



2023 No. 010

Summative Alternate 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: Emily Dickinson
Michele Hardoin
Art Thacker

Prepared under: CN180100.3

Date: February 14, 2023

Summative Alternate 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-3
Review of Summative Alternate ELPAC Documentation.....	ES-3
Summative Alternate ELPAC Alignment Workshop and Outcomes	ES-4
Conclusions	ES-8
Recommendations.....	ES-9
Chapter 1: Introduction.....	1-1
Background.....	1-1
Research Questions	1-4
Organization and Contents of the Alignment Study Report	1-4
Chapter 2: Review of Summative Alternate ELPAC Documentation.....	2-7
Introduction	2-7
Method	2-7
Results	2-8
Summary and Discussion	2-21
Chapter 3: Summative Alternate ELPAC Alignment Workshop and Outcomes	3-23
Introduction	3-23
Summative Alternate ELPAC Alignment Criteria	3-23
Methods	3-27
Results	3-34
Summary and Discussion	3-44
Chapter 4: Conclusions and Recommendations	4-47
References.....	49
Glossary of Acronyms.....	51
Appendix A: Summative Alternate ELPAC Documentation Reviewed by HumRRO	A-1
Appendix B: Materials for Virtual Alignment Workshop	B-1
Appendix C: Summary of Responses to Summative Alternate ELPAC Process Evaluation Survey Questions	C-1
Appendix D: Summative Alternate ELPAC Item-person Maps.....	D-1

List of Tables

Table ES.1. Summative Alternate ELPAC Alignment Results.....	ES-7
Table 2.1 Rating Scale for Evaluating Strength of Evidence for Testing Standards.....	2-8
Table 2.2 Ratings on the Testing Standards for Summative Alternate ELPAC Alignment	2-9
Table 3.1 Summative Alternate ELPAC to ELD Connectors Alignment Criteria.....	3-24
Table 3.2 Organization of ELD Connectors.....	3-26
Table 3.3 Characteristics of Summative Alternate ELPAC Alignment Panelists	3-28
Table 3.4 Panelists' Experience Teaching Students with Disabilities, by Grade Level/Grade Span.....	3-30
Table 3.5 Panelists' Ratings of Overall Quality of Alignment, by Grade Level/Grade Span.....	3-33
Table 3.6 Results for Link to Standards	3-34
Table 3.7 Results for Linguistic Complexity Adequacy.....	3-35
Table 3.8 Summative Alternate ELPAC Grade Two Item-Person Map	3-37
Table 3.9 Results for Range Adequacy Criterion: Parts.....	3-39
Table 3.10 Results for Range Adequacy Criterion: Modes.....	3-39
Table 3.11 Results for Range Adequacy Criterion: Language Processes.....	3-39
Table 3.12 Summary of Results for Range Adequacy Criterion	3-40
Table 3.13 Results for Balance of Knowledge Correspondence: Kindergarten.....	3-41
Table 3.14 Results for Balance of Knowledge Correspondence: Grade One.....	3-41
Table 3.15 Results for Balance of Knowledge Correspondence: Grade Two.....	3-41
Table 3.16 Results for Balance of Knowledge Correspondence: Grades Three through Five Form 1	3-41
Table 3.17 Results for Balance of Knowledge Correspondence: Grades Three through Five Form 2	3-41
Table 3.18 Results for Balance of Knowledge Correspondence: Grades Six through Eight Form 1	3-42
Table 3.19 Results for Balance of Knowledge Correspondence: Grades Six through Eight Form 2.....	3-42
Table 3.20 Results for Balance-of-Knowledge Correspondence: Grades Nine through Twelve Form 1	3-42
Table 3.21 Results for Balance-of-Knowledge Correspondence: Grades Nine through Twelve Form 2	3-42
Table 3.22 Summary of Results for Balance-of-Knowledge Correspondence Criterion.....	3-43
Table 3.23 Summative Alternate ELPAC Alignment Results.....	3-44

Executive Summary

This report summarizes a study of the alignment between the Summative Alternate English Language Proficiency Assessments for California (hereinafter referred to as the Summative Alternate ELPAC) and the California English Language Development (ELD) Connectors (CDE, 2021), which were derived from the 2012 ELD Standards (CDE, 2012). 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

English Learner (ELs) and students with disabilities have been attended to historically as two different student groups in the public school system. Nationwide, in the fall of 2019, the percentage of U.S. public elementary and secondary school students classified as ELs was 10.4 percent, or 5.1 million students (NCES, 2022). That same year, the percentage of U.S. public school students that were receiving special education services was 14 percent, or 7.1 million students (USDOE, 2021a). During the 2018–19 school year, in California public schools specifically, there were approximately 1.195 million ELs (nearly a quarter of the national population of ELs) and over 720,000 students with disabilities enrolled in kindergarten through grade twelve (CDE, 2023a). Intersecting these two separate groups is a group of students that is gaining attention: EL students with disabilities (ELSWDs). Across the U.S., ELs with significant cognitive disabilities may be a small student group in the public school system, as they make up only about 1 or 2 percent of the ELs in each state (Shyyan & Christensen, 2018). However, in California in 2021–22, this translated to nearly 17,000 ELSWDs with particular instruction and assessment needs.¹ In recent years, California’s policy context regarding ELs and students requiring special education services has shifted somewhat, with attention to developing and delivering updated guidance to support the education of ELs with disabilities (CDE, 2019).

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.

ELs in California participate in the Summative ELPAC to measure their annual progress toward English language proficiency (ELP). While the Summative ELPAC is suitable for most ELs, it is not the most appropriate assessment for all EL students. EL students who have been identified as having the most significant cognitive disabilities and who have been found eligible for alternate assessments by their individualized education

¹ CDE reported 16,669 ELSWDs registered for the 2021–22 Alternate ELPAC.

program (IEP) team may participate in the Summative Alternate ELPAC instead. The Summative Alternate ELPAC is designed to allow “for a range of receptive and expressive communication modes, including assistive devices, gestures, and so forth.”²

California identifies students with the most significant cognitive disabilities in need of English language services and supports via the Initial ELPAC or the Initial Alternate ELPAC, and monitors progress toward English language proficiency via the Summative Alternate ELPAC. The Summative Alternate ELPAC is California’s state alternate ELP assessment that EL students with significant cognitive disabilities 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 2021–22 Alternate ELPAC was administered in six grade levels/grade spans—kindergarten, grade one, grade two, grades three through five, grades six through eight, and grades nine through twelve. Subsequent administrations will have a total of seven grade levels/grade spans, with two high school grade spans (9–10 and 11–12) mirroring the structure of the Summative ELPAC. The Summative Alternate ELPAC administration window was open from November 1 through May 31, 2022. Administration is computer based, conducted one-on-one in person via the students’ individually preferred receptive and expressive communication modes (CDE, 2022).

The Summative Alternate ELPAC is designed to measure performance on California’s ELD Connectors. The ELD Connectors were derived from the ELD Standards developed by California educators and published in 2012 after approval by the California State Board of Education. The ELD Connectors reflect the 2012 ELD Standards at reduced breadth, depth, and complexity and represent the highest level of expected performance in ELP for ELs with the most significant cognitive disabilities at a given grade level/grade span.

Chapter 1: Introduction of this report provides the background and rationale for our approach to evaluating the Summative Alternate ELPAC’s alignment to the ELD Connectors. In *Chapter 2: Review of Summative Alternate ELPAC Documentation*, we describe our investigation into the nature of the assessment itself: how the content standards (ELD Connectors) guided the development of the test items (and how the content standards and items should therefore relate to one another) and the interpretations to be made from Summative Alternate ELPAC scores. In *Chapter 3: Summative Alternate ELPAC Alignment Workshop and Outcomes*, we describe how 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. *Chapter 4: Conclusion and Recommendations* briefly summarizes outcomes of HumRRO’s alignment study.

² <https://www.cde.ca.gov/ta/tg/ca/caaiepteamrev.asp>

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 California Department of Education (CDE) identified several research questions to guide the alignment evidence collected. Activities conducted for the Summative Alternate ELPAC Alignment Study were designed to provide information to answer the following research questions:

1. To what extent does the test design of the Summative Alternate ELPAC support the claims to be made about student performance on the assessment?
2. To what extent do the 2021–22 Summative Alternate ELPAC test forms and test items reflect the test design and intended content distributions?
3. Does the Summative Alternate ELPAC include items that cover an appropriate range of linguistic complexity levels to address the English Language Development Connectors?

Review of Summative Alternate ELPAC Documentation

HumRRO researchers collected Summative Alternate ELPAC design and test development materials provided by CDE and Educational Testing Service (ETS) staff, as well as publicly available information about the Summative Alternate ELPAC shared on the CDE website. HumRRO then conducted an evaluation of the alignment of 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 13 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. In the scale, the term “materials” includes 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.

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 13 standards for review. Standard 1.9 is presented below as an example. The full set of rated standards are presented in chapter 2 of this report.

- 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.

All 13 of the identified standards were rated as fully covered based on the available evidence. These results indicate that the Summative Alternate ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to ELD Connectors (see *Chapter 2: Review of Summative Alternate ELPAC Documentation*).

Summative Alternate ELPAC Alignment Workshop and Outcomes

The alignment workshop was designed to collect evidence of whether the Summative Alternate 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 (i.e., familiarity with the Summative Alternate ELPAC tests, student population eligible to take the test, and ELD Connectors) evaluated how well the 2022 test items represent the ELD Connectors.

Alignment Criteria Evaluated

HumRRO developed alignment criteria based on documentation provided by the CDE and ETS. These criteria represent several aspects of the overall alignment of the Summative Alternate ELPAC to the California ELD Connectors. Failure to meet any single criterion does not indicate that the test is invalid or flawed in some way, only that an 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 Alternate ELPAC. We also considered the growing literature on evaluating linguistic difficulty rather than depth-of-knowledge in English language proficiency assessments (Cook, 2006, 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 Alternate ELPAC: Link to Standards, Linguistic Complexity Adequacy, Range Adequacy, and Balance-of-Knowledge Correspondence.

Alignment Workshop Methods

HumRRO conducted a three-day Summative Alternate ELPAC Alignment Study Workshop virtually via Microsoft Teams on August 1–3, 2022. Each workshop day included an eight-hour session. HumRRO worked collaboratively with the CDE to recruit and select a group of 20 educators experienced with the ELD Connectors to serve on six Summative Alternate ELPAC alignment review panels (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, and grades 9–12). The single high school alignment review panel combined grades nine through twelve to correspond with the single high school grade span for the 2021–22 Summative Alternate ELPAC administration (operational field test).

HumRRO developed data collection tools and adapted several other materials to support the data collection process. For example, we used the task type specifications to develop the grade level/grade span-specific linguistic complexity rating aids used during the workshop. 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 Connectors, (b) rating aids for linguistic complexity levels, and (c) a detailed workshop outline and instructions for both panelists and facilitators. The Directions for Administration were provided to panelists in secure electronic format only. ETS created six online test forms for the alignment workshop (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, and grades 9–12) consisting of all the operational 2022 Summative Alternate 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 or grade span panel groups and received more detailed training on the use of materials and CRT, 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 three items. Panelists accessed the items electronically via the CRT and made their independent ratings in their individual electronic spreadsheet. 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.

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 Connector measured by item
 - b. Secondary ELD Connector measured by item, if applicable
 - c. Linguistic complexity level (using a three-point scale). The rating scale was based on the linguistic complexity levels outlined in the task type specifications: (1) Low, (2) Medium, and (3) High.
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.

The HumRRO facilitator recorded the final consensus (or majority) item ratings in a spreadsheet, which was displayed to panelists. Once all consensus ratings were recorded, panelists completed two online surveys: an alignment evaluation survey, and a demographic questionnaire. The alignment evaluation survey was designed to give panelists the opportunity to describe their overall view of the quality of alignment, as well as provide 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 Alternate ELPAC for all grade levels and grade spans. Each criterion for each grade level/span/form reported in table ES.1 is labeled as “Met,” “Not Met,” or “Partially Met.” A criterion may be labeled as “Partially Met” if there were multiple components for the criterion and not all were met (see table 3.1, page 3–24 for full descriptions). These results show that the Summative Alternate ELPAC items are linked to ELD Connectors across all assessments, although grade one panelists aligned several items to one connector and did not indicate any secondary connectors that were not already indicated as primary connectors for other items. The Linguistic Complexity Adequacy criterion showed that most Summative Alternate ELPAC items reflected knowledge, skills, and abilities associated with the Medium linguistic complexity level. The Summative Alternate ELPAC tended to do a good job addressing the organization of the ELD Connectors, with multiple items measuring each Part (Part I: Interacting in Meaningful Ways, Part II: Learning About How English Works, and Part III: Foundational Literacy Skills), Mode (A. Collaborative, B. Interpretive, and C. Productive), and Language Process (A. Structuring Cohesive Texts, B. Expanding and Enriching Ideas, and C. Connecting and Condensing Ideas).

The Summative Alternate ELPAC is for the most part evenly balanced by number of test items for Parts, Modes, and Language Processes.

Table ES.1. Summative Alternate ELPAC Alignment Results

Criterion	K	1	2	3–5 Form 1	3–5 Form 2	6–8 Form 1	6–8 Form 2	9–12 Form 1	9–12 Form 2
Link to Standards	Met	Met							
Linguistic Complexity Adequacy	Not Met	Partially Met	Partially Met	Not Met	Not Met	Not Met	Not Met	Partially Met	Met
Range Adequacy	Met	Partially Met	Met	Met	Met	Met	Met	Met	Met
Balance-of-Knowledge Correspondence	Partially Met	Not Met	Partially Met	Partially Met	Partially Met	Met	Met	Met	Met

While the results include several instances where the Summative Alternate 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 Summative Alternate 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 Alternate ELPAC and the California ELD Connectors. Here we present the conclusions reached for each of the three research questions posed at the beginning of the study:

Research Question 1: To what extent does the test design of the Summative Alternate ELPAC support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprint for the Summative Alternate ELPAC support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment (see table 2.2). Review of 2022 Summative Alternate ELPAC test forms found that all test items are linked to the California ELD Connectors (see table 3.6), thus the Link to Standards criterion was met for all grade levels/grade spans (see table ES.1). The Summative Alternate ELPAC is designed to produce interpretable overall English language proficiency scores.

Research Question 2: To what extent do the 2021–22 Summative Alternate ELPAC test forms and test items reflect the test design and intended content distributions?

Data from the alignment workshop component of the study provide support for the overall alignment of the Summative Alternate ELPAC to the ELD Connectors. All items were rated as measuring at least one ELD Connector, and many were rated as aligned to a secondary ELD Connector. Further, test forms reflected the breadth of the standards by including items from each of the Parts (Part I: Interacting in Meaningful Ways, Part II: Learning About How English Works, and Part III: Foundational Literacy Skills), Modes (A. Collaborative, B. Interpretive, and C. Productive), and Language Processes (A. Structuring Cohesive Texts, B. Expanding and Enriching Ideas, and C. Connecting and Condensing Ideas).

Research Question 3: Does the Summative Alternate ELPAC include items that cover an appropriate range of linguistic complexity levels to address the English Language Development Connectors?

For all grade level/grade span test forms, multiple items were rated at each of the three levels of linguistic complexity. However, across the grade levels/grade spans, more items tended to be rated at the Medium level (see table 3.7). The Linguistic Complexity Adequacy criterion was met by one test form (grade span 9–12 Form 2), partially met by three forms (grade level/grade span 1, 2, and 9–12 Form 1), and not met by five forms (grade level/grade span K, 3–5 Forms 1 and 2, 6–8 Forms 1 and 2) as shown in table ES.1. The Summative Alternate ELPAC test blueprint provides guidance for the number

of each task type that should be included on a test form, and task types may be written to one or more linguistic complexity levels (i.e., Low, Low to Medium, Medium, Medium to High, High). Because of this overlap, it stands to reason that a test form would contain more items at the Medium level. Currently, the blueprint does not specify a distribution of linguistic complexity levels for any of the grade level/grade span test forms.

Recommendations

In this section, we offer three recommendations. These recommendations are based on the results from both the documentation review and the data collected during the alignment workshop.

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

The grade one panel was the only panel that did not identify any items as measuring a secondary connector that they had not already identified as a primary connector for other items. The panel's secondary connector ratings thus did not expand the range of ELD Connectors measured by grade one items. Panelists also rated a large number of items as primarily aligned to one Connector (PI.B.5). This may indicate an issue with the panelists or an issue with the grade one test items. We recommend reviewing these items to determine if there really is a concentration of items measuring this ELD Connector, and/or if other ELD Connectors are being measured via a secondary alignment.

Recommendation 2. Review the linguistic complexity of items at all grade levels to determine if developing additional items at the Low and High linguistic complexity levels is necessary.

Most grade level/grade span test forms did not fully meet the Linguistic Complexity Adequacy criterion. That criterion established linguistic complexity level targets based on the number of each task type presented in the test blueprint. Failure to fully meet the criterion was typically due to a large percentage of items rated at the Medium level (up to 70.8% of items in some forms) and smaller percentages of items rated at the Low (down to 12.5% of items in some forms) and High levels (down to 8.3% of items in some forms).

Analysis of item-person (Wright) maps provided by ETS offers some support for the panelists' findings. Specifically, there are few items on the Summative Alternate ELPAC with difficulties in the part of the scale associated with the top score category (Level 3). Assuming that items with difficulties at the Level 3 performance level require students to demonstrate skills on more linguistically complex content, developing more items at the High linguistic complexity level may be warranted.

Item metadata from ETS includes the linguistic complexity level to which each item was written, based on the task type. CDE should consider reviewing the metadata linguistic complexity categorizations in conjunction with panelists' linguistic complexity consensus ratings to inform the levels, if any, at which additional item development is needed.

Recommendation 3. Refine the test blueprint to specify the number and/or percentage of items at each linguistic complexity level.

It would be beneficial to adjust the test blueprint to include more precise targets for the distribution of linguistic complexity. Currently, the blueprint outlines the number of items representing each task type. Task types may be written to one or more linguistic complexity levels (i.e., Low, Low to Medium, Medium, Medium to High, High). Clearly specified linguistic complexity targets by grade level/grade span test form can better inform item development goals, ensuring that the item bank contains an adequate number of items at each linguistic complexity level and supporting the construction of future test forms that reflect both the breadth and depth of the ELD Connectors.

Chapter 1: Introduction

Human Resources Research Organization (HumRRO) conducted an alignment study of the Summative Alternate English Language Proficiency Assessments for California (Summative Alternate ELPAC) by collecting and evaluating evidence to determine whether the Summative Alternate ELPAC system produces test forms that effectively measure the intended construct, as described in the English Language Development (ELD) Connectors. This report describes the Summative Alternate ELPAC alignment study in detail and summarizes the results.

To evaluate the Summative Alternate ELPAC’s alignment to the ELD Connectors, we first investigated the nature of the assessment itself: how the ELD Connectors guided the development of the test items (and how the ELD Connectors and items should therefore relate to one another) and the interpretations to be made from Summative Alternate ELPAC scores. 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 process also supports federal peer review goals. HumRRO collected data for the second major task of the study by conducting a virtual workshop with subject matter experts who reviewed and rated 2022 operational Summative Alternate ELPAC items.

Background

English Learner (ELs) and students with disabilities have been attended to historically as two different student groups in the public school system. Nationwide, in the fall of 2019, the percentage of U.S. public elementary and secondary school students classified as ELs was 10.4 percent, or 5.1 million students (NCES, 2022). That same year, the percentage of U.S. public school students that were receiving special education services was 14 percent, or 7.1 million students (USDOE, 2021a). During the 2018–19 school year, in California public schools specifically, there were approximately 1.195 million ELs (nearly a quarter of the national population of ELs) and over 720,000 students with disabilities enrolled in kindergarten through grade twelve (CDE, 2023a). Intersecting these two separate groups is a group of students that is gaining attention: EL students with disabilities (ELSWDs). Across the U.S., ELs with significant cognitive disabilities may be a small student group in the public school system, as they make up only about 1 or 2 percent of the ELs in each state (Shyyan & Christensen, 2018). However, in California in 2021–22, this translated to nearly 17,000 students whose particular instruction and assessment needs must be addressed.³ In recent years, California’s policy context regarding ELs and students requiring special education services has shifted somewhat, with attention to developing and delivering updated guidance to support the education of ELs with disabilities (CDE, 2019).

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,

³ CDE reported 16,669 ELSWDs registered for the 2021–22 Alternate ELPAC.

APA, NCME, 2014), hereafter referred to as the *Testing Standards*, as well as the U.S. Department of Education’s (USDOE) *Assessment Peer Review Process* guidance (USDOE, 2021b) 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 which are derived from the domains of Listening, Speaking, Reading, and Writing and address the different proficiency levels of ELs. 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 ELP. Clarification from the U.S. Department of Education’s Office of School Support (2017) indicates this requirement extends to all English learners, even to students with significant cognitive disabilities (Shyyan & Christensen, 2018). 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 (Initial ELPAC or Initial Alternate ELPAC) to newly enrolled students with a language other than English, as indicated on a home language survey; and (2) as a summative assessment (Summative ELPAC or Summative Alternate ELPAC) annually to students who have been identified as EL students until reclassified as fluent English proficient (CDE, 2022). 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 blueprints are designed to be aligned with the 2012 ELD Standards. To review HumRRO’s technical report on the Summative ELPAC alignment study, see <https://www.cde.ca.gov/ta/tg/ep/documents/sumelpacalignmentstudy21.pdf>.

ELs in California participate in the Summative ELPAC to measure their annual progress toward English language proficiency (ELP). While the Summative ELPAC is suitable for most ELs, it is not the most appropriate assessment for all EL students. EL students who have been identified as having the most significant cognitive disabilities and who have been found eligible for alternate assessments by their individualized education

program (IEP) team may participate in the Summative Alternate ELPAC instead. The Summative Alternate ELPAC is designed to allow “for a range of receptive and expressive communication modes, including assistive devices, gestures, and so forth.”⁴

The Summative Alternate ELPAC is designed to align with the 2012 ELD Standards via ELD Connectors, which reduce the depth, breadth, and complexity of the standards, as appropriate for students with the most significant cognitive disabilities. The ELD Connectors were developed through collaboration among California educators, the CDE, and Educational Testing Service (ETS) research and assessment experts, with guidance from a Test Design Advisory Team comprised of nationally recognized experts on the assessment of ELs with the most significant cognitive disabilities.

The Summative Alternate ELPAC is one component of a system that also includes the Initial Alternate ELPAC. While the Initial Alternate ELPAC identifies students who are English learners (ELs) among the population of students with the most significant cognitive disabilities, the Summative Alternate ELPAC assesses the progress of ELs with the most significant cognitive disabilities towards English language proficiency. The 2021–22 Summative Alternate ELPAC was administered to six grade levels or grade spans: kindergarten, grade one, grade two, grades three through five, grades six through eight, and grades nine through twelve. Subsequent administrations will have a total of seven grade levels/grade spans, with two high school grade spans (9–10 and 11–12) mirroring the structure of the Summative ELPAC. The grade-span assessments have the same items for all grades within the span. The Summative Alternate ELPAC administration window was open from November 1, 2021, through May 31, 2022. Administration is computer based, conducted one-on-one in person via the students’ individually preferred receptive and expressive communication modes. (CDE, 2022).

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 Summative Alternate ELPAC is designed to measure students’ language proficiency across four integrated domains: Listening and Reading (Receptive) and Speaking and Writing (Expressive). The individual student will receive an overall score based on a continuous scale for the assessed grade level or grade span, and a corresponding performance level (Level 1: Novice English Learner; Level 2: Intermediate English Learner; and Level 3: Fluent English Proficient). The scale scores are not vertically scaled, so it is not advisable to compare these scores from adjacent grade levels.

For the Summative Alternate ELPAC, evaluating alignment involves examining the test items in terms of their representation of the ELD Connectors, as well as the test’s capacity to indicate students’ readiness for content taught in English.

⁴ <https://www.cde.ca.gov/ta/tg/ca/caaiepteamrev.asp>

Research Questions

Activities conducted for the Summative Alternate ELPAC Alignment Study were designed to provide information to answer three 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 are:

1. To what extent does the test design of the Summative Alternate ELPAC support the claims to be made about student performance on the assessment?
2. To what extent do the 2021–22 Summative Alternate ELPAC test forms and test items reflect the test design and intended content distributions?
3. Does the Summative Alternate ELPAC include items that cover an appropriate range of linguistic complexity levels to address the English Language Development Connectors?

Organization and Contents of the Alignment Study Report

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

- Chapter 2, *Review of Summative Alternate ELPAC Documentation*, presents the methods, rating scale, and data analysis activities HumRRO conducted to evaluate the alignment of development documentation of the Summative Alternate ELPAC to relevant *Testing Standards*. The chapter identifies the list of Summative Alternate 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 Alternate ELPAC documentation.
- Chapter 3, *Summative Alternate ELPAC Alignment Workshop and Outcomes*, presents HumRRO’s method for evaluating the alignment of the 2022 Summative Alternate ELPAC test forms to the ELD Connectors and Summative Alternate 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/grade span (i.e., kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, grades 9–12) for each alignment criterion. The chapter concludes with an overall summary of HumRRO’s evaluation of the alignment of Summative Alternate 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 Alternate ELPAC documentation review and the Summative Alternate ELPAC item ratings by content experts. The chapter offers three recommendations based on HumRRO’s evaluation of the alignment of Summative Alternate ELPAC grade level/grade span test forms.
- Appendix A, *Summative Alternate ELPAC Documents Reviewed by HumRRO*, lists the file names of all documents reviewed for the study. Documents are grouped by these topics of focus: (a) ELD Connectors, (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 two surveys administered online at the conclusion of the workshop (overall debrief and evaluation of alignment workshop training and procedures, demographic information).
- Appendix C, *Summary of Responses to Summative Alternate ELPAC Process Evaluation Survey Questions*, presents tables of panelists’ ratings on the quality of the workshop, including panel facilitation, materials, and processes.
- Appendix D, *Summative Alternate ELPAC Item-Person Maps*, presents ETS’s item-person maps, which display for each grade level and grade span the comparison between Summative Alternate ELPAC item difficulty and student performance.

This page is intentionally blank.

Chapter 2: Review of Summative Alternate ELPAC Documentation

Introduction

To begin the alignment study and build knowledge of the Summative Alternate English Language Proficiency Assessments for California (ELPAC), HumRRO researchers collected and reviewed Summative Alternate 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 Summative Alternate ELPAC shared on the CDE website.

This chapter presents the methods, rating scale, and data analysis activities HumRRO conducted to evaluate the alignment of development documentation of the Summative Alternate ELPAC to relevant *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014), hereafter referred to as *Testing Standards*. We list the Summative Alternate ELPAC documents reviewed for each testing standard and describe the rationale for each of HumRRO's alignment ratings. The chapter concludes with a summary of HumRRO's evaluation of Summative Alternate ELPAC documentation.

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). The complete text of each identified relevant standard is presented in table 2.2. 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 CDE and ETS staff to obtain documentation related to the design and development of the Summative Alternate ELPAC. We also searched ELPAC website pages to identify additional relevant information. The documents generally focus on the following areas: California ELD Connectors; test design; item development information; test modality; test fairness, accessibility, and accommodations; and scoring and administration (see Appendix A).

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 ^a
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 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.

^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.⁵ Standards are numbered to reflect the chapter of the *Testing Standards* 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.

⁵ 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 Alternate 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"> • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • ELD Standards CDE Publication14.pdf • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819.pdf • 363-2019A V3 FOR ARCHIVE Alternate ELPAC Item Review Meeting Plan • 363-2019B V3 FOR ARCHIVE Alt-ELPAC Item Review Meeting Slides_071219 • 445-2022D v3 FOR ARCHIVE Alt ELPAC Standard Setting Technical Report 072022 	<p>5</p>
<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"> • 438-2020 v4 Alt ELPAC Task Type Specifications Preface_011420 • Alternate ELPAC Test Blueprint - ELPAC (CA Dept of Education) • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819 • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • AltELPACblueprint.pdf • ELD Standards CDE Publication14.pdf • 440-2021 v4 FOR ARCHIVE Alt-ELPAC Field Test Specifications 082520 	<p>5</p>

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
<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"> • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 • 359-2021F V3 FOR APPROVAL Alt ELPAC Range PLDs G1 030921 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • ELD Standards CDE Publication14.pdf • 438-2020 v4 Alt ELPAC Task Type Specifications Preface_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Communicate Familiar Topics_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Describe a Routine_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Informational Text_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Literary Text_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Opinion_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Recognize Use Common Words_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_School Activity_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_School Exchange_011420 	<p>5</p>

Table 2.2 (cont.)

Standard	Supporting Documentation	Standard Rating
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"> • Reliability Chapter.docx • 446-2022 v1 FOR REVIEW Alt ELPAC Technical Report-appendix 8.110322.docx 	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"> • Reliability Chapter.docx • 446-2022 v1 FOR REVIEW Alt ELPAC Technical Report-appendix 8.110322.docx 	5
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.	<ul style="list-style-type: none"> • 021318-01 v2 FOR ARCHIVE CAASPP Item Acceptance Criteria for IRC 022118_ • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report _050819 • 441-2020 V3 FOR ARCHIVE Alt ELPAC IWW PPT 021320 • 711-2021-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW General PPT_022120 • Alt-ELPAC_IRM_Guiding Questions • Universal design for item development • 438-2020 v4 Alt ELPAC Task Type Specifications Preface_011420 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • 711-2021A-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW Plan_022720 • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819.pdf 	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"> • Proposed High-Level Test Design for the Alternate ELPAC • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • Universal design for item development.docx • Alt-ELPAC_IRM_Guiding Questions.docx • 711-2021-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW General PPT_022120 	<p>5</p>
<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"> • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 • Proposed High-Level Test Design for the Alternate ELPAC • 440-2021 v4 FOR ARCHIVE Alt-ELPAC Field Test Specifications 082520 • 711-2021-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW General PPT_022120 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819.pdf • ELD Standards CDE Publication14.pdf • AltELPACblueprint.pdf 	<p>5</p>

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"> • Proposed High-Level Test Design for the Alternate ELPAC (CA Dept of Education) • 359-2021F V3 FOR APPROVAL Alt ELPAC Range PLDs G1 030921 • Universal design for item development.docx • Alt-ELPAC_IRM_Guiding Questions.docx • ELD Standards CDE Publication14.pdf • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819.pdf • 440-2021 v4 FOR ARCHIVE Alt-ELPAC Field Test Specifications 082520 	<p>5</p>
<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"> • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819 • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • 23R-673-1095 FOR ARCHIVE Alt ELPAC Connectors Review Mtg Participants Balance Data 011123.xlsx • 23R-673-1095 FOR ARCHIVE Alt ELPAC Connectors Review Mtg Participants Table 011123.docx 	<p>5</p>

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"> • Alternate ELPAC Test Blueprint - ELPAC (CA Dept of Education) • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819 • AltELPACblueprint.pdf 	<p>5</p>
<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"> • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • 440-2021 v4 FOR ARCHIVE Alt-ELPAC Field Test Specifications 082520 • ALT-ELPAC--Practice-Test-Scoring-Guide-Grade-K.2020-21 • ALT-ELPAC--Practice-Test-DFA-Grade-K.2021-22 • ALT-ELPAC--Training-Test-DFA-Grades-K-5.2021-22 • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • Universal design for item development.docx • Alt-ELPAC_IRM_Guiding Questions.docx • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report_050819 	<p>5</p>

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"> • AltELPACblueprint.pdf • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620 • 359-2021F V3 FOR APPROVAL Alt ELPAC Range PLDs G1 030921 • ELD Standards CDE Publication14.pdf 	<p>5</p>

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.

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 covers all aspects of Standard 1.9.

The key element of validation that is informed by expert judgments is the content validity evidence collected via reviews by California educators. The educators are involved in reviewing both the ELD Connectors and the items. Experience of educators is outlined, including experience teaching the standards and ELD Connectors and working with this student population. Item review training processes and procedures are described. The judgment and consensus processes are described in detail, including how dissenting opinions are recorded and recommendations tallied.

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 covers all aspects of Standard 1.11.

The ELD Connectors, which serve as the basis for the Summative Alternate ELPAC, were created based on the ELD Standards considered to be appropriate for the Summative Alternate ELPAC population. The ELD Connectors, each linked to an ELD Standard, were created with input from subject matter expert stakeholders to ensure they provide appropriate levels of challenge and rigor for kindergarten through grade twelve English learners (ELs) with significant cognitive disabilities. The high-level test design document and blueprint illustrate how items are designed to integrate the four language domains (Listening, Reading, Speaking, Writing) and allow for individually preferred modes of communication to accommodate the needs of the target population. Item development documents provide specifications for receptive and expressive item

types as well as rules for linguistic complexity and skills to be targeted by items within particular task types.

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 covers all aspects of Standard 1.12.

EL students with the most significant cognitive disabilities represent a diverse population of students in kindergarten through grade twelve, inclusive of students up to age 22 enrolled in grade twelve who continue to be eligible for special education and ELD services. A wide variety of language- and disability-related needs and alternate ways of communicating require careful thinking about how to measure the English language proficiency needed to communicate in social and academic contexts. California's ELD Standards incorporate an awareness of the cognitive development of students at various ages to make them appropriate at different grade levels. The Summative Alternate ELPAC ELD Connectors Report describes the processes used to develop the ELD Connectors for the Summative Alternate ELPAC. The ELD Connectors were used for key test development tasks such as identifying task types, drafting a test blueprint, and developing test items for the Summative Alternate ELPAC. California educators who participated in the review of the ELD Connectors and in the cognitive lab study contributed additional evidence of the alignment of the tasks and items to the communication modes and the agreement between perceived student skills and test scores.⁶

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 covers all aspects of Standard 2.3.

The Summative Alternate ELPAC 2021–2022 Technical Report contains a chapter dedicated to reliability and validity evidence to support the interpretations and uses of Summative Alternate ELPAC scores. Because only a total score is reported for students, reliability coefficients are only reported for total scores. Reliability evidence includes IRT-based reliability coefficients (i.e., marginal reliability), standard error of measurement (SEM), conditional standard error of measurement (CSEM), and interrater

⁶ The cognitive lab study was conducted by ETS to determine whether the Alternate ELPAC task types are suitable for the intended population. It included observations of test examiners and student interactions with the task types.

reliability statistics (e.g., kappa). Marginal reliability and SEM are reported for all students and for disaggregated groups.

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 covers all aspects of Standard 2.16.

Appendix 8.A of the Summative Alternate ELPAC 2021–22 Technical Report provides a summary of the accuracy and consistency of classifications of students into adjacent performance levels and overall. The reliability and validity chapter of the technical report (chapter 1) explains that accuracy refers to the extent to which students are classified in the same way as they would be if their true scores could be known, whereas consistency refers to the extent to which students would be classified in the same way based on two nonoverlapping, equally difficult forms of the test.

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 covers all aspects of Standard 3.2.

The Summative Alternate ELPAC is intended only for EL students, and potential EL students, who have been identified as having the most significant cognitive disabilities and who have been found eligible for alternate assessments by their individualized education program (IEP) team. The universal design document, guiding questions, and item writer training all additionally address the target population. Additionally, the cognitive lab study documents efforts to evaluate construct-irrelevant features of the test, as identified by target students and test examiners, and the subsequent efforts taken to remove those obstacles.

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 covers all aspects of Standard 3.9.

The cognitive lab study lists the various universal tools, supports, and accommodations available to test takers. Additionally, the study documented which of those resources test examiners and students made use of and the barriers to their use. The study resulted in recommendations for how to improve the accessibility of those resources.

The universal design document, guiding questions, and item writer training all additionally address the target population.

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 covers all aspects of Standard 4.0.

The Alternate ELPAC ELD Connectors Report describes the two sets of experts (national experts on alternative assessment of ELs and educators) involved to ensure the appropriateness of the ELD Connectors relative to the 2012 ELD Standards and the target test population. The high-level test design and item writing documents outline the processes built into item writing and reviewing that account for validity and fairness concerns. The Pilot Cognitive Lab Report describes efforts taken to understand how a sample of target students and their test examiners engage with the test and the necessary changes suggested as a result of what was learned from the study. The documentation of test specifications and development procedures indicate the intended uses of the test have been supported and lead to valid interpretations of test results.

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: 5, Evidence in the materials fully covers all aspects of Standard 4.1.

The proposed high-level test design outlines the purpose, construct definition, target population, and interpretation of the test, including a rationale for the interpretation. The performance level descriptor (PLD) document provides additional relevant information as it identifies the language skills associated with the three performance levels. The Summative Alternate ELPAC operational field test generated item-level statistics that will be used to support the Summative Alternate ELPAC purposes and will also inform the test specifications for the operational versions of both the Initial Alternate ELPAC and Summative Alternate ELPAC. These specifications will include the embedded field test item distribution, the number of linking items, and other related information.

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 covers all aspects of Standard 4.6.

Two different groups of experts were invited to participate in the development of the ELD Connectors, which serve as the basis for the test specifications. The documentation provides the biographies of the national experts and the expertise of the California educators, including details about their educational background, professional credentials, and experience working with students with disabilities and English Learners. The purpose of each of the reviews is defined.

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 covers all aspects of Standard 4.12.

There is substantial evidence to indicate the representation of the content domain was considered throughout the design and development processes. The test blueprint document provides a table that specifies which of the ELD Connectors has primary, secondary, or no alignment to each grade level test. The cognitive lab study also includes a high-level judgment from California educators about the alignment of the items to the two general constructs: receptive and expressive modes of communication. The teachers largely agreed that the items did reflect those modes.

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 covers all aspects of Standard 4.16.

Preparation materials are provided to students, parents or guardians, and test examiners to familiarize them with the test platform, process, scoring criteria, and questions. There are (a) two training tests that provide sample questions, one for kindergarten through grade five and one for grades six through twelve; (b) a practice test for each grade level test that mirrors the operational test length, item types, and directions for administration; and (c) a practice test scoring guide for each grade level test. The pilot cognitive lab study also gathered feedback from test examiners on the use of directions for administration and scoring rubrics. ETS has made recommendations and taken actions to improve those materials for test examiner use.

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 covers all aspects of Standard 12.4.

The test blueprint document provides a table that specifies which of the ELD Connectors has primary, secondary, or no alignment to each grade level test. The PLD document also specifies skills that define student English language abilities at different performance levels for relevant ELD Connectors. The cognitive lab study also includes a high-level judgment from California educators about the alignment of the items to the two general constructs - receptive and expressive modes of communication; the teachers largely agreed that the items did reflect those modes.

Summary and Discussion

All 13 of the identified standards were rated as fully covered based on the available evidence. These results indicate that the Summative Alternate ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to ELD Connectors. Chapter 3 of this report describes the alignment workshop convened to document the extent to which test forms are adequately aligned to the ELD Connectors.

This page is intentionally blank.

Chapter 3: Summative Alternate ELPAC Alignment Workshop and Outcomes

Introduction

The review of Summative Alternate ELPAC materials described in the prior chapter informed HumRRO's plans and preparation for the second major task of the study, the alignment workshop. This chapter presents the four Summative Alternate ELPAC alignment criteria; describes the alignment workshop data collection activities, including panelist recruitment, training, and item rating procedures; and presents results of data analysis. The results section provides outcomes by grade level/grade span for each alignment criterion. The chapter concludes with an overall summary of HumRRO's evaluation of the alignment of the Summative Alternate ELPAC to the ELD Connectors.

Summative Alternate ELPAC Alignment Criteria

Alignment studies 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 Summative Alternate ELPAC by the ELD Connectors. The alignment workshop is designed to evaluate how well the test items represent (align with) the ELD Connectors. For the Summative Alternate ELPAC, four main criteria were evaluated. Table 3.1 provides a brief description of the criteria addressed in this study. HumRRO developed the criteria following best practices for alignment criteria, basing them on the Summative Alternate ELPAC test design with review and input from the CDE's ELPAC Technical Advisory Group.

Table 3.1 Summative Alternate ELPAC to ELD Connectors Alignment Criteria

Criteria	Description	Acceptability of Results***
Link to Standards	At least 90% of items are matched to an ELD Connector. At least 50% of the ELD Connectors are represented by items.	To meet criterion, both components must be acceptable. No partially met option.
Linguistic Complexity Adequacy	At least 20% and no more than 50% of items on a test form are rated at each of the three linguistic complexity levels (Low, Medium, and High).	To fully meet criterion, threshold % must be acceptable for every level (Easy, Medium, and Difficult). Criterion is partially met if threshold % is acceptable for two levels.
Range Adequacy*	<p>Part I: Interacting in Meaningful Ways, Part II: Learning About How English Works, and Part III: Foundational Literacy Skills as outlined in the ELD Connectors are each measured by at least one item.*</p> <p>Each of the Part I Modes (A. Collaborative, B. Interpretive, and C. Productive) and Part II Language Processes (A. Structuring Cohesive Texts, B. Expanding and Enriching Ideas, and C. Connecting and Condensing Ideas) outlined in the ELD Connectors is measured by at least one item.</p>	To fully meet criterion, all three components (Parts, Modes, and Language Processes) must be acceptable or partially acceptable. A component is partially acceptable if there is at least one item measuring two of its three ELD Dimensions (i.e., two of the three Parts, Modes, or Language Processes).
Balance-of-Knowledge Correspondence**	Webb’s balance-of-knowledge correspondence criteria is used, computed for the Part*, Mode, and Language Process components of the ELD Connectors. All must meet Webb’s threshold of 0.70.	To fully meet criterion, balance index must be acceptable for each component (Part, Mode, Language Process). Criterion is partially met if balance index is acceptable for two components.

* Part III only assessed in grade 2 and grades 3–5.

** 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.

*** The definitions of partially met were not included in the alignment criteria but were developed after the analysis.

Link to Standards

All Summative Alternate ELPAC test items should reflect content described by the ELD Connectors. An item that does not directly relate to the ELD Connectors 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 ELD Connectors. The criterion is less than 100 percent only to prevent a single poorly written item from triggering an inappropriate alignment judgment. However, any item that is not directly linked to content from the ELD Connectors should be revised to address the intended content or eliminated from the assessment.

It is also important that the Summative Alternate ELPAC test items reflect the breadth of the ELD Connectors. We would expect at least half of the ELD Connectors for a grade level/grade span to be matched to at least one test item. Coverage of too few of the ELD Connectors would indicate that students’ Summative Alternate ELPAC scores do not adequately reflect the English language proficiency construct as outlined in the ELD Connectors document.

Linguistic Complexity Adequacy

English language proficiency assessment items are often categorized by linguistic difficulty level (Cook, 2006) in alignment studies. The test blueprint for the Summative Alternate ELPAC provide guidelines for the linguistic complexity of items. Specifically, the blueprint outlines the number of items representing each task type. Task types are designed to reflect one or more linguistic complexity levels. This information was used to inform the target percentages for this criterion.

In addition to evaluating the item-level linguistic complexity levels for meeting this criterion, we reviewed item-person (Wright) maps created by ETS. Although item difficulty is distinct from linguistic complexity, we expect the two to be positively correlated. The Wright maps will provide additional evidence that test items reflect the range of student performance on English language proficiency content.

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 ELD Connectors document. For the Summative Alternate ELPAC, we use a Range Adequacy criterion. It requires the Summative Alternate ELPAC to reflect the structure of the ELD Connectors (i.e., “Parts,” “Modes,” and “Language Processes”). Parts include Interacting in Meaningful Ways (Part I), Learning About How English Works (Part II), and Foundational Literacy Skills (Part III). Modes of communication include Collaborative, Interpretive, and Productive. Language Processes include Structuring Cohesive Texts, Expanding and Enriching Ideas, and Connecting and Condensing Ideas. Table 3.1 summarizes the organization of ELD Connectors.

Table 3.2 Organization of ELD Connectors

Part	Clusters of ELD Connectors
I: Interacting in Meaningful Ways	Modes: A. Collaborative, B. Interpretive, and C. Productive
II: Learning About How English Works	Language Processes: A. Structuring Cohesive Texts, B. Expanding and Enriching Ideas, and C. Connecting and Condensing Ideas
III: Foundational Literacy Skills	No clusters (single ELD Connector)

To clarify, the structural title “modes of communication” in the ELD Standards and ELD Connectors is distinct from the test administration guidance on individual preferred student modes of communication. EL students with the most significant cognitive disabilities are assessed via the students’ individually preferred receptive and expressive communication modes to help ensure the maximum participation of all eligible test takers and eliminate the need to provide domain (e.g., speaking, reading, listening, writing) exemptions.

To fully meet the Range Adequacy criterion, the Summative Alternate ELPAC test forms must (a) include at least one item representing each Part assessed for the grade level/ grade span, (b) include at least one item representing each Mode, and (c) include at least one item representing each Language Process. Items can meet these requirements based on either the primary identified ELD Connector or via a secondary identified ELD Connector.

Balance-of-Knowledge Correspondence

Webb’s *Balance of Representation* focuses on content coverage in yet more detail. In this case, the number of items matched to the ELD Connector does matter. The *Balance-of-Knowledge Correspondence* criterion determines whether the assessment measures the ELD Connectors equitably within each content category using only those ELD Connectors identified by panelists as measured by the test items. Based on Webb’s (1997) method, items should be distributed evenly across the ELD Connectors per content category for good balance. The *Balance-of-Knowledge Correspondence* is determined by calculating an index, or score, for each content category. Each category 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 Connector category). The more unevenly the items are distributed, the lower the index.

For the Summative Alternate ELPAC, we computed Webb’s *Balance-of-Knowledge Correspondence* index for Part, Mode, and Language Process (see table 3.1). To meet the *Balance-of-Knowledge Correspondence* criterion, the Summative Alternate ELPAC test forms were required to have balance indexes of greater than 0.70 for all three components. 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 educators to serve on six Summative Alternate ELPAC alignment review panels (kindergarten, grade 1, grade 2, grade 3–5, grades 6–8, and grades 9–12). The single high school alignment review panel combined grades nine through twelve to correspond with the single high school grade span for the 2021–22 Summative Alternate ELPAC administration (operational field test). HumRRO began recruitment activity for the virtual workshop in April 2022 with a statewide *Assessment Spotlight* announcement and link to a recruitment survey. Educators interested in being panelists responded with their contact information, answers to eligibility questions, and preferred virtual workshop dates and times. HumRRO pursued recruitment of educators who met the following selection criteria:

- Had taught English learners with the most significant cognitive disabilities within the last three years.
- Had not participated in item writing or item review activities for Summative Alternate ELPAC test development.
- Had access to required technology (i.e., laptop; high-speed reliable Internet access; camera).
- Had access to an environment in which they could actively participate and listen during group sessions and focus to make individual ratings.
- Had either read or had working knowledge of the ELD Connectors.
- Had experience entering information on a spreadsheet.

Despite a solid response to the recruitment survey (136 respondents), many interested educators were either ineligible due to their participation in the Summative Alternate ELPAC test development activities or had conflicts with the dates and times preferred by the majority of eligible panelists. The CDE asked ETS to send a recruitment email to all Summative Alternate ELPAC test examiners, which resulted in some additional eligible respondents. However, the lack of available panelists for the lowest grades required a change to the workshop plans. HumRRO combined educators with experience in kindergarten and grade two to form a panel that participated in an additional workshop day, allowing the same panel to provide item ratings for both of these grades.

At the conclusion of recruitment in mid-July, at least 6 educators had committed to participate in the virtual workshop for each grade level/grade span panel, for a total of 30 panelists. Prior to the workshop, there were 5 panelist cancellations, and there were an additional 5 cancellations the weekend before or day of the workshop. Due to these cancellations and a limited pool of alternate educators (3 of whom were not familiar with the ELD Connectors), 20 panelists participated in the workshop representing 17 local educational agencies. Eighty percent of the panelists were currently teachers, and the remaining 20 percent reported working as EL coordinator, intervention specialist, educational specialist, and assessment specialist.

Table 3.3 summarizes the final number of educators in each panel. It also presents panelists' responses to two recruitment survey questions: if they had administered the Summative Alternate ELPAC (75–100% of panelists in each grade level/grade span had done so), and their knowledge of the ELD Connectors (four response options). All grade level/grade span panels included educators with a working knowledge of ELD Connectors (participated in professional development or developed curriculum). Panels varied as to the percent who had minor knowledge (read the ELD Connectors), and there was no more than one panelist per grade level/grade span who was “not familiar with the ELD Connectors” prior to the workshop.

Table 3.3 Characteristics of Summative Alternate ELPAC Alignment Panelists

Panel Grade Level/ Grade Span	# in Panel	Gave Alt ELPAC	Minor or No Knowledge of ELD Connectors	Working Knowledge of ELD Connectors
K & 2	4	4	1 Had read the ELD Connectors	3 Participated in professional development related to ELD Connectors 2 Developed curriculum that incorporates ELD Connectors
1	3	3	1 Not familiar with ELD Connectors	2 Participated in professional development related to ELD Connectors
3–5	5	4	3 Had read the ELD Connectors	2 Participated in professional development related to ELD Connectors
6–8	4	4	1 Not familiar with ELD Connectors	3 Participated in professional development related to Connectors 1 Developed curriculum that incorporates ELD Connectors
9–12	4	3	1 Not familiar with ELD Connectors; 2 Had read the ELD Connectors	1 Developed curriculum that incorporates ELD Connectors

Table 3.4 summarizes the panelists' responses to the recruitment survey question, "Indicate which disability categories match those of students you have taught in the last 3 years. Select all that apply." All but 3 of the 17 panelists indicated experience teaching students with intellectual disabilities ("Intellectual disabilities" is shown in bold-faced font in the table), and most panelists had experience teaching students with a variety of physical, emotional, and learning disabilities.

After the workshop, panelists completed a demographic survey that asked them to "select all of the student groups with whom you have worked" (see Appendix B). All 20 panelists (100%) reported teaching English learners, students of color, students from low socioeconomic households, and students receiving free/reduced lunch. Also, 85 percent reported teaching students with the most significant cognitive disabilities, and 15 percent reported teaching students with mild to moderate cognitive disabilities.

These are other summary characteristics of the panelists:

- Highest degree: 70% achieved a Master's, 15% a Bachelor's, 5% a PhD.
- Years of teaching experience: 100% at least 3 years; 50% more than 15 years
- 85% female, 15% male
- 75% non-Hispanic, 25% Hispanic
- 55% self-identified as White, 20% as Black or African American, 5% as Asian, and 10% as multi-racial (White and Filipino, and White and American Indian or Alaskan Native).
- Age: 40% 26–45 years; 50% 46–55 years; 10% 56–65 years

Table 3.4 Panelists' Experience Teaching Students with Disabilities, by Grade Level/Grade Span

Panel Grade Level/Grade Span	Disability Categories of Students Panelists Had Experience Teaching	Number of Panelists
K & 2	Autism; Other health impairment; Specific learning disability; Speech or language impairment;	1
K & 2	Autism; Emotional disturbance; Hard of hearing; Intellectual disabilities ; Multiple disabilities; Other health impairment; Specific learning disability; Speech or language impairment	1
K & 2	Autism; Emotional disturbance; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Other health impairment; Specific learning disability; Speech or language impairment	1
K & 2	Autism; Emotional disturbance; Hard of hearing; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Other health impairment; Specific learning disability; Speech or language impairment; Traumatic brain injury; Visual impairment; Deaf-blindness	1
1	Autism only	1
1	Autism; Emotional disturbance; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Other health impairment; Specific learning disability; Speech or language impairment	1
1, 3–5	Autism; Intellectual disabilities ; Hard of hearing; Speech or language impairment; Multiple disabilities	2
3–5	Autism; Emotional disturbance; Hard of hearing; Intellectual disabilities ; Multiple disabilities; Other health impairment; Specific learning disability; Speech or language impairment	1
3–5	Emotional disturbance; Intellectual disabilities ; Speech or language impairment	1
3–5	Autism; Intellectual disabilities ; Orthopedic impairment; Speech or language impairment; Traumatic brain injury;	1
3–5	Autism; Intellectual disabilities ; Multiple disabilities; Other health impairment; Specific learning disability; Speech or language impairment	1
6–8	Autism; Intellectual disabilities	1
6–8	Autism; Intellectual disabilities ; Multiple disabilities; Other health impairment	1
6–8	Autism; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Traumatic brain injury; Visual impairment	1
6–8, 9–12	Autism; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Speech or language impairment	2
9–12	Autism; Intellectual disabilities ; Multiple disabilities	1
9–12	Hard of hearing only	1
9–12	Autism; Intellectual disabilities ; Multiple disabilities; Orthopedic impairment; Other health impairment; Specific learning disability; Speech or language impairment; Visual impairment	1

Workshop Logistics

HumRRO conducted the Summative Alternate ELPAC Alignment Study Workshop virtually via Microsoft Teams on August 1–3, 2022. Prior to entering the workshop, panelists were required to sign nondisclosure agreements as a condition of participation. Panelists received several calendar invitations that provided links to the whole group training session and grade level/grade span panel meetings, where further training, calibration, and item rating occurred. During the workshop, panels of educators evaluated which ELD Connector(s) aligned with each Summative Alternate ELPAC item and which level of linguistic complexity was required to correctly respond to the item.

Workshop Materials

The 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 six online test forms for the alignment workshop (kindergarten, grade 1, grade 2, grades 3–5, grades 6–8, and grades 9–12) consisting of all the operational 2022 Summative Alternate ELPAC items. ETS also created accounts for HumRRO researchers and participants to securely access the items using the IBIS™ Content Review Tool (CRT). In each panel, at least one educator had prior experience using the 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 Connectors, (b) Linguistic Complexity Rating Aid, and (c) detailed workshop outline and instructions for both panelists and facilitators. For security purposes, Directions for Administration (DFA), which include item scoring information, 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/grade 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 and 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 Connectors, grade level/grade span-specific Linguistic Complexity Rating Aids, 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., 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 three 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 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 independent ratings of (a) the primary ELD Connector measured by the item, (b) the secondary ELD Connector measured by the item, and (c) the linguistic complexity level of the item. The panelists next discussed their ratings for an item, then reached their final consensus/majority rating for that item before moving on to the next.⁷ Once consensus/majority ratings were recorded for that group of items, the facilitator shared the item metadata for those items and then the panel moved on to the next group and repeated this process.

During the workshop, we modified one step in the rating process. Our original process entailed showing item metadata prior to consensus ratings. In some panels, when panelists had trouble coming to consensus, seeing the metadata was helpful in providing a different way for them to see the item, though facilitators reported that panelists did not feel the need to match the metadata. However, we observed during the workshop that this step impeded the progress of several panels' discussions. Specifically, panelists' discussions became focused on the metadata and why they agreed or disagreed with it, rather than on coming to consensus regarding their

⁷ When consensus could not be reached, we recorded the majority rating.

judgments about the item. On day one we revised the process, with approval of the CDE contract monitor, 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 Connector. The facilitator then polled the group to determine consensus on the ratings that had been discussed and recorded the rating in a spreadsheet. If the group could not reach true consensus, the facilitator recorded the rating of the majority of panelists.

Once all consensus ratings were recorded, panelists completed two online Microsoft Forms surveys: an alignment debriefing and workshop evaluation survey, and a demographic survey (See Appendix B for survey questions). Panelists were then released from the workshop.

The alignment debriefing questions gave panelists the opportunity to describe their overall view of the quality of alignment. Table 3.5 presents a summary of panelists’ ratings by grade level/grade span, with most rating the Summative Alternate ELPAC items they reviewed as overall “strongly aligned” to the ELD Connectors. No panelist for any grade level/grade span chose the third rating option, “not at all aligned.”

Table 3.5 Panelists’ Ratings of Overall Quality of Alignment, by Grade Level/Grade Span

Grade Level/ Grade Span	Total Number of Panelists	Panelists Rating “Strongly Aligned”	Panelists Rating “Partially Aligned”
K	4	3	1
1	3	2	1
2	4	2	2
3–5	5	2	3
6–8	4	1	3
9–12	4	4	0

Appendix C presents tables summarizing responses to the questions that elicited feedback about the quality of the workshop, including panelist training, panel facilitation, materials, and processes. Panelists mostly agreed or strongly agreed that all aspects of the workshop were effective, and facilitators effectively led discussions and ensured all perspectives were heard.

Results

This section summarizes the data/information collected during the Summative Alternate 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 Alternate ELPAC, we ask if the test items relate directly to the content standards (ELD Connectors) on which the test is based. To meet this criterion, at least 90 percent of items must be matched to an ELD Connector and at least 50 percent of the ELD Connectors must be matched to at least one item. Table 3.6 presents the findings for all grade levels/grade spans for the Link to Standards criterion for the Summative Alternate ELPAC.

Table 3.6 Results for Link to Standards

Grade Level/ Grade Span	Items Matched to ELD Connector	ELD Connectors with Primary Match to Items	ELD Connectors with Primary or Secondary Match to Items	Acceptable?
K	24/24 (100%)	9/15 (60.0%)	13/15 (87%)	Yes
1	24/24 (100%)	8/16 (50.0%)	8/16 (50.0%)	Yes
2	24/24 (100%)	11/18 (61.1%)	17/18 (94%)	Yes
3–5 Form 1	24/24 (100%)	7/18 (38.9%)	14/18 (78%)	Yes
3–5 Form 2	24/24 (100%)	9/18 (50.0%)	15/18 (83%)	Yes
6–8 Form 1	24/24 (100%)	9/17 (52.9%)	13/17 (77%)	Yes
6–8 Form 2	24/24 (100%)	11/17 (64.7%)	13/17 (77%)	Yes
9–12 Form 1	24/24 (100%)	9/17 (52.9%)	15/17 (88%)	Yes
9–12 Form 2	24/24 (100%)	10/17 (58.8%)	13/17 (76%)	Yes

All Summative Alternate ELPAC items across all grades were aligned to a primary connector, meeting the first component of Criterion 1. All grades also met the second component of Criterion 1, with at least 50 percent of the ELD Connectors having a primary or secondary match to at least one item. We do note that grade 3–5 Form 1 had fewer than 50 percent of ELD Connectors with a primary match to at least one item. This is in part due to panelists identifying no items measuring any PI.A, PI.B, or PII.A Connectors. Also, the grade one panel was the only panel that did not identify any items

as measuring a secondary connector that they had not already identified as a primary connector for other items. The panel’s secondary connector ratings thus did not expand the range of ELD Connectors measured by grade one items.

Criterion 2: Linguistic Complexity Adequacy

This criterion is based on panelists’ ratings of the linguistic complexity of the test items. The linguistic complexity levels describe the characteristics of items at each of three levels of increasing complexity (Low, Medium, and High). To be considered acceptable, each Summative Alternate ELPAC test form must include at least 20 percent and no more than 50 percent of items across the assessment rated at each of the three linguistic complexity 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.

Tables 3.7 summarizes the findings across the grade levels/grade spans for the Linguistic Complexity Adequacy criterion for the Summative Alternate ELPAC. In the “Acceptable?” column, the criterion is labeled Partially if two of the three linguistic complexity levels met the criterion (see table 3.1, page 3–24 for full descriptions).

Table 3.7 Results for Linguistic Complexity Adequacy

Grade Level/ Grade Span	Items Rated Low	Items Rated Medium	Items Rated High	Acceptable?
K (n=24)	3 (12.5%)	19 (79.2%)	2 (8.3%)	No
1 (n=24)	4 (16.7%)	10 (41.7%)	10 (41.7%)	Partially
2 (n=24)	4 (16.7%)	12 (50.0%)	8 (33.3%)	Partially
3–5 Form 1 (n=24)	3 (12.5%)	17 (70.8%)	4 (16.7%)	No
3–5 Form 2 (n=24)	4 (16.7%)	17 (70.8%)	3 (12.5%)	No
6–8 Form 1 (n=24)	6 (20.8%)	15 (62.5%)	4 (16.7%)	No
6–8 Form 2 (n=24)	6 (25.0%)	15 (62.5%)	3 (12.5%)	No
9–12 Form 1 (n=24)	8 (33.3%)	12 (50.0%)	4 (16.7%)	Partially
9–12 Form 2 (n=24)	6 (25.0%)	12 (50.0%)	6 (25.0%)	Yes

The current Summative Alternate ELPAC items tend to be concentrated in the Medium linguistic complexity level. Only grade 9–12 Form 2 had acceptable proportions of items in each of the three levels. Grades one and two, and grade 9–12 Form 1, met target proportions at two of the three levels and are considered to partially meet Criterion 2. All other grade and grade-span test forms did not meet this criterion.

Analysis of Item-Person Maps

To further investigate the Linguistic Complexity Adequacy results, HumRRO requested and obtained item-person (Wright) maps from ETS to demonstrate where items and students perform on the Summative Alternate ELPAC on a common scale. Note that item difficulty and linguistic complexity are not the same, but they can be expected to be positively correlated (e.g., items with higher linguistic complexity tend to be more difficult than items with lower linguistic complexity).

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 the full range of test-takers. 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. Test score information depends on item information from items with difficulties throughout the score range. If there are areas on the scale with lower information, students' scores from those parts of the scale will have higher error associated with them than scores from areas with higher information. Item-person maps produced by ETS for each grade level and grade span of the Summative Alternate ELPAC are presented in Appendix D.

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

The item-person map for grade two is replicated in table 3.8 for illustrative purposes. The number of students scoring at each ability level is presented on the left side of the table. The number of items at each difficulty level is presented on the right side of the table. There are dashed horizontal lines representing each threshold score, the lowest score at which a student would be classified at the next higher overall performance level (i.e., Level 1 to 2 and Level 2 to 3). Therefore, easier items are located toward the bottom of the table, while more difficult items are located toward the top.

The Summative Alternate ELPAC was developed such that students scoring at Level 3 are judged Fluent English Proficient in the assessed English language skills. When we examine the distribution of students, we see a mostly normal curve (depicted vertically on the left side of table 3.8). The curve's highest point, representing the most students, is in Level 2, but there are a substantial number of students in Levels 1 and 3 as well.

If we examine the item locations, we see that most item difficulties are in Levels 1 and 2. There are only two items in Level 3. The Es and Rs on the figure represent expressive or receptive items, respectively. There is only one receptive and one expressive item in Level 3.

Table 3.8 Summative Alternate ELPAC Grade Two Item-Person Map

Number of Students	Students' Ability Estimate*	Logit	Item Difficulty**	Number of Items	Performance Level
4	.	4.0	-	0	
6	.x	3.8	-	0	
0	-	3.6	-	0	
0	-	3.4	-	0	
4	.	3.2	-	0	
0	-	3.0	-	0	
6	.x	2.8	-	0	
0	-	2.6	-	0	
13	.xx	2.4	-	0	
15	xxx	2.2	-	0	
19	.xxx	2.0	-	0	
17	.xxx	1.8	E	1	
16	.xxx	1.6	-	0	
25	xxxxx	1.4	R	1	PL3
32	.xxxxxx	1.2	-	0	
39	.xxxxxxx	1.0	EEEE	4	
38	.xxxxxxx	0.8	REREEEE	7	
51	.xxxxxxxxxx	0.6	ER	2	
48	.xxxxxxxxxx	0.4	ERRRRE	6	
51	.xxxxxxxxxx	0.2	E	1	
54	.xxxxxxxxxx	0.0	EERRRR	6	PL2
50	xxxxxxxxxx	-0.2	RERE	4	
34	.xxxxxx	-0.4	RE	2	
44	.xxxxxxx	-0.6	EE	2	
38	.xxxxxxx	-0.8	RE	2	
35	xxxxxxx	-1.0	R	1	
12	.xx	-1.2	R	1	
8	.x	-1.4	-	0	
14	.xx	-1.6	-	0	
11	.xx	-1.8	ERE	3	
9	.x	-2.0	-	0	
14	.xx	-2.2	-	0	
17	.xxx	-2.4	-	0	
0	-	-2.6	-	0	
0	-	-2.8	-	0	
9	.x	-3.0	-	0	
0	-	-3.2	-	0	
0	-	-3.4	-	0	
18	.xxx	-3.6	-	0	
0	-	-3.8	-	0	
22	.xxxx	-4.0	-	0	

*For each bin in the Students' Ability Estimate column, "x" represents 5 students, "." represents a value in between 1 and 4 students, and no students are denoted as "-".

**For each bin in the item difficulty distribution column, "R" represents a receptive item, "E" represents an expressive item, and no items are denoted as "-".

The other item-person maps for other grade levels follow the same general pattern. Item difficulties tend to be clustered toward the middle of the scale with most items clustered in Level 2 and the top portion of Level 1. This corresponds to the panelists' results for item-level linguistic complexity ratings. Panelists rated most of the items in the lower part of the linguistic complexity scale (Low or Medium), and very few items were rated as "High" linguistic complexity. Note that while the linguistic complexity scale has three levels (Low, Medium, and High) and there are three performance levels for the Summative Alternate ELPAC (1, 2, and 3), those levels were not designed to correspond (e.g., an item with a linguistic complexity of Low does not necessarily correspond to performance Level 1). However, our linguistic complexity results are consistent with item difficulty results.

Criterion 3: Range Adequacy

This criterion has three components that help evaluate how well the Summative Alternate ELPAC addresses the full range of content described by the ELD Connectors. The ELD Connectors are arranged according to three parts, which include Interacting in Meaningful ways, Learning about how English works, and Foundational Literacy Skills; by three Modes, which include Collaborative, Interpretive, and Productive; and by three Language Processes, which include Structuring Cohesive Texts, Expanding and Enriching Ideas, and Connecting and Condensing Ideas. For the assessments to fully represent the ELD Connectors, items should address each of the Parts, each of the Modes, and each of the Language Processes.

The three components of the Range Adequacy criterion are summarized below.

- Each of the Parts (Interacting in Meaningful Ways, Learning About How English Works, and Foundational Literacy Skills) is measured by at least one item.
- Each of the three Modes (Collaborative, Interpretive, and Productive) is measured by at least one item.
- Each of the three Language Processes (Structuring Cohesive Texts, Expanding and Enriching Ideas, and Connecting and Condensing Ideas) is measured by at least one item.

Tables 3.9 through 3.11 present the results (counts of items aligned) for the three components of the Range Adequacy criterion across the grade levels/grade spans, and table 3.12 presents a summary of results. For the Range Adequacy criterion to be fully met, all components had to be acceptable or partially acceptable (see table 3.1 on page 3–24). Grade one partially met and all other test forms fully met this criterion. All test forms except for grade 3–5 Form 2 were acceptable for the first component and included items measuring each Part. Grade 3–5 Form 2 was partially acceptable because it had no item linked to a Part III: Foundational Literacy Skills Connector. It is important to note that there is only a single Part III Connector. All test forms were acceptable for the second component and included items measuring each Mode. All test forms but grade one were acceptable or partially acceptable for the third component

and included items measuring each Language Process. Across the grade levels/grade spans, the largest number of items were aligned to ELD Connectors in Part I: Interacting in Meaningful Ways. Similarly, the largest number of items across grade levels/grade spans were aligned to ELD Connectors in the Interpretive Mode.

Table 3.9 Results for Range Adequacy Criterion: Parts

Part	K	1	2	3–5 Form 1	3–5 Form 2	6–8 Form 1	6–8 Form 2	9–12 Form 1	9–12 Form 2
Part I: Interacting in Meaningful Ways	23	24	21	21	22	22	22	23	24
Part II: Learning about How English Works	12	3	14	10	8	8	7	9	7
Part III: Foundational Literacy Skills	NA	NA	1	1	0	NA	NA	NA	NA

Note. Counts of items include both primary and secondary alignments. A component is partially acceptable if two Parts were measured by at least one item.

Table 3.10 Results for Range Adequacy Criterion: Modes

Mode	K	1	2	3–5 Form 1	3–5 Form 2	6–8 Form 1	6–8 Form 2	9–12 Form 1	9–12 Form 2
Collaborative	6	3	6	2	5	7	9	9	8
Interpretive	17	17	13	15	14	13	13	14	14
Productive	6	11	7	10	11	9	9	5	6

Note. Counts of items include both primary and secondary alignments.

Table 3.11 Results for Range Adequacy Criterion: Language Processes

Language Process	K	1	2	3–5 Form 1	3–5 Form 2	6–8 Form 1	6–8 Form 2	9–12 Form 1	9–12 Form 2
Structuring Cohesive Texts	5	3	7	3	3	5	3	5	2
Expanding and Enriching Ideas	7	0	7	7	5	2	3	4	4
Connecting and Condensing Ideas	0	0	1	1	1	1	1	1	1

Note. Counts of items include both primary and secondary alignments. A component is partially acceptable if two Language Processes were measured by at least one item.

Table 3.12 Summary of Results for Range Adequacy Criterion

Grade Level/ Grade Span/ Test Form	Acceptable for Parts?	Acceptable for Modes?	Acceptable for Language Processes?	Criterion Result
K	Yes	Yes	Partially	Met
1	Yes	Yes	No	Partially Met
2	Yes	Yes	Yes	Met
3–5 Form 1	Yes	Yes	Yes	Met
3–5 Form 2	Partially	Yes	Yes	Met
6–8 Form 1	Yes	Yes	Yes	Met
6–8 Form 2	Yes	Yes	Yes	Met
9–12 Form 1	Yes	Yes	Yes	Met
9–12 Form 2	Yes	Yes	Yes	Met

Criterion 4: Balance-of-Knowledge Correspondence

This criterion includes three components and is evaluated based on the number of items panelists rate as directly and clearly matched to each ELD Connector dimension: Part (e.g., Interacting in Meaningful Ways), Mode (e.g., Productive), or Language Process (e.g., Structuring Cohesive Texts). The Summative Alternate ELPAC Balance-of-Knowledge Correspondence index is computed separately for each component, following the approach used for Webb’s Balance-of Knowledge Representation index. The Balance-of-Knowledge Correspondence index builds upon the Range Adequacy criterion by quantifying the extent to which balance is achieved within the Parts, Modes, and Language Processes. The balance index is computed based on the total number of items that were matched to either Part, Mode, or Language Process. The index ranges from 0 to 1, with 1 indicating perfect balance (e.g., if there were exactly the same number of items for each of the three Modes). Each component of the criterion is considered Acceptable if the calculated balance index is 0.70 or higher (Webb, 2007). To fully meet the Balance of Knowledge criterion, all three components must be rated Acceptable. To partially meet the criterion, two of the components must be rated Acceptable (see table 3.1 on page 3–24). Tables 3.13 through 3.20 provide the balance indexes for each Summative Alternate ELPAC grade and grade-span test form, and table 3.22 presents a summary of the results for the Balance-of-Knowledge Correspondence criterion.

Table 3.13 Results for Balance of Knowledge Correspondence: Kindergarten

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.84	Yes
Modes	0.75	Yes
Language Processes	0.67	No

Table 3.14 Results for Balance of Knowledge Correspondence: Grade One

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.61	No
Modes	0.76	Yes
Language Processes	0.33	No

Table 3.15 Results for Balance of Knowledge Correspondence: Grade Two

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.69	No
Modes	0.83	Yes
Language Processes	0.73	Yes

Table 3.16 Results for Balance of Knowledge Correspondence: Grades Three through Five Form 1

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.68	No
Modes	0.74	Yes
Language Processes	0.70	Yes

Table 3.17 Results for Balance of Knowledge Correspondence: Grades Three through Five Form 2

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.60	No
Modes	0.83	Yes
Language Processes	0.78	Yes

Table 3.18 Results for Balance of Knowledge Correspondence: Grades Six through Eight Form 1

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.77	Yes
Modes	0.89	Yes
Language Processes	0.71	Yes

Table 3.19 Results for Balance of Knowledge Correspondence: Grades Six through Eight Form 2

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.74	Yes
Modes	0.91	Yes
Language Processes	0.81	Yes

Table 3.20 Results for Balance-of-Knowledge Correspondence: Grades Nine through Twelve Form 1

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.78	Yes
Modes	0.83	Yes
Language Processes	0.77	Yes

Table 3.21 Results for Balance-of-Knowledge Correspondence: Grades Nine through Twelve Form 2

ELD Connector Dimension	Balance Index	Acceptable?
Parts	0.73	Yes
Modes	0.83	Yes
Language Processes	0.76	Yes

Table 3.22 Summary of Results for Balance-of-Knowledge Correspondence Criterion

Grade Level/ Grade Span/Test Form	Acceptable for Parts?	Acceptable for Modes?	Acceptable for Language Processes?	Criterion Result
K	Yes	Yes	No	Partially Met
1	No	Yes	No	Not Met
2	No	Yes	Yes	Partially Met
3–5 Form 1	No	Yes	Yes	Partially Met
3–5 Form 2	No	Yes	Yes	Partially Met
6–8 Form 1	Yes	Yes	Yes	Met
6–8 Form 2	Yes	Yes	Yes	Met
9–12 Form 1	Yes	Yes	Yes	Met
9–12 Form 2	Yes	Yes	Yes	Met

Overall, four test forms met the Balance-of-Knowledge Correspondence criterion (6–8 Forms 1 and 2, 9–12 Forms 1 and 2). For the first component, Parts, the grade two and both grade three through five forms (Form 1 and 2) did not meet the minimum balance value and thus partially met this criterion overall (acceptable for Modes and Language Processes). This is mainly a function of the fact that these are the only grades for which Part III is included in the blueprint. Because there is only one Part III Connector, it stands to reason that fewer items would be aligned to Part III. Meeting the balance criterion is thus more challenging for these grades. All grade levels/grade spans met the minimum balance value for the second component, Modes. For the third component, Language Processes, kindergarten did not meet the minimum balance value and thus partially met this criterion overall (acceptable for Parts and Modes). If one item had been rated as measuring a Connecting and Condensing Ideas Connector, the balance index would have been met for Language Processes in kindergarten. Grade one did not meet the minimum balance value for two components (Parts and Language Processes) and thus did not meet this criterion overall (only acceptable for Modes).

Summary and Discussion

Summary Results

Table 3.23 summarizes the alignment criteria results for the Summative Alternate ELPAC for all grade levels/grade spans. These results show that the Summative Alternate ELPAC items are linked to ELD Connectors across all assessments, although grade one panelists did not indicate any secondary connectors not already identified as a primary connector for other items. The Linguistic Complexity Adequacy criterion showed that most Summative Alternate ELPAC items measure the Medium linguistic complexity level. The Summative Alternate ELPAC tended to do a good job addressing the organization of the ELD Connectors, including items measuring all Parts, Modes, and Language Processes in a balanced way.

Each criterion for each grade level/span/form reported in table 3.23 is labeled as “Met,” “Not Met,” or “Partially Met” (see table 3.1, page 3–24 for full descriptions). For example, the criterion Linguistic Complexity Adequacy is partially met for grade one because fewer than 20 percent of grade one items were rated Low, but at least 20 percent and no more than 50 percent of items were rated Medium or High.

Table 3.23 Summative Alternate ELPAC Alignment Results

Criterion	K	1	2	3–5 Form 1	3–5 Form 2	6–8 Form 1	6–8 Form 2	9–12 Form 1	9–12 Form 2
Link to Standards	Met	Met							
Linguistic Complexity Adequacy	Not Met	Partially Met	Partially Met	Not Met	Not Met	Not Met	Not Met	Partially Met	Met
Range Adequacy	Met	Partially Met	Met	Met	Met	Met	Met	Met	Met
Balance-of-Knowledge Correspondence	Partially Met	Not Met	Partially Met	Partially Met	Partially Met	Met	Met	Met	Met

Discussion

While the results presented above include several instances where the Summative Alternate 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 Summative Alternate 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, all grade level/grade span forms fully met the criterion. For Linguistic Complexity Adequacy, only the grade 9–12 Form 2 fully met the criterion. This was in part due to a relatively large number of items across the grade levels/spans rated at the Medium level. For grades one and two, similar percentages of items were rated as Medium and High, but too few items were rated as Low to fully meet the blueprint targets. For grade 9–12 Form 1, similar percentages of items were rated as Low and Medium, but too few items were rated as High to fully meet the blueprint targets.

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 Alternate ELPAC. Most of the items are clustered in scoring Levels 1 and 2, while students' scores are clustered around Level 2, but with substantial numbers of students in both Levels 1 and 3. There were very few items (1 or 2 for some grades or grade spans) with difficulties in Level 3. While the panelists' linguistic complexity results cannot be directly compared to the scoring results (the two data sources refer to different, but correlated, item characteristics), both data sources indicate a lack of higher-level items.

The Range Adequacy criterion was fully met for all but grade one, which partially met this criterion. For grade one, this was due to its not meeting the Language Processes component of the criterion (i.e., no item was linked to Expanding and Enriching Ideas or Connecting and Condensing Ideas). Also, although grade one panelists identified ELD Connectors as secondary alignments for some items, all of these connectors had been identified already as primary alignments for other items. See tables 3.9 through 3.11 for results for each component of the Range Adequacy criterion.

The Balance-of-Knowledge Correspondence criterion was at least partially met for all grade levels/grade spans except grade one. For grade one, this was due to panelists aligning a large number of items to one Connector (PI.B.5) and not rating any items as measuring either of two Language Processes, Expanding and Enriching Ideas or Connecting and Condensing Ideas. Also, although grade one panelists identified ELD Connectors as secondary alignments for some items, all of these connectors had been identified already as primary alignments for other items. For the grade two and grade 3–5 forms, the criterion was partially met due to Part III: Foundational Literacy Skills being included on the blueprint for these tests, but only one item being aligned to a Part III Connector. For kindergarten, the criterion was partially met due to no item being rated as measuring the Language Process Connecting and Condensing Ideas.

This page is intentionally blank.

Chapter 4: Conclusions and Recommendations

This study combined documentation review and item ratings by content experts to evaluate the alignment between the Summative Alternate ELPAC and the California ELD Connectors. The documentation review was successful. Of the 13 identified *Testing Standards* we reviewed in respect to the alignment study, all were rated as fully covered based on the available evidence. These results indicate that the Summative Alternate ELPAC test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to English Language Development (ELD) Connectors.

Data from the alignment workshop component of the study provides support for the alignment of the Summative Alternate ELPAC to California’s ELD Connectors. First, all items on the Summative Alternate ELPAC, across all grade levels/grade spans were rated as aligned to ELD Connectors by panelists. No items were flagged for poor quality or as outside the measurement construct. This represents strong evidence that the Summative Alternate ELPAC does reflect its intended construct. Item ratings indicate that the test forms reflect the organization of the ELD Connectors, covering the ranges of Parts, Modes, and Language Processes, and maintaining balance among the number of items within each Part, Mode, and Language Process. There are some areas for improvement in terms of the linguistic complexity of items, as a large percentage of items were categorized as Medium by panelists. Future item development should ensure an adequate number of items at each level.

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 2021–2022 administration that includes both reliability estimates and classification accuracy results (CDE, 2023b). 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 Alternate ELPAC ranged from 0.86 to 0.88. Classification accuracy ranged from 0.70 to 0.90 (Alternate ELPAC 2021-2022 Technical Report: Appendices, p. 9). Though ETS split the high school grade span (9–12) into two grade spans (9–10 and 11–12) for conducting these analyses, splitting the grade spans should not impact the overall conclusions about reliability and classification accuracy.

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

The grade one panel was the only panel that did not identify any items as measuring a secondary connector that they had not already identified as a primary connector for other items. The panel’s secondary connector ratings thus did not expand the range of ELD Connectors measured by grade one items. Panelists also rated a large number of items as primarily aligned to one ELD Connector (PI.B.5). This may indicate an issue

with the panelists or an issue with the grade one test items. We recommend reviewing these items to determine if there really is a concentration of items measuring this ELD Connector, and/or if other ELD Connectors are being measured via a secondary alignment.

Recommendation 2. Review the linguistic complexity of items at all grade levels to determine if developing additional items at the Low and High linguistic complexity levels is necessary.

Most grade level/grade span test forms did not fully meet the Linguistic Complexity Adequacy criterion. That criterion established linguistic complexity level targets based on the number of each task type presented in the test blueprint. Failure to fully meet the criterion was typically due to a large percentage of items rated at the Medium level (up to 70.8% of items in some forms) and smaller percentages of items rated at the Low (down to 12.5% of items in some forms) and High levels (down to 8.3% of items in some forms).

Analysis of item-person (Wright) maps provided by ETS offers some support for the panelists' findings. Specifically, there are few items on the Summative Alternate ELPAC with difficulties in the part of the scale associated with the top score category (Level 3). Assuming that items with difficulties at the Level 3 performance level require students to demonstrate skills on more linguistically complex content, developing more items at the High linguistic complexity level may be warranted.

Item metadata from ETS include the linguistic complexity level to which each item was written, based on the task type. CDE should consider reviewing the metadata linguistic complexity categorizations in conjunction with panelists' linguistic complexity consensus ratings to inform the levels, if any, at which additional item development is needed.

Recommendation 3. Refine the test blueprint to specify the number and/or percentage of items at each linguistic complexity level.

It would be beneficial to adjust the test blueprint to include more precise targets for the distribution of linguistic complexity. Currently, the blueprint outlines the number of items representing each task type. Task types may be written to one or more linguistic complexity levels (i.e., Low, Low to Medium, Medium, Medium to High, High). Clearly specified linguistic complexity targets by grade level/grade span test form can better inform item development goals, ensuring that the item bank contains an adequate number of items at each linguistic complexity level and supporting the construction of future test forms that reflect both the breadth and depth of the ELD Connectors.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- California Department of Education. (2012). *California English Language Development Standards (Electronic Edition): Kindergarten through Grade 12*.
<https://www.cde.ca.gov/sp/el/er/documents/eldstndpublication14.pdf>.
- California Department of Education. (2019). *California Practitioners' Guide for Educating English Learners with Disabilities*.
<https://www.cde.ca.gov/sp/se/ac/documents/ab2785guide.pdf>
- California Department of Education. (2021). *English Language Development Connectors for the Alternate ELPAC*
<https://www.cde.ca.gov/ta/tg/ep/documents/eldconnectorsaltelpac.docx>
- California Department of Education (2022). *2022–23 English Language Proficiency Assessments for California Information Guide*.
<https://www.cde.ca.gov/ta/tg/ep/documents/elpacinfo22.docx>
- California Department of Education (2023a). *DataQuest* web-based reporting system.
<https://www.cde.ca.gov/ds/ad/dataquest.asp> (accessed January 19, 2023)
- California Department of Education (2023b). *Alternate English Language Proficiency Assessments for California 2021–2022 Technical Report*.
- Cook, H.G. (2006). Aligning English Language Proficiency Tests to English Language Learning Standards. In *Aligning Assessment to Guide the Learning of All Students*. Council of Chief State School Officers.
- 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*. University of Wisconsin.
- Every Student Succeeds Act, 114–95 C.F.R. (2015).
- National Center for Education Statistics. (2022). *English Learners in Public Schools. Condition of Education*. <https://nces.ed.gov/programs/coe/indicator/cqf>
- National Center for Educational Statistics. (n.d.). *Fast Facts: English language learners*.
<https://nces.ed.gov/fastfacts/display.asp?id=96>
- Shyyan, V. V., & Christensen, L. L. (2018). A Framework for Understanding English Learners with Disabilities: Triple the Work (ALTELLA Brief No. 5). *Alternate English Language Learning Assessment Project*.
https://altella.wceruw.org/pubs/ALTELLA_Brief-05_Framework.pdf

- 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. <https://eddataexpress.ed.gov/>
- U.S. Department of Education (2021a). IDEA Section 618 Data Products: Static Tables. <https://www2.ed.gov/programs/osepidea/618-data/static-tables/index.html>
- U.S. Department of Education (n.d.). *Our Nation's English Learners*. <https://www2.ed.gov/datastory/el-characteristics/index.html>
- U.S. Department of Education. (2021b). Understanding the Assessment Peer Review Process. <https://oese.ed.gov/files/2021/07/webinar-1-july-15-2021.pdf>
- 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). 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). University of Wisconsin–Madison, National Institute for Science Education.
- Webb, N. L. (2006). “Identifying content for student achievement tests.” In *Handbook of test development*, edited by S. Downing & T. Haladyna. Erlbaum.
- Webb, N. L. (2007). Issues Related to Judging the Alignment of Curriculum Standards and Assessments. *Applied Measurement in Education*, 20 (1),7–25. <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
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
ELSWD	English Learner Student with Disabilities
ESEA	Elementary and Secondary Education Act
ETS	Educational Testing Service
IEP	Individualized Education Program
ESSA	Every Student Succeeds Act
NCES	National Center for Education Statistics
PLD	Proficiency Level Descriptor (for ELD Standards)

This page is intentionally blank.

Appendix A: Summative Alternate ELPAC Documentation Reviewed by HumRRO

Table A.1. Summative Alternate ELPAC Documents Reviewed

Document Focus	Document File Name
English Language Development Connectors	<ul style="list-style-type: none"> • 041619-01 V2 FOR ARCHIVE Alternate ELPAC ELD Connectors Report _050819 • ELD Standards CDE Publication14.pdf
Test Design	<ul style="list-style-type: none"> • 357-2019 FOR ARCHIVE Alternate ELPAC High-Level Test Design_052819.pdf • Alternate ELPAC Test Blueprint - ELPAC (CA Dept of Education) • AltELPACblueprint.pdf • Proposed High-Level Test Design for the Alternate ELPAC • 23R-673-1095 FOR ARCHIVE Alt ELPAC Connectors Review Mtg Participants Balance Data 011123.xlsx • 23R-673-1095 FOR ARCHIVE Alt ELPAC Connectors Review Mtg Participants Table 011123.docx
Item Development and Information	<ul style="list-style-type: none"> • 021318-01 v2 FOR ARCHIVE CAASPP Item Acceptance Criteria for IRC 022118_ • 362-2020 V3 FOR ARCHIVE Alt ELPAC Item Development Specifications 121719 • 363-2019A V3 FOR ARCHIVE Alternate ELPAC Item Review Meeting Plan • 363-2019B V3 FOR ARCHIVE Alt-ELPAC Item Review Meeting Slides_071219 • 438-2020 v4 Alt ELPAC Task Type Specifications Preface_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Communicate Familiar Topics_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Describe a Routine_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Informational Text_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Literary Text_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Opinion_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_Recognize Use Common Words_011420 • 438-2020 v4 Alt-ELPAC Task Type Specs_School Activity_011420

Table A.1. (cont.)

Document Focus	Document File Name
Item Development and Information (cont.)	<ul style="list-style-type: none"> • 438-2020 v4 Alt-ELPAC Task Type Specs_School Exchange_011420 • 441-2020 V3 FOR ARCHIVE Alt ELPAC IWW PPT 021320 • 711-2021-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW General PPT_022120 • 711-2021A-v4_FOR ARCHIVE_CAA EMS Alt ELPAC IWW Plan_022720 • Alt-ELPAC_IRM_Guiding Questions • 439-2020B V5 FOR ARCHIVE Alt-ELPAC Pilot Cognitive Lab Report_100620
Test Administration	<ul style="list-style-type: none"> • ALT-ELPAC--Practice-Test-DFA-Grade-K.2021-22 • ALT-ELPAC--Training-Test-DFA-Grades-K-5.2021-22
Item Scoring	<ul style="list-style-type: none"> • ALT-ELPAC--Practice-Test-Scoring-Guide-Grade-K.2020-21
Score Reporting	<ul style="list-style-type: none"> • 359-2021F V3 FOR APPROVAL Alt ELPAC Range PLDs G1 030921 • Reliability Chapter.docx • 446-2022 v1 FOR REVIEW Alt ELPAC Technical Report-appendix 8.110322.docx
Accessibility	<ul style="list-style-type: none"> • Universal design for item development
Field Test	<ul style="list-style-type: none"> • 440-2021 v4 FOR ARCHIVE Alt-ELPAC Field Test Specifications 082520
Standard Setting	<ul style="list-style-type: none"> • 445-2022D v3 FOR ARCHIVE Alt ELPAC Standard Setting Technical Report 072022

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

Appendix B: Materials for Virtual Alignment Workshop

List of Materials

- Virtual Workshop Agenda (August 1–3, 2022)
- Panelist Instructions
- Sample Panelist Rating Form
- Linguistic Complexity Rating Aid
- Summative Alternate ELPAC Alignment Overall Debrief Survey and Evaluation
- Summative Alternate ELPAC Alignment Panelist Demographic Information Survey

Virtual Workshop Agenda (August 1–3, 2022)

Day 1 - Monday, August 1, 2022 (All Panels)

8:00 – 8:45 a.m. Join Teams Meeting with All Panelists, HumRRO Facilitators, and California Department of Education Staff

Welcome, logistics, overview of Alt ELPAC, general alignment training

8:45 – 9:00 a.m. BREAK

9:00 – 10:30 a.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 Alt ELPAC items

- Identify Primary ELD Connector and Secondary ELD Connector, if applicable
- Assign linguistic complexity rating

Begin iterative alignment rating process:

- Independent rating
- Discussion and consensus building
- Panel review of metadata
- Final independent and consensus ratings

10:30 – 10:45 p.m. Break
10:45 – 12:00 p.m. Continue iterative alignment rating process
12:00 – 12:45 p.m. Lunch Break
12:45 – 2:30 p.m. Continue iterative alignment rating process
2:30 – 2:45 p.m. Break
2:45 – 4:00 p.m. Continue iterative alignment rating process

Day 2 - Tuesday, August 2 (All Panels)

8:00 – 10:30 a.m. Join Teams Meeting for Assigned Grade Level Panel
Continue iterative alignment rating process
10:30 – 10:45 p.m. Break
10:45 – 12:00 p.m. Continue iterative alignment rating process
12:00 – 12:45 p.m. Lunch Break
12:45 – 3:30 p.m. Continue and complete iterative alignment rating process
3:30 – 3:45 p.m. Break
3:45 – 4:00 p.m. Complete two short online surveys (all but Grade K and 2 Panel):

- Debrief/ Workshop evaluation
- Demographic information

Panels for Kindergarten, Grade 1, Grades 3–5, Grades 6–8, and Grades 9–12 Adjourn

Day 3 - Wednesday, August 3 (Grade 2 Panel Only)

8:00 – 10:30 a.m. Join Teams Meeting for Assigned Grade Level Panel
Continue iterative alignment rating process
10:30 – 10:45 p.m. Break
10:45 – 12:00 p.m. Continue iterative alignment rating process
12:00 – 12:45 p.m. Lunch Break
12:45 – 2:30 p.m. Continue iterative alignment rating process
2:30 – 2:45 p.m. Break
2:45 – 3:45 p.m. Continue and complete alignment rating process
3:45 – 4:00 p.m. Complete two short online surveys:

- Debrief/ Workshop evaluation
- Demographic information

Adjourn

Panelist Instructions

No.	Rating Documents & Tools	Format/Location
1	Summative Alternate ELPAC Panelist Instructions	Read ahead packet
2	Content Review Tool (CRT) <ul style="list-style-type: none"> • Directions for Administration (DFA) • Secure Summative Alternate ELPAC items 	Online via ETS-provided link
3	Summative Alternate ELPAC Rating Form	Online via HumRRO-provided link
4	CA English Language Development (ELD) Connectors	Read ahead packet
5	Linguistic Complexity Rating Aid	Read ahead packet
6	Post-Ratings Activity – Online Survey: Debriefing/Evaluation Form	HumRRO-provided link
7	Post-Ratings Activity – Online Survey: Demographic Questionnaire	HumRRO-provided link

Technology check prior to Workshop (by 7/29): Confirm access to CRT & Teams meeting.

Prior to alignment ratings:

1. Introductions
2. Review of MS Teams features and settings, use of Hand Raising & Chat
3. Review of panelist materials, with discussion of ELD Connectors and Linguistic Complexity Rating Aid
4. Familiarization with logging in and navigating in Content Review Tool (CRT)
 - a. Secure Resource, Directions for Administration (DFA)
 - b. Secure items to be rated.

Rate Summative Alternate ELPAC Items

Orient to Rating Form:

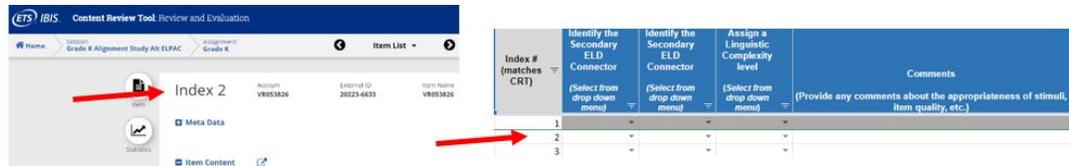
1. You will review several Summative Alternate ELPAC items and will enter the ELD Connector rating(s) and linguistic complexity level rating for each item.
2. Access Summative Alternate ELPAC Rating Form, a panelist-specific Google Sheet file:
 - a. Click link in HumRRO Final Instructions email
 - b. Your Google Sheet will save automatically, no manual saving needed.
3. Review rating categories on rating form
 - a. Confirm your rating form matches the grade level of your panel.
 - b. You will only need to work in the first tab. The other tab is for internal use only.

C	D	E	F	G	H	I	J	K
Item Description	Item Type	Max Points	Mode of Communication	Index # (match CRT)	Identify the Primary ELD Connector (Select from drop down menu)	Identify the Secondary ELD Connector (Select from drop down menu)	Assign a Linguistic Complexity level (Select from drop down menu)	Comments (Provide any comments about the appropriateness of stimuli, item quality, etc.)
	MCSS - Member	1	Receptive	2				
	InlineChoiceList55 - Member	2	Expressive	3				

- c. Columns A through F contain information about each Summative Alternate ELPAC item. Column A (hidden) provides the ETS unique item identifier. Column B (hidden) provides the sequence number that corresponds to the item number on the test form. Column C provides a brief description of the item (masked in light blue here for security). Column D provides the item type. Column E provides the maximum number of points possible on the item. Column F provides the mode of communication.
- d. Column G provides the index number, which corresponds to the number of the item or set leader in the Content Review Tool (CRT). You will rate only items, not set leaders. Rows for set leader are greyed out in Columns H through K.
- e. Column H asks for the primary ELD Connector measured by the item. An example Connector code is P1.A.1. All Connector codes for the grade level/span being rated are presented in a dropdown menu that is accessible by clicking on the arrow that appears to the right of the cell where the rating is to be made. The last option in the menu is “None.”
- f. Column I asks for the secondary ELD Connector measured by the item. Not all items measure a secondary Connector. The dropdown menu of secondary Connector codes is identical to the one for primary Connectors and is accessible by clicking on the arrow that appears to the right of the cell where the rating is to be made. The last option in the menu is “None.”
- g. Column J asks for the linguistic complexity level associated with answering the item correctly or scoring full points on an item that is worth more than one point. The linguistic complexity levels (Low, Medium, High) are presented in a dropdown menu that is accessible by clicking on the arrow that appears to the right of the cell where the rating is to be made.
- h. Column K is available for entering any comments or notes to clarify or qualify any of the other ratings. You may enter comments on set leaders, even though the cell is greyed out.

Make item ratings:

1. Rate the first item independently, all relevant columns.
 - a. Navigate to the first **item** index number in the CRT and confirm that it matches the first index number for an **item** on the rating form (Index 2 in the images below). When the item type is “Set Leader” (Index 1 in the image on the right below) there will be no ratings made (rating cells are greyed out). This is a stimulus that subsequent items refer to. You will review the content of the set leader when you review the item, and you can make comments if you have concerns about the set leader’s quality.



- b. Review the content of the item as well as any related directions, scoring rubrics, or picture cards presented in the DFA.
- c. Review the ELD Connectors.
- d. Using the dropdown menu, rate the primary ELD Connector measured by the item. If you don't think any ELD Connector is measured by the item, choose "None" (last option in dropdown menu).
- e. If applicable, rate the secondary ELD Connector measured by the item, using the dropdown menu. If you don't think a secondary ELD Connector is measured by the item, choose "None" (last option in dropdown menu).
- f. Review the Linguistic Complexity Rating Aid to help you determine the level associated with answering the item correctly (or scoring full points on an item that is worth more than one point).
- g. Using the dropdown menu, rate the linguistic complexity level of the item.
- h. Provide comments, as needed.
 - i. All items have been thoroughly reviewed. Comments are not required, but you may choose to enter comments to provide context for your ratings, list other Connector(s) you strongly considered, or if you do notice an issue related to the quality of the item.

2. After all panelists have rated the first item, the group will discuss their independent ratings. 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.
 - a. Repeat at least 3 times, one item at a time, as instructed by the HumRRO facilitator.
 - b. Panelists should not change ratings after discussion and review unless they are **certain** they made an error (e.g., coding error). Do NOT change independent ratings after discussion.
3. The HumRRO facilitator will next share the item metadata (test developer's assigned ELD Connector(s) and linguistic complexity). 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. Rate all remaining Summative Alternate ELPAC items independently in sets of 4–8 items before discussing and settling on consensus. The HumRRO facilitator will instruct the group on the set of items to be rated. Repeat the process above for each set of items.
5. 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, you will be given a link to complete (anonymously) the following two online questionnaires:
 - a. Debriefing/Evaluation Form
 - b. Demographic Questionnaire

Sample Screenshots of a Panelist Rating Form

Panelists will be given a link to their individual Google Sheet to enter their independent ratings and comments about the Summative Alternate ELPAC items they review.

- Panelists will rate every item on its Primary ELD Connector.
- If applicable for the item, panelists will also rate the Secondary ELD Connector.
- Panelists will use item content as well as item administration and scoring information (from the Directions for Administration, DFA) to assign a Linguistic Complexity Level to each item.

The screen shots below illustrate that panelists will use drop down menus to make their ratings. A comments field is available for all items but is not shown below.

Item Type	Max Points	Mode of Communication	Index # (matches CRT)	Identify the Primary ELD Connector (Select from drop down menu)	Identify the Secondary ELD Connector (Select from drop down menu)	Assign a Linguistic Complexity level (Select from drop down menu)
Set Leader - Leader	~	~	1			
MCSS - Member	1	Receptive	2			
InlineChoiceListSS - Member	2	Expressive	3			
Set Leader - Leader	~	~	4			
MCSS - Member	1	Receptive	5			
InlineChoiceListSS - Member	2	Expressive	6			
Set Leader - Leader	~	~	7			
MCSS - Member	1	Receptive	8			
MCSS - Member	1	Receptive	9			
MCSS - Member	1	Receptive	10			
MCSS - Member	1	Expressive	11			

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

Item Type	Max Points	Mode of Communication	Index # (matches CRT)	Identify the Primary ELD Connector (Select from drop down menu)	Identify the Secondary ELD Connector (Select from drop down menu)	Assign a Linguistic Complexity level (Select from drop down menu)
Set Leader - Leader	~	~	1			
MCSS - Member	1	Receptive	2			
InlineChoiceListSS - Member	2	Expressive	3			

Click the ▼ button to reveal the linguistic complexity levels. Rater selects the linguistic complexity level that

Linguistic Complexity Rating Aid Excerpt (Grade Two)

Low	Medium	High
<ul style="list-style-type: none"> • With prompting and support, use a very limited set of strategies to <ul style="list-style-type: none"> ○ Identify a few key words and phrases from read-alouds, simple written texts, and oral presentations. ○ answer yes/no questions about key details from read-alouds and oral presentations. • With prompting and support, <ul style="list-style-type: none"> ○ listen with occasional participation in short conversations. ○ respond to simple yes/no questions about familiar topics. 	<ul style="list-style-type: none"> • With prompting and support, use an emerging set of strategies to <ul style="list-style-type: none"> ○ identify the main topic or characters from read-alouds, simple written texts, and oral presentations. ○ Sequence information (beginning and end) from read-alouds, simple written texts, and oral presentations. ○ answer yes/no and simple wh-questions about key details from read-alouds and oral presentations. • With prompting and support, <ul style="list-style-type: none"> ○ communicate simple information (e.g., show and tell, tell about a familiar picture) about familiar texts topics, experiences, or events. ○ communicate simple information about an event or topic, with emerging control. ○ express a preference or opinion about a familiar topic or story. ○ provide one reason for the preference about a familiar topic or story. 	<ul style="list-style-type: none"> • With prompting and support, use an increasing range of strategies to <ul style="list-style-type: none"> ○ ask and answer questions about key details from read-alouds, picture books, simple written texts, and oral presentations. ○ identify the main idea from read-alouds, picture books, simple written texts, and oral presentations. ○ retell parts of a story from read-alouds, picture books, simple written texts, and oral presentations. • With prompting and support, <ul style="list-style-type: none"> ○ communicate a few pieces of information about a familiar topic. ○ tell or dictate information (e.g., recount an experience, retell a story, describe a picture) about familiar topics, texts, experiences, or events. ○ express a preference or an opinion about a familiar topic or story. ○ provide more than one reason for the preference or opinion about a familiar topic or story.

Summative Alternate ELPAC Alignment Overall Debrief Survey and Evaluation

The questions and response options below represent the content of an online survey HumRRO administered to the alignment workshop panelists on August 2 and 3, 2022.

1. Please enter the grade level or grade span of your panel for the Alt ELPAC alignment workshop.
 - Kindergarten
 - Grade 1
 - Grade 2
 - Grade 3–5
 - Grade 6–8
 - Grade 9–12

2. Overall, how well were the Alt ELPAC items aligned with the ELD Connectors?
 - Strongly aligned
 - Partially aligned
 - Not at all aligned

3. Please share any additional information on the alignment between the Alt ELPAC items and the ELD Connectors.

4. Indicate the extent to which you agree or disagree with each of the following statements: *

- 4a. The large-group training session effectively outlined the purpose of the alignment workshop.
 - Strongly Disagree
 - Disagree
 - Neither Agree nor Disagree
 - Agree
 - Strongly Agree

- 4b. The large-group training session provided a useful overview of the alignment activities for the workshop.
 - Strongly Disagree
 - Disagree
 - Neither Agree nor Disagree
 - Agree
 - Strongly Agree

4c. The large-group training session clearly described my role as a panelist.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

4d. The large-group training session was well organized.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

4e. The large group training was an effective use of time.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

5. Indicate the extent to which you agree or disagree with each of the following statements:

5a. The hands-on training in my panel helped me better understand the alignment activities.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

5b. Practicing making ratings as a group in my panel helped me better understand the alignment activities.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

5c. The panel-specific hands-on training was well organized.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

5d. The hands-on training in my panel was an effective use of time.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

6. Indicate the extent to which you agree or disagree with each of these statements:

6a. My panel facilitator clearly and promptly addressed my questions

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

6b. My panel facilitator did an effective job of facilitating discussion and ensuring that all panelist' perspectives were heard.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

7. Indicate the extent to which you agree or disagree with each of these statements:

7a. Everyone in my panel had equal opportunity to contribute ideas and opinions.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

7b. My ideas and opinions were listened to and respected by my panel.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

8. Indicate the extent to which you agree or disagree with each of the following statements:

8a. Communications from HumRRO prior to the workshop prepared me for participating.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

8b. The support materials provided to me before the workshop were useful (e.g., Linguistics Complexity Rating Aid).

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

8c. The Google rating sheet was useful for recording alignment ratings.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

9. Indicate the extent to which you agree or disagree with each of the following statements:

9a. It was easy to access the item content and DFA in the CRT.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

9b. The item content and DFA allowed me to effectively accomplish my tasks during the alignment workshop.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

9c. It was easy to access the online evaluation and demographics forms.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

10. Indicate the extent to which you agree or disagree with each of these statements:

10a. The large-group training facilitator was helpful during the workshop.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

10b. The panel facilitator was helpful during the workshop.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

10c. Other HumRRO researchers were helpful during the workshop.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

11. Please use this space for any additional comments you wish to share:

Summative Alternate 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 August 1, 2022.

1. What is your gender?
 - Female
 - Male
 - Non-binary
 - Prefer not to disclose
 - Other

2. What is your age (in years)?
 - 25 or under
 - 26–35
 - 36–45
 - 46–55
 - 56–65
 - 66 or over
 - Prefer not to disclose

3. Choose one ethnic identity
 - Hispanic/Latino
 - Not Hispanic/Latino

4. Choose one or more racial identities (regardless of ethnicity).
 - American Indian or Alaskan Native
 - Asian
 - Black or African American
 - Native Hawaiian or Pacific Islander
 - White
 - Prefer not to disclose
 - Other

5. What is your highest earned degree?
 - Associate’s Degree
 - Baccalaureate Degree
 - Master’s Degree
 - Ph.D. or equivalent (e.g., Ed.D.)
 - Other

6. Use the drop down menu to indicate how many years of teaching experience you have.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15 or more

7. Do you have experience working with students from diverse backgrounds?

- Yes
- No

8. You indicated you have experience working with students from diverse backgrounds. Please select all of the student groups with whom you have worked.

- English language learners
- Students of color
- Students with mild or moderate disabilities
- Students with the most significant cognitive disabilities
- Students from low socioeconomic households
- Students receiving free and/or reduced lunch
- Other

9. Please use this space for any additional comments you wish to share:

This page is intentionally blank.

Appendix C: Summary of Responses to Summative Alternate ELPAC Process Evaluation Survey Questions

Tables C.1 through C.4 summarize responses from 20 workshop panelists to the process evaluation questions of the Summative Alternate ELPAC Evaluation Survey.

Table C.1 Evaluation of Summative Alternate ELPAC Alignment Large Group Training

Evaluative Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
4a. The large-group training session effectively outlined the purpose of the alignment workshop.	5%	0%	0%	30%	65%
4b. The large-group training session provided a useful overview of the alignment activities for the workshop.	5%	0%	5%	25%	65%
4c. The large-group training session clearly described my role as a panelist.	5%	0%	5%	35%	55%
4d. The large-group training session was well organized.	5%	0%	0%	20%	75%
4e. The large-group training was an effective use of time.	5%	0%	15%	10%	70%
10a. The large-group training facilitator was helpful during the workshop.	5%	0%	5%	20%	70%

Table C.2 Evaluation of Hands-on Summative Alternate ELPAC Alignment Training

Evaluative Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
5a. The hands-on training in my panel helped me better understand the alignment activities.	5%	0%	0%	15%	80%
5b. Practicing making ratings as a group in my panel helped me better understand the alignment activities.	5%	0%	0%	15%	80%
5c. The panel-specific hands-on training was well organized.	5%	0%	0%	15%	80%
5d. The hands-on training in my panel was an effective use of time.	5%	0%	0%	15%	80%

Table C.3 Evaluation of Summative Alternate ELPAC Alignment Panel Facilitators

Evaluative Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
6a. My panel facilitator clearly and promptly addressed my questions.	5%	0%	0%	10%	85%
6b. My panel facilitator did an effective job of facilitating discussion and ensuring that all panelists' perspectives were heard.	5%	0%	0%	10%	85%
7a. Everyone in my panel had equal opportunity to contribute ideas and opinions.	5%	0%	0%	10%	85%
7b. My ideas and opinions were listened to and respected by my panel.	5%	0%	0%	5%	90%
10b. The panel facilitator was helpful during the workshop.	5%	0%	0%	15%	80%

Table C.4 Evaluation of Summative Alternate ELPAC Alignment Communications, Materials, and Processes

Evaluative Statement	% Strongly Disagree	% Disagree	% Neither Agree nor Disagree	% Agree	% Strongly Agree
8a. Communications from HumRRO prior to the workshop prepared me for participating.	5%	0%	0%	25%	70%
8b. The support materials provided to me before the workshop were useful (e.g., Linguistic Complexity Rating Aid).	5%	0%	5%	20%	70%
8c. The Google rating sheet was useful for recording alignment ratings.	5%	0%	0%	15%	80%
9a. It was easy to access the item content and DFA in the CRT.	5%	0%	0%	40%	55%
9b. The item content and DFA allowed me to effectively accomplish my tasks during the alignment workshop.	5%	0%	5%	20%	70%
9c. It was easy to access the online evaluation and demographics forms.	5%	0%	0%	10%	85%
10c. Other HumRRO researchers were helpful during the workshop.	5%	0%	20%	10%	65%

This page is intentionally blank.

Appendix D: Summative Alternate ELPAC Item-Person Maps

Tables D.1 through D.6 are called item-person maps and present a comparison of student ability and test item difficulty. Each table presents a grade level or grade span item-person map for the overall scores of the Summative Alternate ELPAC.

Both student ability and item difficulty are presented on the same scale, represented by the *Logit* column at the center of the map. These values are also referred to as *bins*. The number of students scoring at each ability level is presented on the left side of the table. The number of items at each difficulty level is presented on the right side of the table. The students at the top of the table earned the highest scores (highest ability students), while the items at the top of the table are the most difficult. The students at the bottom of the table earned the lowest scores (lowest ability students), and the items at the bottom of the table are easiest. There are horizontal lines representing each threshold score, the lowest score at which a student would be classified at the next higher overall performance level (i.e., solid line for Level 1 to 2 and dashed line for Level 2 to 3).

The Summative Alternate ELPAC performance levels are Novice English Learner (Level 1), Intermediate Learner (Level 2), and Fluent English Proficient (Level 3). Performance levels were identified during a standard setting process that was separate from this study.

Table D.1 Kindergarten Item-Person Map

Number of Students	Students*	Logit	Items**	Number of Items	
4	.	4.0	-	0	
0	-	3.8	-	0	
5	x	3.6	-	0	
0	-	3.4	-	0	
3	.	3.2	-	0	
0	-	3.0	-	0	
1	.	2.8	-	0	
7	.x	2.6	R	1	
7	.x	2.4	-	0	
8	.x	2.2	-	0	
21	.xxxx	2.0	R	1	
12	.xx	1.8	R	1	
26	.xxxxx	1.6	EE	2	PL3
25	xxxxx	1.4	ERE	3	
33	.xxxxxx	1.2	ERREEEE	7	
64	.xxxxxxxxxxxx	1.0	ERRE	4	
45	xxxxxxxxxx	0.8	ERER	4	
47	.xxxxxxxxxx	0.6	EREE	4	
47	.xxxxxxxxxx	0.4	EEERR	5	
34	.xxxxxx	0.2	EE	2	
50	xxxxxxxxxx	0.0	ERER	4	PL2
27	.xxxxx	-0.2	R	1	
26	.xxxxx	-0.4	-	0	
31	.xxxxxx	-0.6	E	1	
27	.xxxxx	-0.8	R	1	
15	xxx	-1.0	RE	2	
19	.xxx	-1.2	E	1	
20	xxxx	-1.4	-	0	
16	.xxx	-1.6	-	0	
0	-	-1.8	R	1	
16	.xxx	-2.0	-	0	
0	-	-2.2	-	0	
19	.xxx	-2.4	-	0	
0	-	-2.6	-	0	
0	-	-2.8	-	0	
0	-	-3.0	-	0	
29	.xxxxx	-3.2	-	0	
0	-	-3.4	-	0	
0	-	-3.6	-	0	
0	-	-3.8	-	0	
43	.xxxxxxxxx	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 5 students, “.” represents a value in between 1 and 4 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

Table D.2 Grade One Item-Person Map

Number of Students	Students*	Logit	Items**	Number of Items
10	XX	4.0	-	0
11	.XX	3.8	-	0
0	-	3.6	-	0
0	-	3.4	-	0
0	-	3.2	-	0
1	.	3.0	-	0
0	-	2.8	-	0
9	.X	2.6	-	0
0	-	2.4	-	0
18	.XXX	2.2	-	0
7	.X	2.0	-	0
20	XXXX	1.8	-	0
33	.XXXXXX	1.6	-	0
13	.XX	1.4	-	0
26	.XXXXX	1.2	RE	2
<hr/>				
34	.XXXXXX	1.0	EEEEER	7
46	.XXXXXXXXXX	0.8	EEE	3
85	XXXXXXXXXXXXXXXXXX	0.6	ERRE	4
59	.XXXXXXXXXXXX	0.4	ERRE	4
61	.XXXXXXXXXXXX	0.2	ERRR	4
79	.XXXXXXXXXXXXXXXXXX	0.0	EEERE	5
44	.XXXXXXXXXX	-0.2	RRERE	5
<hr/>				
47	.XXXXXXXXXX	-0.4	ER	2
38	.XXXXXXXXXX	-0.6	RRE	3
13	.XX	-0.8	R	1
23	.XXXX	-1.0	-	0
15	XXX	-1.2	-	0
27	.XXXXX	-1.4	ER	2
9	.X	-1.6	E	1
9	.X	-1.8	-	0
0	-	-2.0	R	1
15	XXX	-2.2	-	0
11	.XX	-2.4	-	0
0	-	-2.6	-	0
0	-	-2.8	-	0
12	.XX	-3.0	-	0
0	-	-3.2	-	0
0	-	-3.4	-	0
26	.XXXXX	-3.6	-	0
0	-	-3.8	-	0
32	.XXXXXX	-4.0	-	0

PL3

PL2

*For each bin in the theta distribution column, “X” represents 5 students, “.” represents a value in between 1 and 4 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

Table D.3 Grade Two Item-Person Map

Number of Students	Students*	Logit	Items**	Number of Items
4	.	4.0	-	0
6	.X	3.8	-	0
0	-	3.6	-	0
0	-	3.4	-	0
4	.	3.2	-	0
0	-	3.0	-	0
6	.X	2.8	-	0
0	-	2.6	-	0
13	.XX	2.4	-	0
15	XXX	2.2	-	0
19	.XXX	2.0	-	0
17	.XXX	1.8	E	1
16	.XXX	1.6	-	0
25	XXXXX	1.4	R	1
<hr style="border-top: 1px dashed black;"/>				
32	.XXXXXX	1.2	-	0
39	.XXXXXXX	1.0	E E E E	4
38	.XXXXXXX	0.8	R E R E E E E	7
51	.XXXXXXXXXX	0.6	E R	2
48	.XXXXXXXXXX	0.4	E R R R R E	6
51	.XXXXXXXXXX	0.2	E	1
54	.XXXXXXXXXX	0.0	E E R R R R	6
<hr style="border-top: 1px solid black;"/>				
50	XXXXXXXXXX	-0.2	R E R E	4
34	.XXXXXX	-0.4	R E	2
44	.XXXXXXXXXX	-0.6	E E	2
38	.XXXXXXX	-0.8	R E	2
35	XXXXXXX	-1.0	R	1
12	.XX	-1.2	R	1
8	.X	-1.4	-	0
14	.XX	-1.6	-	0
11	.XX	-1.8	E R E	3
9	.X	-2.0	-	0
14	.XX	-2.2	-	0
17	.XXX	-2.4	-	0
0	-	-2.6	-	0
0	-	-2.8	-	0
9	.X	-3.0	-	0
0	-	-3.2	-	0
0	-	-3.4	-	0
18	.XXX	-3.6	-	0
0	-	-3.8	-	0
22	.XXXX	-4.0	-	0

PL3

PL2

*For each bin in the theta distribution column, “X” represents 5 students, “.” represents a value in between 1 and 4 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

Table D.4 Grades Three through Five Item-Person Map

Number of Students	Students *	Logit	Items**	Number of Items
54	.XXXXX	4.0	-	0
0	-	3.8	-	0
0	-	3.6	-	0
20	XX	3.4	-	0
44	.XXXX	3.2	-	0
0	-	3.0	-	0
0	-	2.8	-	0
30	XXX	2.6	-	0
60	XXXXXX	2.4	-	0
47	.XXXX	2.2	-	0
96	.XXXXXXXXXX	2.0	-	0
48	.XXXX	1.8	E	1
46	.XXXX	1.6	-	0
91	.XXXXXXXXXX	1.4	E	1
185	.XXXXXXXXXXXXXXXXXXXX	1.2	E	1
80	XXXXXXXXXX	1.0	E	1
<hr/>				
158	.XXXXXXXXXXXXXXXXXXXX	0.8	EE	2
72	.XXXXXXXXXX	0.6	ERRR	4
149	.XXXXXXXXXXXXXXXXXXXX	0.4	E	1
158	.XXXXXXXXXXXXXXXXXXXX	0.2	RERRE	5
174	.XXXXXXXXXXXXXXXXXXXX	0.0	EEERE	5
157	.XXXXXXXXXXXXXXXXXXXX	-0.2	ERRERERER	9
169	.XXXXXXXXXXXXXXXXXXXX	-0.4	REEEE	5
220	XXXXXXXXXXXXXXXXXXXX	-0.6	REEERE	6
<hr/>				
139	.XXXXXXXXXXXXXXXXXXXX	-0.8	RRRRER	6
131	.XXXXXXXXXXXXXXXXXXXX	-1.0	EEE	3
120	XXXXXXXXXXXXXXXXXXXX	-1.2	EER	3
72	.XXXXXXXXXX	-1.4	R	1
66	.XXXXXX	-1.6	R	1
38	.XXX	-1.8	ER	2
46	.XXXX	-2.0	ERR	3
13	.X	-2.2	RE	2
29	.XX	-2.4	-	0
12	.X	-2.6	-	0
14	.X	-2.8	-	0
20	XX	-3.0	-	0
0	-	-3.2	-	0
17	.X	-3.4	-	0
16	.X	-3.6	-	0
0	-	-3.8	-	0
105	.XXXXXXXXXXXX	-4.0	-	0

PL3

PL2

*For each bin in the theta distribution column, “X” represents 10 students, “.” represents a value in between 1 and 9 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

Table D.5 Grades Six through Eight Item-Person Map

Number of Students	Students *	Logit	Items**	Number of Items	
128	.XXXXXXXXXXXXX	4.0	-	0	
0		- 3.8	-	0	
0		- 3.6	-	0	
0		- 3.4	-	0	
0		- 3.2	-	0	
0		- 3.0	-	0	
78	.XXXXXXX	2.8	-	0	
134	.XXXXXXXXXXXXX	2.6	-	0	
0		- 2.4	-	0	
0		- 2.2	-	0	
77	.XXXXXXX	2.0	-	0	
102	.XXXXXXXXXXXXX	1.8	-	0	
74	.XXXXXXX	1.6	-	0	
79	.XXXXXXX	1.4	-	0	
50	XXXXX	1.2	-	0	
123	.XXXXXXXXXXXXX	1.0	-	0	
89	.XXXXXXX	0.8	-	0	
94	.XXXXXXXXXXXXX	0.6	E R	2	PL3
140	XXXXXXXXXXXXX	0.4	E	1	
69	.XXXXXXX	0.2	R	1	
113	.XXXXXXXXXXXXX	0.0	E E E E	4	
87	.XXXXXXX	-0.2	R R E E E	5	
86	.XXXXXXX	-0.4	E E E E R R	6	
112	.XXXXXXXXXXXXX	-0.6	E E	2	
114	.XXXXXXXXXXXXX	-0.8	R E E R	4	
122	.XXXXXXXXXXXXX	-1.0	E R E E E E E R R R	10	
115	.XXXXXXXXXXXXX	-1.2	E E R R R R E	7	PL2
87	.XXXXXXX	-1.4	E R	2	
89	.XXXXXXX	-1.6	R E R	3	
49	.XXXX	-1.8	R R R R E	5	
52	.XXXXX	-2.0	E	1	
20	XX	-2.2	-	0	
33	.XXX	-2.4	R E	2	
17	.X	-2.6	R	1	
28	.XX	-2.8	E E R	3	
16	.X	-3.0	E	1	
14	.X	-3.2	-	0	
11	.X	-3.4	R R	2	
10	X	-3.6	-	0	
16	.X	-3.8	-	0	
102	.XXXXXXXXXXXXX	-4.0	-	0	

*For each bin in the theta distribution column, “X” represents 10 students, “.” represents a value in between 1 and 9 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

Table D.6 Grades Nine through Twelve Item-Person Map

Number of Students	Students *	Logit	Items**	Number of Items	
360	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	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	
133	.XXXXXXXXXXXXXX	2.6	-	0	
124	.XXXXXXXXXXXXXX	2.4	-	0	
0		- 2.2	-	0	
0		- 2.0	-	0	
119	.XXXXXXXXXXXXXX	1.8	-	0	
114	.XXXXXXXXXXXXXX	1.6	-	0	
101	.XXXXXXXXXXXXXX	1.4	-	0	
74	.XXXXXXX	1.2	-	0	
175	.XXXXXXXXXXXXXXXXXXXX	1.0	E	1	PL3

87	.XXXXXXX	0.8	-	0	
128	.XXXXXXXXXXXXXX	0.6	-	0	
125	.XXXXXXXXXXXXXX	0.4	E	1	
149	.XXXXXXXXXXXXXX	0.2	E E	2	
130	XXXXXXXXXXXXXX	0.0	E E E	3	
177	.XXXXXXXXXXXXXXXXXXXX	-0.2	E E R E	4	
79	.XXXXXXX	-0.4	E R E E E	5	
166	.XXXXXXXXXXXXXXXXXXXX	-0.6	E	1	
136	.XXXXXXXXXXXXXX	-0.8	E R E E R	5	
146	.XXXXXXXXXXXXXX	-1.0	E R R E R R	6	PL2

128	.XXXXXXXXXXXXXX	-1.2	E R R E R	5	
117	.XXXXXXXXXXXXXX	-1.4	E R E R R E R	7	
117	.XXXXXXXXXXXXXX	-1.6	R E R R	4	
103	.XXXXXXXXXXXXXX	-1.8	R E E R	4	
87	.XXXXXXX	-2.0	R R E	3	
48	.XXXX	-2.2	E E R R	4	
54	.XXXXX	-2.4	E E	2	
45	.XXXX	-2.6	R	1	
23	.XX	-2.8	-	0	
22	.XX	-3.0	R	1	
23	.XX	-3.2	R	1	
24	.XX	-3.4	E	1	
4	.	-3.6	-	0	
12	.X	-3.8	-	0	
172	.XXXXXXXXXXXXXXXXXXXX	-4.0	R	1	

*For each bin in the theta distribution column, “X” represents 10 students, “.” represents a value in between 1 and 9 students, and no students are denoted as “-”.

**For each bin in the item difficulty distribution column, “R” represents a receptive item, “E” represents an expressive item, and no items are denoted as “-”.

This page is intentionally blank.