic personal and safety needs and ideas and respond to questions on social and academic topics with phrases and short sentences. Participate in simple, face-to-face conversations with peers and others. 	<ul style="list-style-type: none"> Express a variety of personal needs, ideas, and opinions and respond to questions using short sentences. Initiate simple conversations on social and academic topics. 	<ul style="list-style-type: none"> Express more complex feelings, needs, ideas, and opinions using extended oral and written production; respond to questions using extended discourse. Participate actively in collaborative conversations in all content areas with moderate to light support as appropriate. 	<ul style="list-style-type: none"> Express increasingly complex feelings, needs, ideas, and opinions in a variety of settings; respond to questions using extended and more elaborate discourse. Initiate and sustain dialogue on a variety of grade-level academic and social topics. 	<ul style="list-style-type: none"> Participate fully in all collaborative conversations in all content areas at grade level, with occasional support as necessary. Participate fully in both academic and non-academic settings requiring English.

Mode of Communication	Level 1 - Early Emerging	Level 2 - Exiting Emerging	Level 3 - Early Expanding	Level 4 - Exiting Expanding	Level 5 - Early Bridging	Level 6 - Exiting Bridging
Interpretive	<ul style="list-style-type: none"> • Comprehend frequently occurring words and basic phrases in immediate physical surroundings. • Read very brief grade-appropriate text with simple sentences and familiar vocabulary, supported by graphics or pictures. • Comprehend familiar words, phrases, and questions drawn from content areas. 	<ul style="list-style-type: none"> • Comprehend a sequence of information on familiar topics as presented through stories and face-to-face conversation. • Read brief grade-appropriate text with simple sentences and mostly familiar vocabulary, supported by graphics or pictures. • Demonstrate understanding of words and phrases from previously learned content material. 	<ul style="list-style-type: none"> • Comprehend information on familiar topics and on some unfamiliar topics in contextualized settings. • Read independently a variety of grade-appropriate text with simple sentences. • Read more complex text supported by graphics or pictures. • Comprehend basic concepts in content areas. 	<ul style="list-style-type: none"> • Comprehend detailed information with fewer contextual clues on unfamiliar topics. • Read increasingly complex grade-level text while relying on context and prior knowledge to obtain meaning from print. • Read technical text on familiar topics supported by pictures or graphics. 	<ul style="list-style-type: none"> • Comprehend concrete and many abstract topics and begin to recognize language subtleties in a variety of communication settings. • Read increasingly complex text at grade level • Read technical text supported by pictures or graphics. 	<ul style="list-style-type: none"> • Comprehend concrete and abstract topics and recognize language subtleties in a variety of communication settings. • Read, with limited comprehension difficulty, a variety of grade-level and technical texts in all content areas.

Mode of Communication	Level 1 - Early Emerging	Level 2 - Exiting Emerging	Level 3 - Early Expanding	Level 4 - Exiting Expanding	Level 5 - Early Bridging	Level 6 - Exiting Bridging
Productive	<ul style="list-style-type: none"> • Produce learned words and phrases and use gestures to communicate basic information. • Express ideas using visuals such as drawings, charts, or graphic organizers. • Write or use familiar words and phrases related to everyday and academic topics. 	<ul style="list-style-type: none"> • Produce basic statements and ask questions in direct informational exchanges on familiar and routine subjects. Express ideas using information and short responses within structured contexts. • Write or use learned vocabulary drawn from academic content areas. 	<ul style="list-style-type: none"> • Produce sustained informational exchanges with others on an expanding variety of topics. • Express ideas in highly structured and scaffolded academic interactions. • Write or use expanded vocabulary to provide information and extended responses in contextualized settings. 	<ul style="list-style-type: none"> • Produce, initiate, and sustain spontaneous interactions on a variety of topics. Write and express ideas to meet most social and academic needs through the recombination of learned vocabulary and structures with support. 	<ul style="list-style-type: none"> • Produce, initiate, and sustain interactions with increasing awareness of tailoring language to specific purposes and audiences. • Write and express ideas to meet increasingly complex academic demands for specific purposes and audiences. 	<ul style="list-style-type: none"> • Produce, initiate, and sustain extended interactions tailored to specific purposes and audiences. • Write and express ideas to meet a variety of social needs and academic demands for specific purposes and audiences.

Summative ELPAC Alignment Overall Debrief Survey

The questions and response options below represent the content of an online survey HumRRO administered to the alignment workshop panelists on March 26, 2021.

1. Grade level panel
 - Grade K
 - Grade 1
 - Grade 2
 - Grades 3-5
 - Grades 6-8
 - Grades 9-10
 - Grades 11-12
2. Did the items you reviewed generally represent the content in ELD Standards that you expected to be covered?
 - Yes
 - No
3. (If answer to 2. was No) Briefly describe what content seemed underrepresented or overrepresented.
4. Did the items generally reflect the range of English language complexity (as reflected in the ELD Proficiency Levels) that you expected?
 - Yes
 - No
5. (If answer to 4. was No) Briefly describe which (if any) of the levels (Emerging, Expanding, Bridging) you found over- or under-represented.
6. Did the items you reviewed generally allow students to demonstrate skills in listening, speaking, reading, and writing in English?
 - Yes
 - No
7. (If answer to 6. was No) Please explain why you believe the items did not generally allow students to demonstrate performance in English language development.

8. What is your general opinion of the alignment between the Summative ELPAC items you reviewed and the California English Language Development Standards?
 - Excellent
 - Good
 - Limited
 - Weak

9. (If answer to 8. was Weak) Please explain and provide some examples to illustrate where you found weak alignment between the Summative ELPAC items you reviewed and the California ELD Standards.

Evaluation: Virtual Alignment Workshop Training and Procedures

The questions and response options below represent the content of an online survey HumRRO administered to the alignment workshop panelists on March 26, 2021.

1. I received all necessary information prior to the virtual workshop to fully participate as a panelist.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree
2. Please suggest areas for improvement.
3. The training presentation in the large group provided useful information about the Summative ELPAC and HumRRO's alignment method.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree
4. Please suggest areas for improvement.
5. After the additional training in my small group, I felt prepared to review and rate test items.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree

6. Please suggest areas for improvement.
7. HumRRO staff seemed knowledgeable of the Summative ELPAC and alignment steps.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree
8. Please suggest areas for improvement.
9. The Panelist Instruction document was clear, understandable, and useful in performing the alignment steps.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree
10. Please suggest areas for improvement.
11. The Google Sheets file was understandable and relatively easy to use to enter item ratings.
 - Strongly Disagree
 - Disagree
 - Somewhat Disagree
 - Somewhat Agree
 - Agree
 - Strongly Agree
12. Please suggest areas for improvement.

13. The process for reaching consensus ratings was conducted fairly.

- Strongly Disagree
- Disagree
- Somewhat Disagree
- Somewhat Agree
- Agree
- Strongly Agree

14. Please suggest areas for improvement.

15. Additional feedback:

Summative ELPAC Alignment Panelist Demographic Information Survey

The questions and response options below represent the content of an online survey HumRRO administered to the alignment workshop panelists on March 26, 2021.

1. Panelist Name (Optional)
2. Grade level panel
 - Grade K
 - Grade 1
 - Grade 2
 - Grades 3-5
 - Grades 6-8
 - Grades 9-10
 - Grades 11-12
3. Region of residence:
 - Northern California
 - Central California
 - Southern California
4. Do you teach in or work for a School District/local educational district?
 - Yes
 - No
5. District name:
6. Gender:
 - Male
 - Female
 - Non-Binary
 - Prefer not to say
7. Ethnicity:
 - Hispanic or Latino
 - Not Hispanic or Latino

8. Race – select all that apply
 - American Indian or Alaska Native
 - Asian
 - Black or African American
 - Native Hawaiian or Other Pacific Islander
 - White
 - Other
9. Brief Description of Current Professional Role(s)
10. Indicate your total years of experience as a teacher of English learners and the content area(s) taught:
11. List all roles you've held as an EL educator in the last three years:
12. Describe your experience working with the California English Language Development Standards prior to this workshop:
13. Have you taught English learners from diverse socioeconomic and cultural backgrounds?
 - Yes
 - No
14. Please provide information about your experience teaching English learners from diverse socioeconomic and cultural backgrounds:
15. Have you taught English learners with disabilities?
 - Yes
 - No

This page is intentionally blank.

Appendix C: Summative ELPAC Item-person Maps

Table C.1 Kindergarten Item-Person Map, Oral	C-2
Table C.2 Kindergarten Item-Person Map, Written.....	C-4
Table C.3 Grade One Item-Person Map, Oral.....	C-7
Table C.4 Grade One Item-Person Map, Written	C-9
Table C.5 Grade Two Item-Person Map, Oral.....	C-11
Table C.6 Grade Two Item-Person Map, Written	C-13
Table C.7 Grade Three Item-Person Map, Oral	C-15
Table C.8 Grade Three Item-Person Map, Written.....	C-17
Table C.9 Grade Four Item-Person Map, Oral	C-19
Table C.10 Grade Four Item-Person Map, Written.....	C-21
Table C.11 Grade Five Item-Person Map, Oral.....	C-23
Table C.12 Grade Five Item-Person Map, Written	C-25
Table C.13 Grade Six Item-Person Map, Oral.....	C-27
Table C.14 Grade Six Item-Person Map, Written	C-29
Table C.15 Grade Seven Item-Person Map, Oral	C-31
Table C.16 Grade Seven Item-Person Map, Written.....	C-33
Table C.17 Grade Eight Item-Person Map, Oral	C-35
Table C.18 Grade Eight Item-Person Map, Written.....	C-37
Table C.19 Grade Nine-Ten Item-Person Map, Oral.....	C-39
Table C.20 Grade Nine-Ten Item-Person Map, Written	C-41
Table C.21 Grade Eleven-Twelve Item-Person Map, Oral	C-43
Table C.22 Grade Eleven-Twelve Item-Person Map, Written.....	C-46

This page is intentionally blank.

Appendix C: Summative ELPAC Item-person Maps

Tables C.1 through C.22 are called item-person maps and present a comparison of student ability and test item difficulty. The left side of the map shows the distribution of student ability levels, or the *Theta Distribution*. The right side of the map shows the distribution of item difficulty levels. Each table presents a grade or grade span item-person map for the oral language (Listening and Speaking) or written language (Reading and Writing) scores of the Summative ELPAC. Each table spans across two or three pages, but if you were to put the pages together, both the student ability and item difficulty distributions would take on a bell curve shape (oriented vertically).

Both student ability and item difficulty are presented on the same scale, represented by the *Value* column at the center of the map. These values are also referred to as *bins*. The students at the top of the map earned the highest scores (highest ability students), while the items at the top of the map are the most difficult. The students at the bottom of the map earned the lowest scores (lowest ability students), and the items at the bottom of the map are easiest. When students and items are directly opposite each other on the map, the difficulty of the items and the ability of the students are comparable. Students and items are comparable when a student at a given ability level has about a 50 percent probability of correctly answering an item at that level of difficulty.

The last column in Tables C.1 through C.22 presents the theta cuts between Summative ELPAC performance levels. The ELPAC performance levels are Beginning to Develop (Level 1), Somewhat Developed (Level 2), Moderately Developed (Level 3), and Well Developed (Level 4). Performance levels were identified during a standard setting process that was separate from this study.

Table C.1 Kindergarten Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
168	.X	4.0	-	0	
0		3.8	-	0	
0		3.6	-	0	
0		3.4	-	0	
0		3.2	-	0	
0		3.0	-	0	
0		2.8	-	0	
0		2.6	-	0	
52	.	2.4	-	0	
18	.	2.2	-	0	
0		2.0	-	0	
101	.X	1.8	-	0	
29	.	1.6	-	0	
6	.	1.4	-	0	
96	.	1.2	-	0	
171	.X	1.0	-	0	
136	.X	0.8	-	0	
303	.XXX	0.6	-	0	
304	.XXX	0.4	-	0	
363	.XXX	0.2	O	1	
532	.XXXXX	0.0	-	0	
693	.XXXXXX	-0.2	-	0	
989	.XXXXXXXXX	-0.4	OPP	3	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.1 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1304	.XXXXXXXXXXXXXXXXXX	-0.6	-	0	Level 4 (-0.75)
1650	.XXXXXXXXXXXXXXXXXX	-0.8	P	1	
1894	.XXXXXXXXXXXXXXXXXX	-1.0	-	0	
2364	.XXXXXXXXXXXXXXXXXX	-1.2	P	1	
2622	.XXXXXXXXXXXXXXXXXX	-1.4	-	0	
2730	.XXXXXXXXXXXXXXXXXX	-1.6	P	1	
2891	.XXXXXXXXXXXXXXXXXX	-1.8	O	1	Level 3 (-1.96)
2873	.XXXXXXXXXXXXXXXXXX	-2.0	OPP	3	
2579	.XXXXXXXXXXXXXXXXXX	-2.2	O	1	
2480	.XXXXXXXXXXXXXXXXXX	-2.4	PPP	3	
2112	.XXXXXXXXXXXXXXXXXX	-2.6	OOPPP	6	
1584	.XXXXXXXXXXXXXXXXXX	-2.8	OPPP	4	Level 2 (-2.94)
1146	.XXXXXXXXXXXX	-3.0	O	1	
837	.XXXXXXXXXX	-3.2	O	1	
575	.XXXXX	-3.4	OOOOP	5	
461	.XXXX	-3.6	OOOOP	5	
380	.XXX	-3.8	O	1	
262	.XX	-4.0	OP	2	
229	.XX	-4.2	P	1	
159	.X	-4.4	O	1	
111	.X	-4.6	-	0	
71	.	-4.8	-	0	
41	.	-5.0	-	0	
54	.	-5.2	-	0	
314	.XXX	-5.4	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.2 Kindergarten Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1726	.XXXXXXXXXXXXXXXXXX	4.0	-	0	
0		3.8	-	0	
0		3.6	-	0	
0		3.4	-	0	
0		3.2	-	0	
0		3.0	-	0	
0		2.8	-	0	
0		2.6	-	0	
0		2.4	-	0	
0		2.2	-	0	
0		2.0	-	0	
0		1.8	-	0	
0		1.6	-	0	
0		1.4	-	0	
0		1.2	-	0	
0		1.0	-	0	
0		0.8	-	0	
0		0.6	-	0	
0		0.4	-	0	
0		0.2	-	0	
0		0.0	-	0	
0		-0.2	-	0	
0		-0.4	-	0	
517	.XXXXX	-0.6	-	0	
501	.XXXXX	-0.8	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.2 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
184	.X	-1.0	-	0	
218	.XX	-1.2	-	0	Level 4 (-1.26)
428	.XXXX	-1.4	-	0	
578	.XXXXX	-1.6	-	0	
658	.XXXXXX	-1.8	-	0	
1092	.XXXXXXXXXX	-2.0	-	0	
1250	.XXXXXXXXXXXX	-2.2	-	0	
1312	.XXXXXXXXXXXXX	-2.4	O	1	
1646	.XXXXXXXXXXXXXXXX	-2.6	P	1	Level 3 (-2.62)
1764	.XXXXXXXXXXXXXXXXX	-2.8	P	1	
1646	.XXXXXXXXXXXXXXXXX	-3.0	-	0	
1603	.XXXXXXXXXXXXXXXXX	-3.2	OP	2	
1762	.XXXXXXXXXXXXXXXXX	-3.4	OPPP	4	
2020	.XXXXXXXXXXXXXXXXX	-3.6	-	0	
1776	.XXXXXXXXXXXXXXXXX	-3.8	OOPP	5	
1491	.XXXXXXXXXXXXXXXXX	-4.0	OPP	3	Level 2 (-4.27)
1302	.XXXXXXXXXXXXXXXXX	-4.2	OP	2	
1135	.XXXXXXXXXXXXX	-4.4	P	1	
908	.XXXXXXXXXXXXX	-4.6	O	1	
668	.XXXXXX	-4.8	O	1	
464	.XXXX	-5.0	OO	2	
364	.XXX	-5.2	OOP	3	
261	.XX	-5.4	-	0	
196	.X	-5.6	P	1	
109	.X	-5.8	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.2 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
69	.	-6.0	O	1	
56	.	-6.2	-	0	
39	.	-6.4	-	0	
27	.	-6.6	-	0	
25	.	-6.8	-	0	
1	.	-7.0	-	0	
150	.X	-7.2	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.3 Grade One Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
109	.X	4.0	-	0	
0		3.8	-	0	
0		3.6	-	0	
0		3.4	-	0	
0		3.2	-	0	
59	.	3.0	-	0	
0		2.8	-	0	
0		2.6	-	0	
14	.	2.4	-	0	
36	.	2.2	-	0	
51	.	2.0	-	0	
28	.	1.8	-	0	
46	.	1.6	-	0	
63	.	1.4	-	0	
268	.XX	1.2	-	0	
239	.XX	1.0	P	1	
277	.XX	0.8	-	0	
430	.XXXX	0.6	-	0	
683	.XXXXXX	0.4	-	0	
809	.XXXXXXXX	0.2	-	0	
1194	.XXXXXXXXXXXX	0.0	PP	2	
1423	.XXXXXXXXXXXXXXXX	-0.2	-	0	Level 4 (-0.38)
1701	.XXXXXXXXXXXXXXXXXXXX	-0.4	O	1	
2069	.XXXXXXXXXXXXXXXXXXXX	-0.6	P	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.3 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
2303	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.8	-	0	
2414	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.0	P	1	
2394	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.2	-	0	Level 3 (-1.39)
2396	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.4	O	1	
2179	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.6	P	1	
1906	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.8	OOP	3	
1536	.XXXXXXXXXXXXXXXXXXXX	-2.0	-	0	
1213	.XXXXXXXXXXXX	-2.2	OOOP	4	
878	.XXXXXXX	-2.4	OOOOPPPP	8	Level 2 (-2.41)
591	.XXXXX	-2.6	OPPP	4	
407	.XXXX	-2.8	P	1	
263	.XX	-3.0	OP	2	
223	.XX	-3.2	OOOP	4	
158	.X	-3.4	OOO	3	
111	.X	-3.6	OOOP	4	
99	.	-3.8	P	1	
56	.	-4.0	-	0	
45	.	-4.2	O	1	
20	.	-4.4	O	1	
188	.X	-4.6	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.4 Grade One Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
421	.XXXX	4.0	-	0	
0		3.8	-	0	
0		3.6	-	0	
0		3.4	-	0	
0		3.2	-	0	
0		3.0	-	0	
0		2.8	-	0	
0		2.6	-	0	
0		2.4	-	0	
0		2.2	-	0	
0		2.0	-	0	
0		1.8	-	0	
0		1.6	-	0	
0		1.4	-	0	
1	.	1.2	-	0	
113	.X	1.0	-	0	
381	.XXX	0.8	-	0	
88	.	0.6	-	0	
187	.X	0.4	-	0	
416	.XXXX	0.2	-	0	Level 4 (0.20)
426	.XXXX	0.0	-	0	
671	.XXXXXX	-0.2	-	0	
684	.XXXXXX	-0.4	-	0	
939	.XXXXXXXXXX	-0.6	-	0	
1024	.XXXXXXXXXX	-0.8	P	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.4 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1167	.XXXXXXXXXXXX	-1.0	PP	2	
1246	.XXXXXXXXXXXX	-1.2	O	1	Level 3 (-1.37)
1257	.XXXXXXXXXXXX	-1.4	OOP	3	
1441	.XXXXXXXXXXXX	-1.6	OOP	3	
1525	.XXXXXXXXXXXX	-1.8	O	1	
1598	.XXXXXXXXXXXX	-2.0	OO	2	
1514	.XXXXXXXXXXXX	-2.2	OOOP	5	
1409	.XXXXXXXXXXXX	-2.4	OPP	3	Level 2 (-2.53)
1187	.XXXXXXXXXXXX	-2.6	OOPP	4	
905	.XXXXXXXXXX	-2.8	OOOP	5	
717	.XXXXXXX	-3.0	OOOP	4	
577	.XXXXXX	-3.2	OOP	3	
513	.XXXXX	-3.4	P	1	
450	.XXXX	-3.6	OOOOPP	6	
353	.XXX	-3.8	O	1	
243	.XX	-4.0	OP	2	
171	.X	-4.2	P	1	
87	.	-4.4	-	0	
59	.	-4.6	O	1	
78	.	-4.8	O	1	
20	.	-5.0	-	0	
80	.	-5.2	-	0	
5	.	-5.4	-	0	
134	.X	-5.6	PP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.5 Grade Two Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
223	.XX	4.0	O	1	
0		3.8	-	0	
0		3.6	-	0	
25	.	3.4	-	0	
0		3.2	-	0	
21	.	3.0	-	0	
33	.	2.8	-	0	
41	.	2.6	-	0	
36	.	2.4	-	0	
123	.X	2.2	-	0	
74	.	2.0	-	0	
117	.X	1.8	-	0	
410	.XXXX	1.6	-	0	
288	.XX	1.4	-	0	
470	.XXXX	1.2	-	0	
613	.XXXXXX	1.0	P	1	
942	.XXXXXXXXXX	0.8	P	1	
1171	.XXXXXXXXXXXX	0.6	-	0	
1478	.XXXXXXXXXXXXXXXX	0.4	-	0	
1700	XXXXXXXXXXXXXXXXXXXX	0.2	-	0	Level 4 (0.03)
2044	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.0	-	0	
2338	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.2	P	1	
2452	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.4	PP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.5 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
2572	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.6	O	1	
2594	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.8	-	0	
2327	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.0	-	0	Level 3 (-1.18)
2082	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.2	-	0	
1789	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.4	-	0	
1442	.XXXXXXXXXXXXXXXXXXXX	-1.6	OOPP	4	
1123	.XXXXXXXXXXXX	-1.8	OOPPPP	6	
819	.XXXXXXX	-2.0	P	1	
571	.XXXXX	-2.2	OPP	3	Level 2 (-2.28)
412	.XXXX	-2.4	OOOP	4	
275	.XX	-2.6	OOOPPP	6	
204	.XX	-2.8	PPPP	4	
157	.X	-3.0	OPP	3	
117	.X	-3.2	OOO	3	
102	.X	-3.4	OOP	3	
96	.	-3.6	OOOP	4	
67	.	-3.8	OO	2	
267	.XX	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.6 Grade Two Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
225	.XX	4.0	-	0	
0		3.8	-	0	
113	.X	3.6	-	0	
69	.	3.4	-	0	
0		3.2	-	0	
0		3.0	-	0	
0		2.8	-	0	
2	.	2.6	-	0	
39	.	2.4	-	0	
23	.	2.2	-	0	
45	.	2.0	-	0	
235	.XX	1.8	-	0	
113	.X	1.6	-	0	
251	.XX	1.4	-	0	
277	.XX	1.2	-	0	Level 4 (1.07)
424	.XXXX	1.0	O	1	
565	.XXXXX	0.8	-	0	
674	.XXXXXX	0.6	P	1	
866	.XXXXXXXX	0.4	-	0	
1028	.XXXXXXXXXX	0.2	-	0	
1180	.XXXXXXXXXXX	0.0	-	0	
1312	.XXXXXXXXXXXX	-0.2	-	0	
1370	.XXXXXXXXXXXX	-0.4	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.6 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1447	.XXXXXXXXXXXXXXXX	-0.6	PPPP	4	
1450	.XXXXXXXXXXXXXXXX	-0.8	PPP	3	Level 3 (-0.81)
1372	.XXXXXXXXXXXXXXXX	-1.0	O	1	
1201	.XXXXXXXXXXXXXXXX	-1.2	OP	2	
1166	.XXXXXXXXXXXXXXXX	-1.4	OOOOO	5	
1024	.XXXXXXXXXXXX	-1.6	OOOOOOO	7	
1109	.XXXXXXXXXXXXXXXX	-1.8	OOO	3	
984	.XXXXXXXXXX	-2.0	OOOOOOP	7	Level 2 (-2.03)
767	.XXXXXXX	-2.2	OOOPPPP	7	
595	.XXXXX	-2.4	OOPPPP	7	
457	.XXXX	-2.6	OOPP	4	
409	.XXX	-2.8	OOP	3	
298	.XX	-3.0	O	1	
182	.X	-3.2	OP	2	
83	.	-3.4	-	0	
33	.	-3.6	-	0	
25	.	-3.8	-	0	
99	.	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.7 Grade Three Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
87	.	4.0	-	0	
16	.	3.8	-	0	
14	.	3.6	O	1	
23	.	3.4	-	0	
38	.	3.2	-	0	
32	.	3.0	-	0	
59	.	2.8	-	0	
74	.	2.6	-	0	
107	.X	2.4	-	0	
154	.X	2.2	-	0	
166	.X	2.0	-	0	
229	.XX	1.8	-	0	
343	.XXX	1.6	-	0	
468	.XXXX	1.4	-	0	
585	.XXXXX	1.2	OP	2	
887	.XXXXXXXX	1.0	P	1	
1105	.XXXXXXXXXXXX	0.8	-	0	
1391	.XXXXXXXXXXXXXXXX	0.6	O	1	
1803	.XXXXXXXXXXXXXXXXXXXX	0.4	-	0	
2190	.XXXXXXXXXXXXXXXXXXXXXXXX	0.2	O	1	Level 4 (0.09)
2561	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.0	OOP	3	
2852	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.2	OPP	3	
3039	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.4	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.7 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
3012	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.6	P	1	
2870	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.8	-	0	
2616	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.0	PP	2	Level 3 (-1.02)
2211	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.2	OP	2	
1861	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.4	O	1	
1515	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.6	OOOP	4	Level 2 (-1.77)
1114	.XXXXXXXXXXXX	-1.8	PPP	3	
781	.XXXXXXX	-2.0	PPP	3	
585	.XXXXX	-2.2	OP	2	
371	.XXX	-2.4	PPP	3	
239	.XX	-2.6	OOPPP	6	
172	.X	-2.8	OP	2	
115	.X	-3.0	OPPPP	5	
83	.	-3.2	OO	2	
68	.	-3.4	OO	2	
82	.	-3.6	O	1	
82	.	-3.8	-	0	
618	.XXXXXX	-4.0	O	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.8 Grade Three Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
27	.	4.0	-	0	
3	.	3.8	-	0	
4	.	3.6	O	1	
7	.	3.4	O	1	
12	.	3.2	-	0	
13	.	3.0	O	1	
27	.	2.8	-	0	
25	.	2.6	-	0	
43	.	2.4	-	0	
74	.	2.2	-	0	
110	.X	2.0	OO	2	
155	.X	1.8	O	1	
224	.XX	1.6	O	1	
310	.XXX	1.4	P	1	
527	.XXXXX	1.2	P	1	Level 4 (1.16)
711	.XXXXXXXX	1.0	O	1	
1028	.XXXXXXXXXXXX	0.8	O	1	
1260	.XXXXXXXXXXXXXXXX	0.6	-	0	
1650	.XXXXXXXXXXXXXXXXXXXX	0.4	P	1	
2067	.XXXXXXXXXXXXXXXXXXXXXXXX	0.2	OOP	3	
2207	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.0	OOOOOPP	7	Level 3 (-0.07)
2521	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.2	OOO	3	
2680	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.4	OP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.8 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
2891	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.6	OPP	3	
2943	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.8	O	1	
2798	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.0	-	0	
2624	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.2	OP	2	Level 2 (-1.31)
2322	.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.4	O	1	
1889	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-1.6	P	1	
1488	.XXXXXXXXXXXXXXXXXXXX	-1.8	PPP	3	
1225	.XXXXXXXXXXXX	-2.0	OP	2	
903	.XXXXXXXXXX	-2.2	O	1	
660	.XXXXXX	-2.4	P	1	
523	.XXXXX	-2.6	P	1	
315	.XXX	-2.8	-	0	
193	.X	-3.0	-	0	
107	.X	-3.2	-	0	
68	.	-3.4	-	0	
32	.	-3.6	-	0	
22	.	-3.8	-	0	
38	.	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.9 Grade Four Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
153	.X	4.0	-	0	
41	.	3.8	-	0	
39	.	3.6	O	1	
40	.	3.4	-	0	
65	.	3.2	-	0	
68	.	3.0	-	0	
124	.X	2.8	-	0	
130	.X	2.6	-	0	
147	.X	2.4	-	0	
215	.XX	2.2	-	0	
317	.XXX	2.0	-	0	
395	.XXX	1.8	-	0	
529	.XXXXX	1.6	-	0	
686	.XXXXXX	1.4	-	0	
898	.XXXXXXXX	1.2	OP	2	
1127	.XXXXXXXXXXX	1.0	P	1	
1390	.XXXXXXXXXXXX	0.8	-	0	
1669	.XXXXXXXXXXXXXXXX	0.6	O	1	
1937	.XXXXXXXXXXXXXXXXXXXX	0.4	-	0	Level 4 (0.32)
2245	.XXXXXXXXXXXXXXXXXXXXXXXX	0.2	O	1	
2242	.XXXXXXXXXXXXXXXXXXXXXXXX	0.0	OOP	3	
2348	.XXXXXXXXXXXXXXXXXXXXXXXX	-0.2	OPP	3	
2297	.XXXXXXXXXXXXXXXXXXXXXXXX	-0.4	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.9 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1989	.XXXXXXXXXXXXXXXXXXXXX	-0.6	P	1	
1798	.XXXXXXXXXXXXXXXXXXXXX	-0.8	-	0	Level 3 (-0.89)
1380	.XXXXXXXXXXXXXXXXXXXXX	-1.0	PP	2	
1186	.XXXXXXXXXXXXXXXXXXXXX	-1.2	OP	2	
884	.XXXXXXXXXXXX	-1.4	O	1	
652	.XXXXXX	-1.6	OOOP	4	Level 2 (-1.67)
467	.XXXX	-1.8	PPP	3	
354	.XXX	-2.0	PPP	3	
256	.XX	-2.2	OP	2	
187	.X	-2.4	PPP	3	
130	.X	-2.6	OOOPPP	6	
113	.X	-2.8	OP	2	
80	.	-3.0	OPPPP	5	
46	.	-3.2	OO	2	
55	.	-3.4	OO	2	
63	.	-3.6	O	1	
61	.	-3.8	-	0	
474	.XXXX	-4.0	O	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.10 Grade Four Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
42	.	4.0	-	0	
7	.	3.8	-	0	
11	.	3.6	O	1	
25	.	3.4	O	1	
32	.	3.2	-	0	
38	.	3.0	O	1	
59	.	2.8	-	0	
69	.	2.6	-	0	
115	.X	2.4	-	0	
155	.X	2.2	-	0	
261	.XX	2.0	OO	2	
365	.XXX	1.8	O	1	Level 4 (1.62)
546	.XXXXX	1.6	O	1	
782	.XXXXXXXX	1.4	P	1	
1023	.XXXXXXXXXXXX	1.2	P	1	
1202	.XXXXXXXXXXXXXX	1.0	O	1	
1534	.XXXXXXXXXXXXXXXX	0.8	O	1	
1715	.XXXXXXXXXXXXXXXXXX	0.6	-	0	
1963	.XXXXXXXXXXXXXXXXXXXX	0.4	P	1	Level 3 (0.32)
2066	.XXXXXXXXXXXXXXXXXXXXXX	0.2	OOP	3	
2063	.XXXXXXXXXXXXXXXXXXXXXX	0.0	OOOOOPP	7	
2049	.XXXXXXXXXXXXXXXXXXXXXX	-0.2	OOO	3	
2004	.XXXXXXXXXXXXXXXXXXXXXX	-0.4	OP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.10 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1850	.XXXXXXXXXXXXXXXXXXXX	-0.6	OPP	3	
1794	.XXXXXXXXXXXXXXXXXXXX	-0.8	O	1	Level 2 (-0.87)
1544	.XXXXXXXXXXXXXXXXXXXX	-1.0	-	0	
1203	.XXXXXXXXXXXXXXXX	-1.2	OP	2	
1029	.XXXXXXXXXXXX	-1.4	O	1	
810	.XXXXXXXXXX	-1.6	P	1	
653	.XXXXXX	-1.8	PPP	3	
489	.XXXX	-2.0	OP	2	
394	.XXX	-2.2	O	1	
293	.XX	-2.4	P	1	
229	.XX	-2.6	P	1	
168	.X	-2.8	-	0	
96	.	-3.0	-	0	
63	.	-3.2	-	0	
30	.	-3.4	-	0	
14	.	-3.6	-	0	
5	.	-3.8	-	0	
13	.	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.11 Grade Five Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
336	.XXX	4.0	-	0	
81	.	3.8	-	0	
84	.	3.6	O	1	
78	.	3.4	-	0	
123	.X	3.2	-	0	
152	.X	3.0	-	0	
202	.XX	2.8	-	0	
260	.XX	2.6	-	0	
297	.XX	2.4	-	0	
324	.XXX	2.2	-	0	
511	.XXXXX	2.0	-	0	
639	.XXXXXX	1.8	-	0	
806	.XXXXXXXX	1.6	-	0	
1022	.XXXXXXXXXX	1.4	-	0	
1312	.XXXXXXXXXXXXXX	1.2	OP	2	
1481	.XXXXXXXXXXXXXXXX	1.0	P	1	
1787	.XXXXXXXXXXXXXXXXXXXX	0.8	-	0	
2075	.XXXXXXXXXXXXXXXXXXXXXXXX	0.6	O	1	Level 4 (0.58)
2205	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.4	-	0	
2164	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.2	O	1	
2306	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	0.0	OOP	3	
2122	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.2	OPP	3	
1902	.XXXXXXXXXXXXXXXXXXXXXXXXXXXX	-0.4	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.11 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1623	.XXXXXXXXXXXXXXXXXX	-0.6	P	1	Level 3 (-0.76)
1343	.XXXXXXXXXXXXXXXXXX	-0.8	-	0	
1045	.XXXXXXXXXXXX	-1.0	PP	2	
846	.XXXXXXXXXX	-1.2	OP	2	
561	.XXXXX	-1.4	O	1	Level 2 (-1.47)
436	.XXXX	-1.6	OOOP	4	
319	.XXX	-1.8	PPP	3	
231	.XX	-2.0	PPP	3	
166	.X	-2.2	OP	2	
133	.X	-2.4	PPP	3	
92	.	-2.6	OOOPPP	6	
85	.	-2.8	OP	2	
70	.	-3.0	OPPPP	5	
52	.	-3.2	OO	2	
52	.	-3.4	OO	2	
67	.	-3.6	O	1	
84	.	-3.8	-	0	
475	.XXXX	-4.0	O	1	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.12 Grade Five Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
156	.X	4.0	-	0	
26	.	3.8	-	0	
40	.	3.6	O	1	
76	.	3.4	O	1	
78	.	3.2	-	0	
140	.X	3.0	O	1	
146	.X	2.8	-	0	
254	.XX	2.6	-	0	
338	.XXX	2.4	-	0	
440	.XXXX	2.2	-	0	
661	.XXXXXX	2.0	OO	2	Level 4 (1.91)
871	.XXXXXXXX	1.8	O	1	
1222	.XXXXXXXXXXXX	1.6	O	1	
1444	.XXXXXXXXXXXX	1.4	P	1	
1755	.XXXXXXXXXXXX	1.2	P	1	
1914	.XXXXXXXXXXXX	1.0	O	1	Level 3 (0.96)
2114	.XXXXXXXXXXXX	0.8	O	1	
2134	.XXXXXXXXXXXX	0.6	-	0	
2100	XXXXXXXXXXXX	0.4	P	1	
1977	.XXXXXXXXXXXX	0.2	OOP	3	
1885	.XXXXXXXXXXXX	0.0	OOOOOPP	7	
1590	.XXXXXXXXXXXX	-0.2	OOO	3	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.12 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1469	.XXXXXXXXXXXXXXXX	-0.4	OP	2	
1339	.XXXXXXXXXXXXXXXX	-0.6	OPP	3	Level 2 (-0.65)
1092	.XXXXXXXXXX	-0.8	O	1	
908	.XXXXXXXXXX	-1.0	-	0	
768	.XXXXXXX	-1.2	OP	2	
631	.XXXXXX	-1.4	O	1	
512	.XXXXX	-1.6	P	1	
384	.XXX	-1.8	PPP	3	
321	.XXX	-2.0	OP	2	
266	.XX	-2.2	O	1	
221	.XX	-2.4	P	1	
162	.X	-2.6	P	1	
108	.X	-2.8	-	0	
62	.	-3.0	-	0	
40	.	-3.2	-	0	
19	.	-3.4	-	0	
14	.	-3.6	-	0	
5	.	-3.8	-	0	
8	.	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.13 Grade Six Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
471	.XXXX	4.0	O	1	
48	.	3.8	-	0	
94	.	3.6	-	0	
114	.X	3.4	-	0	
126	.X	3.2	-	0	
180	.X	3.0	-	0	
219	.XX	2.8	-	0	
298	.XX	2.6	-	0	
379	.XXX	2.4	-	0	
416	.XXXX	2.2	-	0	
578	.XXXXX	2.0	-	0	
691	.XXXXXX	1.8	-	0	
961	.XXXXXXXX	1.6	-	0	
1097	.XXXXXXXXXX	1.4	OP	2	
1307	.XXXXXXXXXXXX	1.2	OPP	3	
1396	.XXXXXXXXXXXX	1.0	-	0	Level 4 (0.81)
1557	.XXXXXXXXXXXX	0.8	O	1	
1741	.XXXXXXXXXXXX	0.6	-	0	
1740	.XXXXXXXXXXXX	0.4	P	1	
1758	.XXXXXXXXXXXX	0.2	P	1	
1740	.XXXXXXXXXXXX	0.0	P	1	
1571	.XXXXXXXXXXXX	-0.2	P	1	
1363	.XXXXXXXXXXXX	-0.4	OO	2	Level 3 (-0.59)

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.13 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1198	.XXXXXXXXXXXX	-0.6	OOP	3	
969	.XXXXXXXXXXXX	-0.8	PP	2	
697	.XXXXXX	-1.0	OO	2	
614	.XXXXXX	-1.2	P	1	
450	.XXXX	-1.4	OP	2	Level 2 (-1.40)
346	.XXX	-1.6	P	1	
260	.XX	-1.8	PP	2	
209	.XX	-2.0	PPPP	4	
164	.X	-2.2	PPP	3	
102	.X	-2.4	PPP	3	
111	.X	-2.6	OOP	3	
81	.	-2.8	OO	2	
79	.	-3.0	O	1	
65	.	-3.2	OOOP	5	
68	.	-3.4	-	0	
44	.	-3.6	OO	2	
51	.	-3.8	-	0	
394	.XXX	-4.0	OOO	3	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.14 Grade Six Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
52	.	4.0	OO	2	
10	.	3.8	-	0	
24	.	3.6	OO	2	
19	.	3.4	-	0	
69	.	3.2	O	1	
103	.X	3.0	-	0	
140	.X	2.8	OO	2	
176	.X	2.6	-	0	
241	.XX	2.4	O	1	
319	.XXX	2.2	-	0	
448	.XXXX	2.0	OOP	3	Level 4 (2.05)
643	.XXXXXX	1.8	O	1	
829	.XXXXXXXX	1.6	OP	2	
1096	.XXXXXXXXXX	1.4	OOO	3	
1252	.XXXXXXXXXXXX	1.2	OP	2	
1524	.XXXXXXXXXXXXXXXX	1.0	-	0	Level 3 (0.97)
1671	.XXXXXXXXXXXXXXXXXXXX	0.8	O	1	
1801	.XXXXXXXXXXXXXXXXXXXX	0.6	O	1	
1931	.XXXXXXXXXXXXXXXXXXXX	0.4	OOOOP	5	
1954	.XXXXXXXXXXXXXXXXXXXX	0.2	P	1	
1782	.XXXXXXXXXXXXXXXXXXXX	0.0	P	1	
1664	.XXXXXXXXXXXXXXXXXXXX	-0.2	O	1	Level 2 (-0.33)
1446	.XXXXXXXXXXXXXXXXXXXX	-0.4	O	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.14 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1291	.XXXXXXXXXXXXX	-0.6	OPPP	4	
989	.XXXXXXXXXX	-0.8	P	1	
885	.XXXXXXXX	-1.0	O	1	
652	.XXXXXX	-1.2	P	1	
518	.XXXXX	-1.4	PP	2	
374	.XXX	-1.6	P	1	
328	.XXX	-1.8	P	1	
222	.XX	-2.0	P	1	
207	.XX	-2.2	-	0	
168	.X	-2.4	P	1	
114	.X	-2.6	-	0	
83	.	-2.8	-	0	
55	.	-3.0	-	0	
33	.	-3.2	-	0	
32	.	-3.4	-	0	
14	.	-3.6	-	0	
7	.	-3.8	-	0	
22	.	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.15 Grade Seven Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
649	.XXXXXX	4.0	O	1	
81	.	3.8	-	0	
107	.X	3.6	-	0	
123	.X	3.4	-	0	
158	.X	3.2	-	0	
179	.X	3.0	-	0	
294	.XX	2.8	-	0	
317	.XXX	2.6	-	0	
444	.XXXX	2.4	-	0	
497	.XXXX	2.2	-	0	
646	.XXXXXX	2.0	-	0	
795	.XXXXXXXX	1.8	-	0	
929	.XXXXXXXXXX	1.6	-	0	
1050	.XXXXXXXXXXXX	1.4	OP	2	
1239	.XXXXXXXXXXXXXX	1.2	OPP	3	Level 4 (1.10)
1361	.XXXXXXXXXXXXXXX	1.0	-	0	
1486	.XXXXXXXXXXXXXXX	0.8	O	1	
1404	.XXXXXXXXXXXXXXX	0.6	-	0	
1529	.XXXXXXXXXXXXXXX	0.4	P	1	
1478	.XXXXXXXXXXXXXXX	0.2	P	1	
1355	.XXXXXXXXXXXXXXX	0.0	P	1	
1113	.XXXXXXXXXXXXXXX	-0.2	P	1	Level 3 (-0.26)
966	.XXXXXXXXXX	-0.4	OO	2	
882	.XXXXXXXXXX	-0.6	OOP	3	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.15 (cont.)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
696	.XXXXXX	-0.8	PP	2	
564	.XXXXX	-1.0	OO	2	
446	.XXXX	-1.2	P	1	Level 2 (-1.28)
378	.XXX	-1.4	OP	2	
274	.XX	-1.6	P	1	
228	.XX	-1.8	PP	2	
202	.XX	-2.0	PPPP	4	
128	.X	-2.2	PPP	3	
121	.X	-2.4	PPP	3	
101	.X	-2.6	OOP	3	
97	.	-2.8	OO	2	
73	.	-3.0	O	1	
64	.	-3.2	OOOPP	5	
62	.	-3.4	-	0	
66	.	-3.6	OO	2	
52	.	-3.8	-	0	
394	.XXX	-4.0	OOO	3	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.16 Grade Seven Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
83	.	4.0	OO	2	
32	.	3.8	-	0	
41	.	3.6	OO	2	
66	.	3.4	-	0	
118	.X	3.2	O	1	
151	.X	3.0	-	0	
224	.XX	2.8	OO	2	
242	.XX	2.6	-	0	
333	.XXX	2.4	O	1	
509	.XXXXX	2.2	-	0	Level 4 (2.20)
624	.XXXXXX	2.0	OOP	3	
741	.XXXXXXX	1.8	O	1	
868	.XXXXXXXX	1.6	OP	2	
1114	.XXXXXXXXXXXX	1.4	OOO	3	
1175	.XXXXXXXXXXXX	1.2	OP	2	Level 3 (1.13)
1326	.XXXXXXXXXXXXXX	1.0	-	0	
1451	.XXXXXXXXXXXXXX	0.8	O	1	
1491	.XXXXXXXXXXXXXX	0.6	O	1	
1511	.XXXXXXXXXXXXXX	0.4	OOOOP	5	
1484	.XXXXXXXXXXXXXX	0.2	P	1	
1306	.XXXXXXXXXXXXXX	0.0	P	1	Level 2 (-0.19)
1244	.XXXXXXXXXXXXXX	-0.2	O	1	
1023	.XXXXXXXXXXXX	-0.4	O	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.16 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
901	.XXXXXXXXXX	-0.6	OPPP	4	
761	.XXXXXXX	-0.8	P	1	
560	.XXXXX	-1.0	O	1	
494	.XXXX	-1.2	P	1	
457	.XXXX	-1.4	PP	2	
354	.XXX	-1.6	P	1	
248	.XX	-1.8	P	1	
206	.XX	-2.0	P	1	
160	.X	-2.2	-	0	
128	.X	-2.4	P	1	
104	.X	-2.6	-	0	
80	.	-2.8	-	0	
49	.	-3.0	-	0	
31	.	-3.2	-	0	
23	.	-3.4	-	0	
16	.	-3.6	-	0	
10	.	-3.8	-	0	
20	.	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.17 Grade Eight Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
800	XXXXXXXXX	4.0	O	1	
86	.	3.8	-	0	
102	.X	3.6	-	0	
162	.X	3.4	-	0	
156	.X	3.2	-	0	
222	.XX	3.0	-	0	
267	.XX	2.8	-	0	
339	.XXX	2.6	-	0	
467	.XXXX	2.4	-	0	
449	.XXXX	2.2	-	0	
607	.XXXXXX	2.0	-	0	
741	.XXXXXX	1.8	-	0	
894	.XXXXXXXXX	1.6	-	0	Level 4 (1.44)
930	.XXXXXXXXXX	1.4	OP	2	
1023	.XXXXXXXXXX	1.2	OPP	3	
1139	.XXXXXXXXXX	1.0	-	0	
1183	.XXXXXXXXXX	0.8	O	1	
1214	.XXXXXXXXXX	0.6	-	0	
1147	.XXXXXXXXXX	0.4	P	1	
1078	.XXXXXXXXXX	0.2	P	1	
971	.XXXXXXXXXX	0.0	P	1	Level 3 (-0.08)
825	.XXXXXXXXX	-0.2	P	1	
748	.XXXXXXX	-0.4	OO	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.17 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
574	.XXXXX	-0.6	OOP	3	
516	.XXXXX	-0.8	PP	2	
382	.XXX	-1.0	OO	2	Level 2 (-1.15)
350	.XXX	-1.2	P	1	
266	.XX	-1.4	OP	2	
224	.XX	-1.6	P	1	
148	.X	-1.8	PP	2	
132	.X	-2.0	PPPP	4	
123	.X	-2.2	PPP	3	
99	.	-2.4	PPP	3	
97	.	-2.6	OOP	3	
75	.	-2.8	OO	2	
67	.	-3.0	O	1	
56	.	-3.2	OOOPP	5	
47	.	-3.4	-	0	
36	.	-3.6	OO	2	
39	.	-3.8	-	0	
340	.XXX	-4.0	OOO	3	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.18 Grade Eight Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
155	.X	4.0	OO	2	
51	.	3.8	-	0	
52	.	3.6	OO	2	
110	.X	3.4	-	0	
158	.X	3.2	O	1	
238	.XX	3.0	-	0	
286	.XX	2.8	OO	2	
373	.XXX	2.6	-	0	Level 4 (2.53)
453	.XXXX	2.4	O	1	
578	.XXXXX	2.2	-	0	
660	.XXXXXX	2.0	OOP	3	
795	.XXXXXXX	1.8	O	1	
974	.XXXXXXXXX	1.6	OP	2	
1123	.XXXXXXXXXXX	1.4	OOO	3	Level 3 (1.29)
1095	.XXXXXXXXXXX	1.2	OP	2	
1239	.XXXXXXXXXXX	1.0	-	0	
1207	.XXXXXXXXXXX	0.8	O	1	
1213	.XXXXXXXXXXX	0.6	O	1	
1167	.XXXXXXXXXXX	0.4	OOOOP	5	
1087	.XXXXXXXXXXX	0.2	P	1	
956	.XXXXXXXXXXX	0.0	P	1	Level 2 (-0.05)
850	.XXXXXXXXXXX	-0.2	O	1	
778	.XXXXXXXXXXX	-0.4	O	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.18 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
642	.XXXXXX	-0.6	OPPP	4	
544	.XXXXX	-0.8	P	1	
428	.XXXX	-1.0	O	1	
354	.XXX	-1.2	P	1	
315	.XXX	-1.4	PP	2	
251	.XX	-1.6	P	1	
181	.X	-1.8	P	1	
145	.X	-2.0	P	1	
117	.X	-2.2	-	0	
99	.	-2.4	P	1	
74	.	-2.6	-	0	
42	.	-2.8	-	0	
36	.	-3.0	-	0	
32	.	-3.2	-	0	
20	.	-3.4	-	0	
14	.	-3.6	-	0	
10	.	-3.8	-	0	
9	.	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.19 Grade Nine-Ten Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
943	.XXXXXXXXXX	4.0	OO	2	
160	.X	3.8	-	0	
160	.X	3.6	-	0	
236	.XX	3.4	-	0	
279	.XX	3.2	-	0	
350	.XXX	3.0	-	0	
386	.XXX	2.8	-	0	
508	.XXXXX	2.6	P	1	
572	.XXXXX	2.4	-	0	
751	.XXXXXXX	2.2	-	0	
846	.XXXXXXX	2.0	-	0	
1004	.XXXXXXXXXX	1.8	P	1	Level 4 (1.68)
1216	.XXXXXXXXXXXXXX	1.6	-	0	
1304	.XXXXXXXXXXXXXX	1.4	-	0	
1426	.XXXXXXXXXXXXXX	1.2	-	0	
1495	.XXXXXXXXXXXXXX	1.0	P	1	
1569	.XXXXXXXXXXXXXX	0.8	-	0	
1621	.XXXXXXXXXXXXXX	0.6	O	1	
1662	.XXXXXXXXXXXXXX	0.4	OP	2	
1630	.XXXXXXXXXXXXXX	0.2	OO	2	Level 3 (0.08)
1469	.XXXXXXXXXXXXXX	0.0	OPPPP	5	
1360	.XXXXXXXXXXXXXX	-0.2	OOP	3	
1164	.XXXXXXXXXXXXX	-0.4	P	1	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.19 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1009	.XXXXXXXXXX	-0.6	-	0	
866	.XXXXXXXX	-0.8	P	1	
695	.XXXXXX	-1.0	-	0	Level 2 (-1.05)
615	.XXXXXX	-1.2	OOP	3	
491	.XXXX	-1.4	PP	2	
461	.XXXX	-1.6	PPP	3	
349	.XXX	-1.8	O	1	
279	.XX	-2.0	PPP	3	
259	.XX	-2.2	OPPP	4	
206	.XX	-2.4	OPPPP	5	
185	.X	-2.6	-	0	
163	.X	-2.8	P	1	
173	.X	-3.0	O	1	
118	.X	-3.2	OO	2	
137	.X	-3.4	OO	2	
137	.X	-3.6	O	1	
135	.X	-3.8	OO	2	
1486	.XXXXXXXXXXXXXXXX	-4.0	OO	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.20 Grade Nine-Ten Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
488	.XXXX	4.0	O	1	
170	.X	3.8	-	0	
239	.XX	3.6	-	0	
292	.XX	3.4	-	0	
390	.XXX	3.2	-	0	Level 4 (3.09)
508	.XXXXX	3.0	O	1	
624	.XXXXXX	2.8	-	0	
756	.XXXXXXX	2.6	-	0	
872	.XXXXXXXX	2.4	P	1	
1069	.XXXXXXXXXX	2.2	-	0	
1187	.XXXXXXXXXX	2.0	P	1	
1274	.XXXXXXXXXX	1.8	OOP	3	Level 3 (1.70)
1369	.XXXXXXXXXX	1.6	O	1	
1396	.XXXXXXXXXX	1.4	O	1	
1594	.XXXXXXXXXX	1.2	OOOO	4	
1557	.XXXXXXXXXX	1.0	OO	2	
1561	.XXXXXXXXXX	0.8	OOP	3	
1535	.XXXXXXXXXX	0.6	OOP	3	
1614	.XXXXXXXXXX	0.4	OO	2	
1613	.XXXXXXXXXX	0.2	O	1	Level 2 (0.20)
1495	.XXXXXXXXXX	0.0	OOPP	4	
1437	.XXXXXXXXXX	-0.2	OPP	3	
1252	.XXXXXXXXXX	-0.4	OOOO	4	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.20 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1135	.XXXXXXXXXXXX	-0.6	-	0	
976	.XXXXXXXXXXXX	-0.8	P	1	
814	.XXXXXXXXXXXX	-1.0	PP	2	
708	.XXXXXXX	-1.2	-	0	
575	.XXXXXX	-1.4	P	1	
470	.XXXX	-1.6	P	1	
381	.XXX	-1.8	-	0	
279	.XX	-2.0	-	0	
176	.X	-2.2	PP	2	
142	.X	-2.4	-	0	
83	.	-2.6	P	1	
46	.	-2.8	-	0	
37	.	-3.0	-	0	
23	.	-3.2	-	0	
8	.	-3.4	-	0	
1	.	-3.6	-	0	
4	.	-3.8	-	0	
8	.	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 100 students and “.” represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, “O” represents one dichotomous item, “P” represents one category threshold of one polytomous item, and no items are denoted as “-”.

Table C.21 Grade Eleven-Twelve Item-Person Map, Oral

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
149	.X	7.4	-	0	
0		7.2	-	0	
25	.	7.0	-	0	
21	.	6.8	-	0	
4	.	6.6	-	0	
22	.	6.4	-	0	
57	.	6.2	-	0	
5	.	6.0	-	0	
36	.	5.8	-	0	
50	.	5.6	-	0	
31	.	5.4	-	0	
29	.	5.2	-	0	
42	.	5.0	-	0	
57	.	4.8	-	0	
54	.	4.6	-	0	
107	.X	4.4	-	0	
131	.X	4.2	-	0	
122	.X	4.0	-	0	
163	.X	3.8	-	0	
195	.X	3.6	-	0	
219	.XX	3.4	-	0	
280	.XX	3.2	-	0	
358	.XXX	3.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.21 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
390	.XXX	2.8	-	0	
464	.XXXX	2.6	-	0	
576	.XXXXX	2.4	-	0	
680	.XXXXXX	2.2	P	1	
740	.XXXXXXX	2.0	P	1	
933	.XXXXXXXXX	1.8	-	0	Level 4 (1.79)
981	.XXXXXXXXX	1.6	-	0	
1054	.XXXXXXXXX	1.4	-	0	
1183	.XXXXXXXXX	1.2	-	0	
1261	.XXXXXXXXX	1.0	O	1	
1316	.XXXXXXXXX	0.8	-	0	
1344	.XXXXXXXXX	0.6	P	1	
1197	.XXXXXXXXX	0.4	P	1	
1195	.XXXXXXXXX	0.2	OP	2	Level 3 (0.13)
1128	.XXXXXXXXX	0.0	P	1	
1073	.XXXXXXXXX	-0.2	OOPPP	5	
952	.XXXXXXXXX	-0.4	OOO	3	
794	.XXXXXXX	-0.6	O	1	
713	.XXXXXXX	-0.8	-	0	Level 2 (-0.92)
603	.XXXXXX	-1.0	OP	2	
551	.XXXXX	-1.2	OOP	3	
436	.XXXX	-1.4	OOOP	4	
385	.XXX	-1.6	OOPP	4	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.21 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
322	.XXX	-1.8	OPPP	4	
282	.XX	-2.0	PP	2	
221	.XX	-2.2	OOPPPP	6	
206	.XX	-2.4	O	1	
180	.X	-2.6	OPPP	4	
156	.X	-2.8	OP	2	
127	.X	-3.0	-	0	
128	.X	-3.2	OP	2	
111	.X	-3.4	-	0	
101	.X	-3.6	O	1	
100	X	-3.8	-	0	
96	.	-4.0	-	0	
127	.X	-4.2	-	0	
111	.X	-4.4	-	0	
106	.X	-4.6	-	0	
86	.	-4.8	-	0	
65	.	-5.0	-	0	
35	.	-5.2	-	0	
14	.	-5.4	-	0	
13	.	-5.6	-	0	
10	.	-5.8	-	0	
5	.	-6.0	-	0	
16	.	-6.2	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.22 Grade Eleven-Twelve Item-Person Map, Written

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
98	.	6.0	-	0	
8	.	5.8	-	0	
28	.	5.6	-	0	
24	.	5.4	-	0	
29	.	5.2	O	1	
38	.	5.0	-	0	
51	.	4.8	-	0	
67	.	4.6	-	0	
100	X	4.4	-	0	
108	.X	4.2	-	0	
140	.X	4.0	-	0	
194	.X	3.8	OO	2	
238	.XX	3.6	-	0	Level 4 (3.44)
291	.XX	3.4	-	0	
396	.XXX	3.2	O	1	
465	.XXXX	3.0	-	0	
579	.XXXXX	2.8	P	1	
709	.XXXXXXX	2.6	-	0	
769	.XXXXXXX	2.4	-	0	
835	.XXXXXXX	2.2	-	0	Level 3 (2.13)
953	.XXXXXXXXX	2.0	OP	2	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.22 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
1096	.XXXXXXXXXX	1.8	P	1	
1169	.XXXXXXXXXX	1.6	-	0	
1218	.XXXXXXXXXX	1.4	OOO	3	
1180	.XXXXXXXXXX	1.2	OOO	3	
1279	.XXXXXXXXXX	1.0	OPP	3	
1328	.XXXXXXXXXX	0.8	P	1	
1320	.XXXXXXXXXX	0.6	OOOOO	5	Level 2 (0.45)
1333	.XXXXXXXXXX	0.4	OOOPP	5	
1216	.XXXXXXXXXX	0.2	-	0	
1188	.XXXXXXXXXX	0.0	OOP	3	
1048	.XXXXXXXXXX	-0.2	O	1	
939	.XXXXXXXXXX	-0.4	-	0	
886	.XXXXXXX	-0.6	OOP	3	
724	.XXXXXXX	-0.8	-	0	
622	.XXXXXX	-1.0	P	1	
485	.XXXX	-1.2	P	1	
376	.XXX	-1.4	O	1	
302	.XXX	-1.6	PP	2	
217	.XX	-1.8	P	1	
147	.X	-2.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".

Table C.22 (cont'd)

Number of Students	Theta Distribution*	Value	Item Difficulty Distribution**	Number of Items	ELPAC Level Theta Cut
106	.X	-2.2	-	0	
55	.	-2.4	-	0	
48	.	-2.6	P	1	
31	.	-2.8	-	0	
13	.	-3.0	P	1	
17	.	-3.2	-	0	
5	.	-3.4	-	0	
2	.	-3.6	-	0	
3	.	-3.8	-	0	
6	.	-4.0	-	0	

*For each bin in the theta distribution column, "X" represents 100 students and "." represents a value between 1 and 99 students.

**For each bin in the item difficulty distribution column, "O" represents one dichotomous item, "P" represents one category threshold of one polytomous item, and no items are denoted as "-".