

# Final Report for the Alternate English Language Proficiency Assessments for California Pilot Using Cognitive Lab Methodology Study

**Contract No. CN140284**

**Prepared for the California Department of Education by Educational Testing Service**

**Submitted October 6, 2020**



**Educational Testing Service**

**660 Rosedale Road**

**Princeton, NJ 08541**

Table of Contents

[1 Executive Summary 6](#_Toc58566362)

[1.1 Preliminary Report Summary 7](#_Toc58566363)

[1.2 Final Study Recommendations 8](#_Toc58566364)

[1.3 Conclusion 10](#_Toc58566365)

[2 Background 11](#_Toc58566366)

[3 Purposes and Goals 13](#_Toc58566367)

[4 Areas of Investigation and Research Questions 14](#_Toc58566368)

[4.1 Interaction with Task Types 14](#_Toc58566369)

[4.2 Accessibility 14](#_Toc58566370)

[4.3 Test Administration Materials 15](#_Toc58566371)

[5 Pilot Test Materials 16](#_Toc58566372)

[5.1 Task Types 17](#_Toc58566373)

[5.2 Item Types 17](#_Toc58566374)

[5.3 Pilot Test Blueprint 18](#_Toc58566375)

[5.4 Pilot Test Forms Development 18](#_Toc58566376)

[5.5 *DFA* Development 19](#_Toc58566377)

[5.6 Accessibility Resources 20](#_Toc58566378)

[6 Methodology 21](#_Toc58566379)

[6.1 Sampling and Recruitment 22](#_Toc58566380)

[6.2 Participants 25](#_Toc58566381)

[6.3 Study Instruments 27](#_Toc58566382)

[6.4 Procedures 29](#_Toc58566383)

[6.5 Data Analysis 32](#_Toc58566384)

[7 Findings 34](#_Toc58566385)

[7.1 Interaction with Task Types 34](#_Toc58566386)

[7.2 Accessibility 58](#_Toc58566387)

[7.3 Test Administration 67](#_Toc58566388)

[8 Limitations of the Study 71](#_Toc58566389)

[9 Recommendations 72](#_Toc58566390)

[9.1 Recommendations on Interactions with Task Types 73](#_Toc58566391)

[9.2 Recommendations for Accessibility 77](#_Toc58566392)

[9.3 Recommendations for Test Administration 79](#_Toc58566393)

[References 81](#_Toc58566394)

[Appendices 85](#_Toc58566395)

[Appendix A 85](#_Toc58566396)

[Appendix B 86](#_Toc58566397)

[Appendix C 91](#_Toc58566398)

List of Tables

[Table 1. Number of Pilot Items Administered 19](#_Toc58566452)

[Table 2. Proposed Target Sampling Matrix 24](#_Toc58566453)

[Table 3. Participating Students by Grade Level and Grade Span 25](#_Toc58566454)

[Table 4. Home Languages 25](#_Toc58566455)

[Table 5. Primary Disability 26](#_Toc58566456)

[Table 6. Participant’s Special Education Setting 26](#_Toc58566457)

[Table 7. Participants’ Hearing 27](#_Toc58566458)

[Table 8. Participants’ Vision 27](#_Toc58566459)

[Table 9. Group 1, Task Type 1, *Recognize and Use Common Words* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings from Each Test Administration 36](#_Toc58566460)

[Table 10. Group 1, Task Type 2, *Understand a School Activity* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 37](#_Toc58566461)

[Table 11. Group 1, Task Type 3, *Describe a Routine* (Medium Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 37](#_Toc58566462)

[Table 12. Group 1, Task Type 4, *Interact with an Informational Text* (High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 37](#_Toc58566463)

[Table 13. Group 2, Task Type 1, *Communicate About Familiar Topics* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 38](#_Toc58566464)

[Table 14. Group 2, Task Type 2, *Understand a School Exchange* (Low-to-Medium Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 38](#_Toc58566465)

[Table 15. Group 2, Task Type 3, *Understand and Express an Opinion* (Medium-to-High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 39](#_Toc58566466)

[Table 16. Group 2, Task Type 4, *Interact with a Literary Text* (Medium-to-High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration 39](#_Toc58566467)

[Table 17. Test Examiner Ratings of Students’ English Language Proficiency 42](#_Toc58566468)

[Table 18. Score Summary for All Grade Levels and Grade Spans 43](#_Toc58566469)

[Table 19. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans) 44](#_Toc58566470)

[Table 20. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans) 44](#_Toc58566471)

[Table 21. Students’ Listening Skills 46](#_Toc58566472)

[Table 22. Students’ Reading Skills 47](#_Toc58566473)

[Table 23. Students’ Speaking Skills 47](#_Toc58566474)

[Table 24. Students’ Writing Skills 48](#_Toc58566475)

[Table 25. Were the Tasks Appropriate for Your Students? Frequencies and Percentages by Grade Level or Grade Span 49](#_Toc58566476)

[Table 26. Were the Stories or Passages Appropriate in Length, Complexity, and Topic? Frequencies and Percentages by Grade Level or Grade Span 49](#_Toc58566477)

[Table 27. Do You Ask Students These Types of Questions During Instruction? Frequencies and Percentages by Grade Level or Grade Span 49](#_Toc58566478)

[Table 28. Does This Item Measure Receptive Skills (Listening and Reading)? Average Percentages of All Items by Grade Level or Grade Span 50](#_Toc58566479)

[Table 29. Does This Item Measure Expressive Skills (Speaking and Writing)? Average Percentages of All Items by Grade Level or Grade Span 50](#_Toc58566480)

[Table 30. Was the Item for This Student (Easy, About Right, Hard)? Percent Average of Frequency of Selection Across All Items by Grade Level or Grade Span 51](#_Toc58566481)

[Table 31. Did This Student Perform as Expected? Percent Average of Frequency of Selection by Grade Level or Grade Span 51](#_Toc58566482)

[Table 32. Test Administration Times in Minutes (Group 1) 53](#_Toc58566483)

[Table 33. Test Administration Times in Minutes (Group 2) 53](#_Toc58566484)

[Table 34. Students’ Primary Communication Mode 55](#_Toc58566485)

[Table 35. English Only Students’ Primary Communication Mode 56](#_Toc58566486)

[Table 36. Total Scores Compared by Verbal and Nonverbal Students (Group 1) 57](#_Toc58566487)

[Table 37. Total Scores Compared by Verbal and Nonverbal Students (Group 2) 57](#_Toc58566488)

[Table 38. Accessibility—Universal Tools: Average Percentage of Use Across All Items and Grade Levels and Grade Spans 59](#_Toc58566489)

[Table 39. Accessibility—Designated Supports: Average Percentage of Use Across All Items and Grade Levels and Grade Spans 60](#_Toc58566490)

[Table 40. Accessibility—Accommodations: Average Percentage of Use Across All Items and Grade Levels and Grade Spans 62](#_Toc58566491)

[Table 41. Issues with Assistive Technologies? Frequencies and Percentages by Grade Level or Grade Span 62](#_Toc58566492)

[Table 42. How Familiar Are Test Examiners with the Accessibility Resources for the Alternate ELPAC? Frequencies and Percentages by Grade Level or Grade Span 63](#_Toc58566493)

[Table 43. Overall, Did Your Student Have Any Difficulties Accessing the Test? Frequencies and Percentages by Grade Level or Grade Span 64](#_Toc58566494)

[Table 44. What Best Describes Your Student’s Computer Use? 64](#_Toc58566495)

[Table 45. What Best Describes Your Student’s Tablet Use? 65](#_Toc58566496)

[Table 46. What Best Describes Your Student’s Need for Access When Using Technology? 65](#_Toc58566497)

[Table 47. Any Issues with Internet Connectivity? Frequencies and Percentages by Grade Level or Grade Span 65](#_Toc58566498)

[Table 48. Were the Task Types Appropriate for Students Across a Wide Range of Communication Modes? Frequencies and Percentages by Grade Level or Grade Span 67](#_Toc58566499)

[Table 49. Were the *Directions for Administration* Clear and Easy to Use? Frequencies and Percentages by Grade Level or Grade Span 68](#_Toc58566500)

[Table 50. Were the Rubrics Easy to Understand? Frequencies and Percentages by Grade Level or Grade Span 68](#_Toc58566501)

[Table 51. Test Examiner Self-Reported Issues with Scoring: Frequencies and Percentages by Grade Level or Grade Span 70](#_Toc58566502)

[Table 52. Did the Test Examiner Assign the Correct Score? 70](#_Toc58566503)

[Table 53. Overall, How Would You Rate the Amount of Time Needed for Taking the Test, Not Including the Probing Questions? Frequencies and Percentages by Grade Level or Grade Span 70](#_Toc58566504)

[Table A1. Alternate ELPAC Task Types Included in Pilot 79](#_Toc52354543)

[Table B1. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans) 86](#_Toc52354556)

[Table B2. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans) 86](#_Toc52354557)

[Table B3. Receptive Scores by Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans) 87](#_Toc52354558)

[Table B4. Receptive Scores by Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans) 87](#_Toc52354559)

[Table B5. Expressive Scores by Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans) 88](#_Toc52354560)

[Table B6. Expressive Scores by Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans) 88](#_Toc52354561)

[Table B7. Total Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans) 88](#_Toc52354562)

[Table B8. Total Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans) 89](#_Toc52354563)

[Table B9. Receptive Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans) 89](#_Toc52354564)

[Table B10. Receptive Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans) 89](#_Toc52354565)

[Table B11. Expressive Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans) 90](#_Toc52354566)

[Table B12. Expressive Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans) 90](#_Toc52354567)

[Table C1. Frequencies and Percentages of Students’ Listening Skills by English Language Proficiency 91](#_Toc52354568)

[Table C2. Frequencies and Percentages of Students’ Reading Skills by English Language Proficiency 91](#_Toc52354569)

[Table C3. Frequencies and Percentages of Students’ Speaking Skills by English Language Proficiency 92](#_Toc52354570)

[Table C4. Frequencies and Percentages of Students’ Writing Skills by English Language Proficiency 92](#_Toc52354571)

## Executive Summary

The Alternate English Language Proficiency Assessments for California (ELPAC) will be the required state test for identifying whether a student is an English learner (EL) and annually thereafter measuring English language proficiency (ELP) for eligible students with the most significant cognitive disabilities as identified by an individualized education program (IEP) team beginning in school year 2020–21. State and federal laws require that local educational agencies (LEAs) administer a state test of ELP to eligible students in kindergarten through grade twelve.

For students with the most significant cognitive disabilities, the Initial Alternate ELPAC will provide information to determine a student’s initial classification as an EL or as initial fluent English proficient. Identified ELs with the most significant cognitive disabilities will take annually the Summative Alternate ELPAC, which will provide information on student progress toward ELP and support decisions on reclassification as fluent English proficient.

The Alternate ELPAC will align to the 2012 *California English Language Development Standards* (2012 ELD Standards) via the English Language Development (ELD) Connectors. The ELD Connectors offer a reduction in the depth, breadth, and complexity of the standards as appropriate for students with the most significant cognitive disabilities. The Alternate ELPAC will assess students’ ELP skills in the four recognized language domains (i.e., Listening, Speaking, Reading, Writing) via receptive (Listening, Reading) and expressive (Speaking, Writing) modes of communication.

Under the direction of the California Department of Education (CDE), testing contractor Educational Testing Service (ETS) used an evidence-centered design approach to design and develop the Alternate ELPAC starting in the fall of 2018. This work has been done in consultation with the Alternate ELPAC Test Design Advisory Team (TDAT), the ELPAC Technical Advisory Group (TAG), and committees of California educators with expertise in special education and the education of ELs and additional stakeholders.

ETS conducted a pilot using cognitive laboratory methodology (hereafter, pilot study) in 23 schools across California from January 14–31, 2020. The overall purpose of the study was to determine whether the task types are suitable for the intended population. The specific goals of the study were as follows:

* Examine test examiner and student interaction with proposed task types to determine whether the task types elicit the intended knowledge and skills while minimizing construct-irrelevant interference
* Collect evidence of the use of accessibility resources used for test administration and determine whether the test design supports access
* Examine test administration practices to determine if the assessment is administered with fidelity to the *Directions for Administration (DFA)*

This executive summary presents both preliminary report summary findings, which were included in preliminary findings (ETS 2020a) and used to support development of the Alternate ELPAC test blueprint (ETS 2020b), and the final study recommendations. The final study recommendations are presented here for the first time and are being used to support assessment design and field test development by informing decision making for refining task types, developing future items, and improving accessibility and test administration materials. This report may also be used as evidence for technical requirements under peer review (U.S. Department of Education, 2018).

### Preliminary Report Summary

A preliminary review of the data on the investigation of task types was presented to the California State Board of Education (SBE) at the May 2020 meeting. The report, *Preliminary Findings on the Development and Evaluation of Task Types for the Alternate English Language Proficiency Assessments for California*,summarized the steps in the development of the task types and provided recommendations based on the pilot to support the proposed test blueprint to the SBE (ETS, 2020a). The recommendations in that preliminary report were approved by the CDE. In this final report, those recommendations are presented along with additional information on the methodology. The recommendations are contextualized in the presentation of the study findings. Collectively, the evidence in the preliminary report suggested that the tasks types were generally suitable for their intended purposes, also revealing some opportunities for refinement and improvement in specific areas. Accordingly, five evidence-based recommendations were proposed in the report. A summary of those approved actions is presented here:

**Approved Action:** The test blueprint was adjusted by adding *Recognize and Use Common Words* as the first task type across grade levels and grade spans. The inclusion of *Recognize and Use Common Words* as an initial task type across all grade levels and grade spans creates a more gradual progression in difficulty for students and provides more information about the students’ ability to take the assessment. This was an important consideration for providing more opportunities for access for students who perform at low levels of linguistic complexity and are at beginning levels of language development.

**Approved Action:** The multiple-selection, multiple-choice item type was removed and replaced with items using a single-selection, multiple-choice format. Students had difficulty understanding the directions for the multiple-selection, multiple-choice item type, which presented students with three options and asked them to select the two that were correct.

**Approved Action:** The kindergarten and grade one *Interact with an Informational Text* had one item aligned with Part III, a foundational literacy connector which was replaced with an item assessing a different standard. Of concern was the fact that some kindergarten and grade one students in the Alternate ELPAC population may not have had an opportunity to learn the skills assessed by this connector. For grade two and grades three through five, one item aligned to the PIII connector was maintained.

**Approved Action:** Minor adjustments to the task types were made, including providing considerations of alternate response options for expressive items to be more inclusive for students who are not verbal communicators. This includes students who are presymbolic communicators, who use few conventional symbols (words, signs, or pictures), or who use augmentative and alternative communication (AAC) devices. Guidance on how to provide optional individualization (adapt to the individual needs of a student) has been added for rubric-scored expressive items that either allow for the use of real objects or provide picture card responses.

**Approved Action:** Guidance was added to the *DFAs* to support the use of objects, manipulatives, and picture cards. Guidance on how to provide optional individualization has been added for stories, passages, and items in which real objects or manipulatives could be provided to the student. Directions regarding Optional Individualization have been included in the *DFA*. A list of additional materials, including any objects that can be used on the assessment, has been included in the *DFA*.

Other recommendations based on task type investigation are included in this final report. The findings are presented in [Section 7](#_Findings), and more detailed explanations for recommendations are presented in [Section 9](#_Recommendations).

### Final Study Recommendations

The study investigated test examiner and student interaction of proposed task types to determine whether the task types elicit the intended knowledge and skills with minimal construct-irrelevant interference. The study also examined whether the test design supported access and whether the test was administered with fidelity to the *DFA*. [Section 9](#_Toc49113528) includes a detailed discussion, including any follow-up actions that occurred as a result of the study. A summary of the recommendations follows.

#### Interaction with Task Types

##### Minimize Construct Irrelevance

The investigation of test examiner and student interaction with the task types found construct-irrelevant factors. Recommendations to improve construct relevance include updating the text complexity guidelines for future item writers, provide explicit training on construct-irrelevant factors including passage length, text complexity, and topic appropriateness. Additionally, guidance on the use of visuals should be considered.

Construct relevance may be improved by considerations of item writer recruitment. Consider targeted recruitment of certified special educators based on experience with serving students with the most significant cognitive disabilities who take alternate assessments. It is important to consider other specialists with direct target student experience who provide services and have expertise in the areas of English language development (ELD), speech language, assistive technology, vision impairment, and hearing impairment.

##### Improve Face Validity

The Alternate ELPAC combines the Listening, Speaking, Reading, and Writing domains into the categories of receptive (Listening and Reading) and expressive (Speaking and Writing) skills. To improve face validity (the appearance of whether or not the test measures what it claims to measure), consider how the test design of the Alternate ELPAC will be communicated effectively with educators, parents, and a variety of other stakeholders.

##### Future Study Recommendations

Future study recommendations include studies of opportunity to learn, including the explicit use of the ELD standards via the connectors for the target population and how students will receive instructional support for learning English. Additional studies recommended to include topics on young English learners, ELs with recent arrival status, and other topics to support validity.

#### Recommendations for Accessibility

##### Test Examiner Guidance on Accessibility Resources

To support test examiners and the field in better understanding accessibility resources for students taking the Alternate ELPAC, consider ways to support individual decision making for test examiners in the selection and use of the accessibility resources. Consider providing guidance and multiple avenues of training to support test examiners in the selection, provision, use, and monitoring of accessibility resources.

##### Accessibility: Presentation of Test Content

Because access to test content was enhanced for the students when test examiners knew how to use the universal tools, consider providing multiple opportunities for students and test examiners to practice the use of the universal tools—expandable items, passages, and zoom features.

##### Accessibility Resources and Standardization

Provide test examiners with explicit guidance on the provision of accessibility resources that support standardized test administration. Some accessibility resources, such as print on demand and alternate response options, were used in a manner that presented content in ways that undermine standardized test administration. Consider training that models the correct usage of accessibility resources for test administration, and specifically provide examples of allowed and nonallowed uses of accessibility resources.

##### Test Delivery and Accessibility

Consider improving the accessibility of the test content on the test delivery platform by adjusting the vertical layout of items, so that all answer choices are visible on one screen. Consider enlarging the [**NEXT**] button to promote access and student independence.

#### Recommendations for Test Administration

##### Test Examiner Training

Since test examiners expressed the need for more time to become familiar with the test materials, consider providing test administration training in multiple ways (e.g., online, face-to-face, video). Consider including a mock test administration and materials review prior to actual test administration. Consider the delivery of test materials in advance of test administration.

##### Practice Items

To prepare students and test examiners for future test administration, consider providing practice tests for test examiners to become familiar with test administration practices. Consider making practice test forms available on the testing interface, including sample items, scoring rubrics and *DFAs*. It is strongly recommended that test examiners practice instructionally with students.

##### Scoring Rubrics

Because test examiners expressed uncertainty with using the scoring rubrics and requested clarity on how and when modeling is permissible, consider practice with the scoring rubrics and demonstrate the modeling rubric flow chart as part of the test administration training. Consider developing materials, such as a short video or resource guide, to demonstrate both correct and incorrect provision of modeling support.

### Conclusion

The final report for the pilot study that follows includes a detailed description of the study’s methods, limitations, findings, and recommendations.

## Background

State and federal law require that LEAs administer a state test of ELP to eligible students in kindergarten through grade twelve. The Alternate ELPAC will be the required state test for initially identifying students as ELs or as initial fluent English proficient (IFEP), as well as measuring ELP for eligible students identified by the students’ IEP team with the most significant cognitive disabilities as per the (*California Code of Regulations* 850[v]) beginning in school year 2020–2021. ELs with the most significant cognitive disabilities represent a diverse population of students in kindergarten through grade twelve, inclusive of students up to age twenty-two enrolled in grade twelve who continue to be eligible for special education and ELD services.

For students with the most significant cognitive disabilities, the Initial Alternate ELPAC will provide information to determine a student’s initial classification as an EL or as initial fluent English proficient. Identified ELs with the most significant cognitive disabilities will take the Summative Alternate ELPAC annually, which will provide information on annual student progress toward ELP and support decisions on reclassification as fluent English proficient.

The Alternate ELPAC will align to the 2012 *California English Language Development Standards* (CDE, 2012) via the ELD Connectors (ETS, 2019b). The ELD Connectors will offer a reduction in the depth, breadth, and complexity of the standards as appropriate for students with the most significant cognitive disabilities. The Alternate ELPAC will assess students’ ELP skills in the four recognized language domains (i.e., Listening, Speaking, Reading, and Writing) in an integrated manner using receptive (Listening, Reading) and expressive (Speaking, Writing) task types. This approach allows flexibility in supporting student responses to items using individually preferred communication modes.

The Alternate ELPAC test design was developed using the evidence-centered design (ECD) approach (Mislevy, Almond & Lukas, 2003; Zieky, 2014). The Alternate ELPAC is an online linear test consistent with the California Alternate Assessments (CAAs). The Alternate ELPAC is organized into grade levels and grade spans consistent with the ELPAC: For the initial assessment, the organization is kindergarten (including transitional kindergarten), grade one, grade two, and grade spans three through five, six through eight, and nine through twelve; for the summative assessment, the high school grades have separate tests for nine and ten and eleven and twelve. The Alternate ELPAC assessment is delivered under untimed testing conditions with a one-on-one administration model that allows for the test examiner to interact with the computer on behalf of the student, as appropriate to the student’s individual needs and abilities.

ETS, under the direction of the CDE and in partnership with the Sacramento County Office of Education (SCOE), conducted a small-scale pilot of the Alternate ELPAC to investigate the proposed task types, use of accessibility resources, and the test administration processes. This study used cognitive lab research methodology to provide richly detailed evidence obtained through structured interview and observation techniques that can be used to support validity evidence and inform the next steps of Alternate ELPAC development.

A report titled *Preliminary Findings on the Development and Evaluation of Task Types for the Alternate English Language Proficiency Assessments for California* was produced and submitted to the CDE and the California SBE in May 2020 (ETS, 2020b). The purpose of that document was to summarize the steps in the development of the task types and provide preliminary recommendations based on the pilot findings to support submission of the proposed test blueprint (ETS, 2020a). Recommendations from that report, which was approved by the SBE during the May 2020 meeting, are included in [section 9](#_Recommendations).

## Purposes and Goals

The overall purpose of the study was to determine whether the Alternate ELPAC task types are suitable for the intended population. The validity evaluation of an assessment should consider whether the assessment is appropriate for the intended population (Pellegrino et al., 2001). Pilot testing allows test developers to identify items with features students perceive as potentially confusing. Identifying items and item features that may unintentionally influence and inhibit the performance of students during pilot testing can improve the quality of test items and other test materials before they are used in a large-scale administration such as a field test and can reduce the unnecessary costs required to make changes to test forms and procedures. The present study was not designed to generate item-level statistics. The goals of the pilot study utilizing cognitive lab methodology were as follows:

* To investigate test examiner and student interactions with the task types to determine whether the task types elicited the intended knowledge, skills, and abilities without construct-irrelevant interference
* To investigate accessibility considerations, to study the usage of the accessibility resources, and to determine whether the test design supports access for a range of communication modes
* To investigate test administration practices to determine if the use of all test materials supports a standardized test administration and fidelity of test scores to make recommendations for the field test

These study findings provide evidence for recommendations that are important for refining task types, informing item development, and improving accessibility and test administration materials for the Alternate ELPAC. Study findings may be used as evidence for technical requirements under peer review (U.S. Department of Education, 2018).

## Areas of Investigation and Research Questions

The Alternate ELPAC pilot used a cognitive lab methodology to examine student responses and test examiner interactions during test administration. There were three areas of investigation: interaction with task types, accessibility, and test administration materials and processes.

### Interaction with Task Types

The first area of investigation focused on collecting evidence of student interactions with the task types. These interactions were investigated to determine whether the task types elicited the intended knowledge and skills while minimizing construct-irrelevant interference. This area of investigation gathered evidence of students’ interaction with the task types to determine whether the task types were suitable for the target population. This area of investigation also focused on whether the task types were effective for measuring the ELP of students in this population.

It was important to determine if the items elicited the intended responses, which included an examination of the students’ communication mode responses to items designed to measure receptive or expressive skills.

The following research questions were posed to investigate the interaction with task types:

* Do the task types gather appropriate evidence about the students’ English language skills?
* Are the task types appropriate for the target population?
* Do the task types measure the intended response process? If not, how can they be improved?
* What are the communication modes students used to respond to the task types?

### Accessibility

The second area of investigation was accessibility, including the use of accessibility resources on the Alternate ELPAC. Accessibility resources included embedded and non-embedded universal tools, designated supports, and accommodations. The study examined other response options including the usage of other augmentative and alternative communication (AAC) devices. Some accessibility resources used were available to all students or were selectively available depending on a student’s IEP.

The following research questions were posed to investigate accessibility:

* Which accessibility resources from the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing* were used during administration?
* To what extent do accessibility resources support access for the individual students as they interact with, and respond to, each task type?
* What other materials or resources were used to provide access to support student responses to the task types?
* For students who independently use a computer or tablet, to what extent is the interface accessible?
* How do teachers perceive the effectiveness of task types and administration for students who participate using a range of communication modes?

### Test Administration Materials

The third area of investigation was to evaluate the clarity of test administration materials, which included the *DFA* and scoring rubrics, to refine materials for future test administration. Test examiners were closely observed using the materials and interviewed about the test administration experience using these materials.

The following research questions were used to investigate test administration materials:

* To what extent are the *DFA* and scoring rubrics clear for test examiners? What recommendations do test examiners have for the *DFA* and scoring rubrics?
* How do test examiners use the *DFA* during administration? What improvements could be made to the *DFA* for clarity and ease of use?
* How do test examiners use the scoring rubrics? What improvements could be made to the scoring rubrics for appropriateness, clarity, and ease of use?

## Pilot Test Materials

The Alternate ELPAC task types as presented in the pilot were developed in reference to California’s ELD Connectors. The task types were developed by ETS assessment development staff with expertise in two complementary areas of assessment: alternate assessment and ELP assessment. This team began by reviewing existing task types and item types from relevant assessments of ELP—including the ELPAC task types used to assess the ELP of students who have not been identified as having the most significant cognitive disabilities—and the CAA for English language arts/literacy (ELA), which is used to measure the reading and writing skills of students who have been identified as having the most significant cognitive disabilities.

A priority for task type development for the pilot was to ensure that the task types would reflect the three test design principles documented in the *Alternate ELPAC High-Level Test Design (HLTD)*:

* The test must be designed to ensure that the intended test-taking population is able to demonstrate ELP.
* The test design must be tailored to the range of needs of the students with the most significant cognitive disabilities, including maximum accessibility as well as ensuring linguistic and cultural fairness and sensitivity.
* The test design must take into consideration the testing burden for students and test examiners.

Two particular factors influenced the design of the task types:

* As many students with the most significant cognitive disabilities may have individually preferred communication modes, the items within the Alternate ELPAC task types are classified more generally as receptive or expressive rather than more specifically as the Listening, Reading, Speaking, or Writing domains. This allows flexibility for students to respond to test items using the same communication modes that the students most commonly use in the classroom.
* Alternate ELPAC task types are categorized in terms of linguistic complexity (low, medium, or high) and are presented in order of increasing complexity to help ensure that all eligible students are able to demonstrate the students’ level of ELP.

Development of items for the pilot study was guided by working versions of test design documentation, including draft versions of task type specifications and the test blueprint. These design documents were updated in an iterative manner as item development for the pilot moved forward to reflect input from reviews by CDE staff and consultations with advisory groups such as the Alternate ELPAC TDAT and the ELPAC TAG, as well as additional stakeholders, such as California educators.

The current versions of these test design documents, reflecting input from the groups noted previously as well as updates made based on the findings of the pilot study, are the *Alternate ELPAC Task Type Specifications* (ETS, 2019a), which include details on the topics of communication modes, receptive and expressive items, linguistic complexity, and the Alternate ELPAC Test Blueprint (ETS 2020a).

### Task Types

The creation of eight Alternate ELPAC task types served as the basis of item development for the pilot. [Appendix A](#_Appendix_A) provides the names and a brief description of each of these task types as administered in the pilot. In accordance with the *Alternate ELPAC HLTD*, all the task types are integrated; that is, each task type includes a stimulus that is followed by one or more receptive (Listening, Reading) items and by one or more expressive (Speaking, Writing) items (ETS, 2019). Task types are presented in order of complexity.

### Item Types

The term “item type” is used to describe test items as categorized according to their properties in the technological platform system supporting the test, including response modes. This is in contrast to the term “task type,” which is being used to describe test items in terms of their content and the evidence they gather about student language proficiency.

The following is a list of item types, with brief descriptions, that were included in the pilot and cognitive lab study; the number of items with technology-enhanced interactions was minimized to ensure maximum accessibility for the diverse test-taking population:

**Multiple Choice** requires a test taker to select from a list of choices (machine scored):

1. **Single Select**—the test taker must select one correct option.
2. **Multiple Select**—the test taker must select more than one correct option.

**Multiple Choice Single Select Partial Credit** requires a test examiner to assess a text taker’s expressive response following a rubric in the *DFA* (constructed response [CR], locally scored). The response is assigned a full, partial, or zero score based on the rubric. For the operational field test, this item type has been replaced with Inline Choice List Single Select, which allows for the constructed-response options to be listed in a drop-down list that will be accessed by the test examiner only.

**Match Single Select** requires the test takerto move a single object, such as a word or an image, by selecting and dragging the object into a specific order or to a specific location in a table, passage, or image (technology-enhanced interaction, machine scored). Note that the pilot study included a limited number of match single select items, as will the operational field test. Item statistics on these technology-enhanced items will be evaluated before additional match single select items are developed for future use.

These item types are included with a ***set leader***. The set leader contains stimulus material such as a passage or image that applies to a set of associated items.

### Pilot Test Blueprint

ETS developed a preliminary test blueprint that guided item and form development for the pilot study.

As a key part of the analysis of the pilot test, recommendations were made regarding which task types were suitable for use on the field test forms and how many items in each task type were needed to appropriately sample the 2012 ELD Standards via the ELD Connectors (ETS, 2019b), support reliable score reporting, and provide a test form that was of appropriate length. The preliminary test blueprint was revised accordingly following CDE’s approval of recommendations regarding the disposition of task types and item types, leading to the development of the Alternate ELPAC Test Blueprint (ETS, 2020), which was approved by the SBE in May 2020.

### Pilot Test Forms Development

One pilot test form was assembled and administered at each of the following grade levels and grade spans: kindergarten, one, two, three through five, six through eight, and nine through twelve (refer to table 1). The pilot test forms were assembled so that they included a balance of receptive and expressive items and covered a variety of item types, item difficulties, linguistic complexity levels, and key distributions. The forms were evaluated via the standard ETS review process and reviewed by the CDE in ETS’ Item Banking Information System (IBIS). Once approved, the forms went through an internal review before moving forward to American Institutes for Research (AIR), now Cambium Assessment, for final user acceptance testing, followed by the launch of the pilot test administration. ETS developed pilot test items and delivered them to the CDE via IBIS. Standard assessment procedures were followed for item development, inclusive of reviews by the CDE and security measures.

The goal was to create pilot test forms containing a representative sample of the items being considered for use in the Alternate ELPAC task types (i.e., to enable all potential items and task types to be administered to students), but not necessarily to create pilot test forms that would closely resemble the intended operational forms. All items were reviewed for alignment to the ELD connectors and sensitivity by a panel of California educators as well as by the CDE during an Item Review Meeting (IRM). One sample item, shared across grade levels, was also developed. Prior to inclusion in the pilot, the sample item was reviewed by California educators and the CDE during the IRM.

Table 1 shows the distribution of the items administered according to task type and grade level or grade span.

Table 1. Number of Pilot Items Administered

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Task Type** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–‍12** | **Total** |
| *Recognize and Use Common Words* | 2 | 0 | 2 | 0 | 2 | 0 | 6 |
| *Communicate About Familiar Topics* | 0 | 2 | 0 | 2 | 0 | 2 | 6 |
| *Understand a School Activity* | 4 | 0 | 4 | 0 | 4 | 0 | 12 |
| *Understand a School Exchange* | 0 | 4 | 0 | 4 | 0 | 4 | 12 |
| *Describe a Routine* | 3 | 0 | 3 | 0 | 3 | 0 | 9 |
| *Understand and Express an Opinion* | 0 | 4 | 0 | 4 | 0 | 4 | 12 |
| *Interact with a Literary Text* | 0 | 4 | 0 | 4 | 0 | 4 | 12 |
| *Interact with an Informational Text* | 4 | 0 | 4 | 0 | 4 | 0 | 12 |
| **Total Number of Items Administered** | 13 | 14 | 13 | 14 | 13 | 14 | 81 |

**Note:** Because there is one set of ELD Connectors for grades nine through twelve, pilot forms for grades nine and ten and eleven and twelve were combined.

### *DFA* Development

The *DFA* was developed to guide test examiners in the administration of the pilot items and the local scoring of CR items. The content of the *DFA* was limited to test administration and scoring procedures that were particular to a grade level or grade span. *DFAs* contained secure test content and were labeled accordingly. Test examiners were trained to keep *DFAs* secure at all times.

With respect to scoring, the *DFA* provided instructions for scoring CR items (i.e., expressive items with a rubric) locally during the pilot. It should be noted that the goal of providing scoring instructions was to examine the clarity of the *DFA* and the scoring rubrics and gain feedback about the directions and procedures. The reliability or the psychometric properties of scores is not within the scope of the study.

Standard assessment procedures were followed for *DFA* development. The grade-level and grade-span *DFAs* were developed as PDFs that were printed and distributed to test examiners prior to administration.

### Accessibility Resources

Accessibility resources are identified in the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing* (ETS, 2019c) document. Students’ use of accessibility resources was one of the major areas of investigation for the study. Some examples of resources that were used in the study are: zoom (universal tool), print on demand (designated support), and alternate response options (accommodation). In the instance that test examiners and students used resources otherwise not specified in the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing*, such as translation of test items or hand-over-hand prompting, ETS documented and reported any unlisted resources.

## Methodology

The Alternate ELPAC pilot was conducted using cognitive lab methodology. Cognitive lab methodology has been recognized as an effective tool for the development of assessments for diverse populations (e.g., American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) and is recommended as a source for obtaining response process evidence for alternate assessments (Karvonen, 2019). Cognitive laboratory methodology techniques include observation and interviewing. Cognitive interviewing is a method to identify sources of confusion in assessment items and to assess validity evidence on the basis of content and response processes. Validity evidence based on response processes refers to whether students respond to an item in the intended way, including documenting individual responses and difficulties to identify potential themes across a group of respondents (Miller, Willson, Chepp, & Ryan, 2014). The cognitive lab methodology for this study included close observation of student-test examiner test-taking processes using a familiar test examiner to collect evidence of students’ responses to content followed by cognitive interviews.

Cognitive laboratories and think-aloud interviews allow researchers “to present observable indicators (evidence) of phenomena that are technically unobservable” (Leighton, 2017, p. 14). A recent study recognized the importance of including test takers themselves as critical to the development and improvement of English language tests (Winke et al., 2018). The study used concurrent verbal probing, a recognized approach to elicit desired evidence from the students themselves. Probing introduces less interference than a true think-aloud (Willis, 2005). The think-aloud approach was not used with the students in this study due to considerations of burden on cognition, working memory, and processing skills.

Another hallmark of cognitive lab methodology is the observation component. Each research team was comprised of two ETS staff with expertise in ELD and special education for students with the most significant cognitive disabilities. While the test examiner administered the test and probing questions to the student, ETS trained research teams systematically collected data using a structured protocol. After each individual student test administration session, a short interview was conducted with test examiners. The purpose of the interview was to gather evidence on test examiners’ feedback on the task types, *DFA*, and accessibility considerations for students.

The sources of data for the study included a student Background Information Questionnaire (BIQ) and a test administration protocol that contained the embedded research questions within the *DFA*. Both the student BIQ and the test administration protocol are sources of evidence from the test examiner. Student probing questions were embedded into the test administration protocol and are sources of evidence from the students who participated in test administration. A combined observation and interview protocol are sources of evidence collected by the research team.

The triangulated data (student responses, test examiner interactions, ratings and interviews, and research team observations) was qualitatively analyzed and summarized for recurring patterns in response to the study’s research questions. Data analysis included descriptive and nonparametric statistics. Detailed information about the sampling plan, study instruments, study procedures, and data analysis are included in the following subsections.

### Sampling and Recruitment

The existing literature on conducting cognitive labs places an emphasis on purposeful, targeted sampling rather than on sampling size (Beatty & Willis, 2007; Willis, 2005). Purposeful sampling is used most often when a difficult-to-reach population needs to be measured. Research suggests that samples of 5 to 15 interviews are common practice for cognitive labs (e.g., Blair & Conrad, 2011; Peterson, Peterson, & Powell, 2017; Willis, 2005). Some researchers argue that even one to three cases can provide critical information (e.g., Hix & Hartson, 1993; Willis, 2005). The cognitive lab methodology relies on identifying characteristics of the groups most relevant to the study and conducts interviews with members of each until relatively few new insights are obtained (Beatty & Willis, 2007).

Recruitment was critical for the success of this study. ETS, in collaboration with SCOE, worked with the CDE to determine specific LEAs to conduct the study and to recruit test examiners and individual students who met the sampling criteria. The recruitment of the sample included consideration of representations from the southern, central, and northern regions of California. Oversampling was desirable as attrition rates for this population were anticipated to be high. Sampling considerations included the following:

Participation was on a voluntary basis.

Parent consent was required for participation in the study.

Time for preparing test examiners was limited.

The main sampling criteria for selection of test examiners was familiarity with the student (Fuchs & Fuchs, 1989). During recruitment, the benefits of participating in the study (e.g., opportunities to experience the Alternate ELPAC prior to operational testing and providing feedback to impact the development) were communicated with potential participants. In the case a substitute teacher was needed to cover participating educators’ time, the cost of substitutes was provided to participating LEAs. Schools, teachers, and students received nonmonetary incentives for participation. SCOE and ETS staff recruited and conducted the Alternate ELPAC test examiner trainings.

In this study, a purposeful sampling plan was used to deliberately include a variety of ELs with significant cognitive disabilities across kindergarten through grade twelve, and background characteristics including the full range of ELP. ETS endeavored to stratify the sample using the characteristics from the student BIQ. It was important to ensure that the sample of the study had a variety of backgrounds representative of the test takers for the operational test. The student BIQ was used to screen students to select a range of those who best fit the characteristics of interest:

* Home languages
* Range of communication modes
* Recent arrival status
* Country of birth
* Primary disability
* Special education setting
* Experience with compatible testing devices
* Previous performance on the CAA for ELA for students in grades three through twelve
* Range of test examiner ratings of ELP (low, medium, and high)
* ELD service delivery and provision of services

The targeted sample had a total of 66 students who are English learners with significant cognitive disabilities, of which 54 were ELs with significant cognitive disabilities, and 12 ELs with recent arrival status, and all were eligible for alternate assessment. Recent arrival status is given to students who are recently arrived ELs that have been in U.S. schools for less than 12 months. Additionally, 12 students with significant cognitive disabilities who were non-ELs across kindergarten through grade twelve were invited to participate and served as a control group. The sample is a relatively large sample for this sort of study, and that robust sample makes the study findings stronger. Such a large sample also represents a significant investment of time and effort made to get data from a wide and representative group of students.

Table 2 summarizes the targeted sample size by grade level or grade span and types of students. In total, 54 ELs with significant cognitive disabilities (ELSCDs) were invited to participate in the study; 12 students with significant cognitive disabilities (SCD) who were non-ELs were added, for a total of 66 study participants across kindergarten through grade twelve.

Table 2. Proposed Target Sampling Matrix

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Participants** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–‍12** | **Total** |
| ELSCDs | 8 | 8 | 8 | 6 | 6 | 6 | 42 |
| ELSCDs with Recent Arrival Status | 2 | 2 | 2 | 2 | 2 | 2 | 12 |
| SCD Non-ELs | 2 | 2 | 2 | 2 | 2 | 2 | 12 |
| **Target Total** | 12 | 12 | 12 | 10 | 10 | 10 | 66 |

It is important to note that the proposed sample was expected to be challenging to identify and recruit. ETS and SCOE made efforts to recruit a sample of students who matched the targeted sample distribution and collaborated closely with the CDE and appropriate stakeholders (e.g., the Alternate ELPAC TDAT and the ELPAC TAG) to identify appropriate measures to ensure the validity of the study results against the possibility of a gap between targeted and actual recruitment.

ETS used oversampling to ensure variation in the sample and to negotiate the risk of absenteeism on the day of testing. ETS advised that the test examiners should be familiar with the student and the student’s accommodations. SCOE identified 278 eligible students across eight participating districts in northern, central, and southern California. Each school was contacted for study participation and test examiners were recruited who had eligible students. Each participating test examiner was requested to complete the student BIQ survey and obtain consent forms for the eligible participating students. This effort resulted in the identification of 35 schools and 144 eligible student participants. From this pool, 71 study participants in 23 schools were observed.

The selection of participants was based on the representativeness of the characteristics using the student BIQ. The identification and selection of study participants sought a maximum variability in the small sample using the characteristics of students with diverse home languages across a variety of primary disabilities, representation of a full range of communication modes, a range of English language ability ratings (low, medium, and high), and special education settings for inclusion criteria. Chi-square analyses and associated effect size measures (refer to Steinberg, Brenneman, Castellano, Lin, & Miller, 2014 for a description of those calculated here) detected no significant association between teacher-reported English language ability ratings and study participation relative to the full SCOE pool (p = .08; Cramer’s *V* = .21 and considered moderate). A similar analysis detected no significant association between communication mode (verbal vs. non-verbal) and study participation relative to the full SCOE pool (p = .08; Goodman and Kruskal tau = .02).

### Participants

Seventy-one students with significant cognitive disabilities were observed; 58 were designated ELs and 13 were non-ELs, all with significant cognitive disabilities. Despite attempts to include recently arrived students in the sample, only three kindergarten students participated. The participants are presented in table 3, which summarizes the sample by grade level or grade span and by category of student.

Table 3. Participating Students by Grade Level and Grade Span

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Participants** | **K** | **Gr 1** | **Gr 2** | **Gr 3–‍5** | **Gr 6–‍8** | **Gr 9–‍12** | **Total** |
| ELSCDs | 7 | 10 | 10 | 8 | 10 | 10 | 55 |
| ELSCD Recent Arrival Status | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| SCD (non-EL) | 3 | 2 | 2 | 2 | 2 | 2 | 13 |
| **Total** | 13 | 12 | 12 | 10 | 12 | 12 | 71 |

Participation exceeded the proposed plan, and a total of 71 students participated. This report reflects the data that was obtained. Data from the student BIQ was missing for some students, and this is reflected in the tables where all totals do not equal 71.

Home languages were reported for 70 of the 71 participants. The home languages reported included Spanish (69.01%), English (18.31%), Vietnamese (2.82%), Russian (2.82%), Arabic (2.82%), Filipino (1.41) and Punjabi (1.41%), as presented in table 4.

Table 4. Home Languages

|  |  |  |
| --- | --- | --- |
| **Home Language** | **Frequency** | **Percent** |
| Arabic | 2 | 2.82 |
| Filipino | 1 | 1.41 |
| Punjabi | 1 | 1.41 |
| Russian | 2 | 2.82 |
| Spanish | 49 | 69.01 |
| Vietnamese | 2 | 2.82 |
| English | 13 | 18.31 |
| Missing | 1 | 1.41 |
| **Total** | 71 | 100.01 |

**Note:** Percentages do not equal 100 percent due to rounding.

Students with disabilities who receive special education are assigned a primary disability code. Primary disability codes were reported for 68 of the 71 participants. The participation of students with the primary disability codes at the highest levels was intellectual disability (49.30.47%), autism (19.72%), and multiple disabilities (16.90%).

These findings are consistent with the most common disability categories among students with the most significant cognitive disabilities on other alternate assessments (Nash, Clark, & Karvonen, 2016; Thurlow, Wu, Quenemoen, & Towles, 2016). Other primary disability codes are included in table 5.

Table 5. Primary Disability

|  |  |  |
| --- | --- | --- |
| **Primary Disability** | **Frequency** | **Percent** |
| Autism | 14 | 19.72 |
| Intellectual Disability | 35 | 49.30 |
| Multiple Disabilities | 12 | 16.90 |
| Other Health Impairment | 3 | 4.23 |
| Orthopedic Impairment | 1 | 1.41 |
| Specific Learning Disability | 1 | 1.41 |
| Speech or Language Impairment | 2 | 2.82 |
| Missing | 3 | 4.23 |
| **Total** | 71 | 100.02 |

**Note:** Percentages do not equal 100% due to rounding.

The special education settings were reported for 68 of the 71 participants. The special education setting category with the largest number of participants was students being in a regular class less than 40 percent of the day (63.40%). Students in this category are primarily served within self-contained classrooms with full-time special education instruction on a regular school campus. The least restrictive setting category obtained for study participants was regular classroom for 40 to 79 percent of the day (16.90%). Students in this category typically receive services in resource rooms with special education and related services or are in resource rooms with part-time instruction in a regular class. None of the students that participated in this study were identified in the category of being in the regular classroom 80 to 100 percent of the school day. The category of students in a separate school setting contained 12.70 percent of students for this study. Students identified in this category receive educational programs in public separate day schools, such as a special district school (refer to table 6).

Table 6. Participant’s Special Education Setting

|  |  |  |
| --- | --- | --- |
| **Special Education Setting** | **Frequency** | **Percent** |
| Regular class (more than 40–79% of the day) | 12 | 16.90 |
| Regular class (less than 40% of the day) | 45 | 63.40 |
| Separate schools | 9 | 12.70 |
| Missing | 5 | 7.04 |
| **Total** | 71 | 100.04 |

**Note:** Percentages do not equal 100 percent due to rounding.

Information on participants’ hearing was reported for 68 of the 71 participants. The majority of the students’ hearing was rated within normal limits (93 %), while the remaining students corrected with an aide (2.9%). No students who were deaf participated in the study (refer to table 7).

Table 7. Participants’ Hearing

|  |  |  |
| --- | --- | --- |
| **Student’s Hearing** | **Frequency** | **Percent** |
| Normal | 66 | 93.00 |
| Hard of Hearing, loss corrected with aide | 2 | 2.81 |
| Missing | 3 | 4.22 |
| **Total** | 71 | 100.03 |

**Note:** Percentages do not equal 100 percent due to rounding.

Information on participant’s vision was reported for 66 of the 71 participants. The majority of the students’ vision was rated normal (76.10%), though some students’ vision (12.67%) was reported as corrected vision within normal limits. Two participants had vision impairment, and one participant was blind. No students in the study used braille (refer to table 8).

Table 8. Participants’ Vision

|  |  |  |
| --- | --- | --- |
| **Student’s Vision** | **Frequency** | **Percent** |
| Normal | 54 | 76.10 |
| Corrected vision in normal limits | 9 | 12.67 |
| Vision impairment or low vision | 1 | 1.41 |
| Vision impairment for daily activities | 1 | 1.41 |
| Blind | 1 | 1.41 |
| Missing | 5 | 7.04 |
| **Total** | 71 | 100.04 |

**Note:** Percentages do not equal 100 percent due to rounding.

### Study Instruments

ETS developed the following study instruments in addition to the test materials:

Study recruitment letters and consent forms

Teacher and student BIQ

Observation and interview protocols

Test administration protocol for test examiners (embedded into the *DFA*)

Nondisclosure agreement (NDA)

Each document was reviewed and approved by the CDE. A description of each instrument is provided in the following.

#### Study Letters and Forms

The recruitment letters and consent forms introduced the context and purpose of the study to potential participants and parents or guardians of students. The letters articulated participants’ tasks, the participants’ rights as research participants, and any procedures that were taken to safeguard personal information. The letters described the benefits for the participants as well as the participants’ right to decline to participate in the study at any time without fear. The consent forms were signed and returned to ETS staff (before the cognitive lab was administered to a student) for processing and safeguarded in a locked file cabinet and are to be destroyed at the end of the study (five years after the end of the data collection procedures).

The purpose of the NDA was to maintain the security and confidentiality of the Alternate ELPAC content from any nonstudy personnel or nonstudy participants. The NDA was completed by test examiners as well as any other personnel who may have needed to be in the testing room but were not participating in the cognitive lab session (e.g., technology, support staff). Precautions were taken to secure the confidentiality and privacy of the participants, and unique identifiers were destroyed at the end of the study.

#### Student BIQ

The learning characteristics of the assessed population have significant implications for an assessment’s validity (Pellegrino et al., 2001). The students’ BIQ collected relevant information about the study participants from test examiners (e.g., grade level, gender, age, home language, recent arrival status, experience using testing compatible devices, primary disability, and level of communication). The questionnaire was completed by a teacher who is familiar with the participating student. ETS used this information to support the recruitment efforts to meet the proposed sampling targets and to describe the demographic characteristics of study participants.

An examination of the characteristics of the target population helped guide decisions for elements such as mode of delivery and allowable accommodations to support the construct definition and the intended assessment purpose (Winter et al., 2018). Attending to these needs through the principles of universal design helped ensure that information was presented in the most accessible formats (CAST, 2018). It also allowed students to demonstrate English language skills across the test in manners most appropriate for each individual student using any accommodations, such as assistive technology, AAC, or other manual communication systems.

#### Observation and Interview Protocols

The observation and interview protocols were used by ETS research staff. The observation protocol was a structured form using the same questions for each test administration. The use of teams was recommended so that at the outset at least two people agreed that a condition existed and also agreed on its rating. Researchers collected data simultaneously, with one researcher focused on the test examiner interactions and the other on student responses. ETS staff systematically noted the observations of the test administration and observed students’ and test examiners’ performance. The protocol was used for the recording of notes and ratings on the student and test examiner interaction with the task types, accessibility options, and fidelity of test administration. Quality control checks were built into the initial training, with practice exercises conducted so that observers’ understanding and abilities were assessed.

The interview section of the protocol was a semi-structured form that used a standardized list of questions and allowed for probing questions that guided the test examiner interview process. The interview protocol included questions about the appropriateness of the task types, accessibility, and test administration recommendations. Interviews were conducted conversationally with one respondent at a time, employing a blend of closed and open-ended questions, often accompanied by follow-up using why or how questions in terms of the insights and information gained. The qualitative information was used to provide insight and elaboration across experiences, identify patterns, offer insight into rare experiences, and illustrate findings.

#### Test Administration Protocol

ETS designed a structured research protocol and embedded the research questions into the *DFA*. Questions for the teacher (open ended, rating form) were embedded in the *DFAs* after each test question. The collection of the data concurrently during test administration was for collecting immediate responses in the moment. The standardized protocol collected information about the students’ response processes for the test items and elicited immediate feedback from the test examiner after each item. The embedded research protocol also included a set of questions for test examiners to respond to regarding students’ performance, item level information, and the task types. Data was collected to provide evidence of how students interacted with and responded to each task type, to determine whether the items elicited the intended responses. The protocol included a rating form where test examiners entered the evaluation of students’ interaction with the task types during the test administration. The embedded research protocol included a set of probing questions to be administered to the students by the test examiners immediately after each test item was completed. Student responses were collected for each item.

#### Test Examiner Questionnaire

A survey was collected from test examiners prior to observation. The survey was conducted to learn about the characteristics of test examiners (including educational background, experience with the target population and test platform, and whether the test examiners themselves are multilingual).

### Procedures

Prior to conducting the study, the study instruments were reviewed and approved by ETS’ internal Institutional Review Board (IRB), also known as the Committee for Prior Review of Research. The ETS team prepared study materials, which were reviewed and approved by the IRB. The IRB reviewed the study design to ensure the rights of human subjects involved in the study were adequately protected, and the research study would not cause harm to any participants.

Additionally, test examiners were verified school- or LEA-level employees who already had appropriate fingerprinting and background clearance to work with students at school sites. ETS observers obtained appropriate security clearance through ETS’ Working with Minors Policy (ETS, 2019) to work with students at the school sites. The test examiners administered the pilot test individually to students in one-on-one cognitive labs with two ETS staff present. Three test administrations per team was the maximum per day.

#### Before Data Collection

Before administering the study, ETS and SCOE worked with each LEA to identify an appropriate LEA liaison who helped arrange logistics at the testing sites as well as identify teachers who were familiar with the student participants as potential test examiners.

At the same time, ETS research teams were trained to conduct the cognitive labs in four online sessions and a one-day, face-to-face meeting prior to data collection for calibration of study instruments and team roles. A total of 12 hours of training for research team members was provided. Training sessions included cognitive lab methodology, observation and interview techniques, and a mock test administration session using the study protocols. During the sessions, the study materials, including the training test and *DFAs*, were reviewed in depth, and items were reviewed to train and calibrate observers’ ratings prior to the data collection efforts.

ETS led the development for test administration training and collaborated with SCOE for live and online trainings for participating test examiners. SCOE contracted local test examiners to administer the Alternate ELPAC pilot task types in one-on-one cognitive lab settings. Prior to the data collection, test examiners were trained to use the *DFA* and review and practice the embedded test questions.

Test examiners were asked in advance to get site approval for testing locations at the school. ETS recommended that students be familiar with the selected setting to be used as a testing site (a classroom where students had received routine instruction was recommended). Test examiners were responsible for securing the equipment for testing (e.g., desktop computers, Chromebooks), and other hardware (e.g., AAC devices). The test examiner confirmed that all equipment was in place prior to test administration. Test examiners identified in advance if substitute teacher reimbursement was needed.

The test examiner was responsible for distributing and collecting consent forms from parents and guardians. Signed consent forms were returned to ETS or SCOE prior to testing; passive consent was not allowed. The collected consent forms were stored at the ETS office in a locked filing cabinet; all personally identifiable information was replaced with a unique study identification (ID) used for study purposes only.

Test examiners were asked to complete a background questionnaire for themselves and one for each potential study participant. Examiners may have had more than one student for the study. The forms were password-protected to safeguard any personally identifiable information. Once completed forms are received, names were removed from the forms and replaced with unique student IDs.

#### During Data Collection

The study was conducted January 14–31, 2020 over 12 school days (i.e., a three-week testing window). All hard copies of the consent forms were collected prior to test administration. The test examiner, assisted by ETS staff members, set up the students’ technology, started the test session, and opened up the secure browsers to prepare for testing. The test examiner considered how students were scheduled and navigated them to the designated testing location to start the pilot administration. A critical part of this step included the provision of the individual statewide student identifiers (SSIDs) so that participants could be logged on to the secure testing browser. The SSIDs were provided by the test examiner who assisted with the study.

Each student session ranged between 9 and 64 minutes, with an average time being 24 minutes for kindergarten through grade twelve. Breaks were given as needed. In each session, a test examiner familiar to the student who administered the assessment and ETS staff members were present. The cognitive labs were conducted in English; efforts were made to allow students’ preferred languages during interviews to the extent possible (e.g., via familiar test examiners or paraprofessionals who may have spoken the specific languages). Students were given instructions on how to respond to the probing questions and allowed to practice to gain familiarity with the process.

Once logged on to the secure testing session, the test examiners reviewed the study’s procedures and secured student assent to continue. Next, test examiners began the administration of the Alternate ELPAC items. During the test administration, the test examiners also followed standardized item prompts in the protocol. The prompts guided the test examiner to elicit the desired response processes from the student. Test examiners asked probing questions to the students during the test administration after each item. Students were offered a break after each task type or set of items, and the probing questions were completed.

All *DFAs* were collected at the end of the test administration. The observers immediately reviewed the *DFAs* to determine areas of focus for the interview with each test examiner. After each test administration, ETS researchers conducted a short 30‑minute interview with test examiners to gather additional evidence on test examiners’ judgments on student responses to the task types, accessibility of the items, and clarity of test administration materials, and test examiners were given an opportunity to elaborate on any recommendations.

The principal investigator supervised team members and conducted daily check-ins with the research teams for consistency, for data quality, and to record progress. The recruitment matrix was reviewed daily to determine that the selection of students from the pool was meeting the study requirements. The number of test administrations was communicated, and weekly meetings were conducted with ETS staff members.

### Data Analysis

To reduce potential bias from the use of any single data source, observer, or theoretical position, multiple data sources were collected through the study. Data included test examiner and students’ background information, test examiner ratings recorded in the embedded research questions in the *DFA*, observation and interview protocols, and research team notes. Students’ responses were collected, and scores were examined to support the interpretation of whether the task types were suitable for the population. Other areas of focus included student interaction with the task types and the students’ test-taking processes, accessibility resource use, and the suitability of test administration materials.

This study used a grounded theory approach to analyze the data (Charmaz, 2002; Miles & Huberman, 1994; Patton, 2015). The ETS team entered the data into a single Excel file per instrument to record and organize it for analysis. ETS compiled the data into an overall Excel spreadsheet for cleaning and further organization. The principal investigator conducted a read-through of the database. The ETS team qualitatively analyzed the data by the protocol topics organized by the research questions regarding the task types, accessibility, test administration, and areas for improvement. Attribute coding was created from the study protocol and used to analyze and interpret the data (Saldaña, 2014).

Data from the test examiner interview included qualitative data that was examined, and open codes were created. Open codes were developed to code qualitative data that did not fit the a priori codes. Coding was led by the principal investigator of the study and corroborated as a group for accuracy and clarity in interpretation with the cross-functional team of ETS observers who were in the field. This iterative process was conducted until 100 percent agreement was reached. Focused coding was conducted using frequencies and emerging themes were examined for the possibility of relationships among variables and to build insight from the data, identify patterns, trends, and summarize recurring themes. Examples of open codes that emerged in the data included *AAC* *use*, *modeling rubrics*, and *training needs*. The qualitative data in the report is shared as anecdotal and observational information to support findings when warranted by the obtained evidence. Care was taken to ensure identifying information from students was removed from any examples.

Data was also categorized using closed codes that corresponded with the a priori categories in the ETS observation protocol, the student BIQ, and the test examiners’ embedded research questions in the *DFA*. The closed codes consisted of examples like *item difficulty*, *accessibility use*, and *clarity of rubric*. The quantitative data was analyzed by each grade level to obtain descriptive statistics by grade level and grade span; analyses of receptive, expressive, and overall scores were conducted. For the purposes of the data analysis, the missing data and items not administered, or research questions skipped, were coded as did not administer (DNA) and were treated similarly as missing in the data set.

The data from the analysis represents the aggregate analysis. Reporting data in the aggregate is important to minimize the risk of identifying information from students. Data was examined by student characteristics from the student BIQ. Descriptive statistics from protocols were obtained at each grade level or grade span.

Exploratory investigations using nonparametric inferential statistics were conducted. Of particular interest was whether scores obtained were related to communication mode. For example, the communication mode responses were collapsed to provide a verbal status for interpreting overall scores and an analysis of receptive and expressive scores. The section on findings presents the results of these investigations.

## Findings

This section presents the study findings, which are organized according to the three main areas of investigation of interaction with task types, accessibility, and test administration. The findings presented are important validity evidence of the pilot, which considers whether the assessment is appropriate for the intended population. Data sources in this section include quantitative and qualitative data from across the research questions embedded in the *DFA*, observation and interview protocols, and the student BIQ.

### Interaction with Task Types

The first area of investigation focused on collecting evidence of test examiner and student interactions with the task types. The goal of this investigation was to determine whether the task types elicited the intended knowledge and skills with minimal construct-irrelevant interference. This area of investigation also gathered evidence of students’ interaction with the task types to determine whether the task types were suitable for the target population and whether the task types were effective for measuring the levels of ELP of students in this population. It was important to collect evidence from a range of diverse students as the students interacted with and responded to each item in a task type. The evidence, which included an examination of the students’ responses to items designed to measure receptive or expressive skills, determined whether the items elicited the intended responses.

This cognitive lab relied on the expert judgment of educators familiar with the students to administer the pilot form. The educators provided ratings of student performance on the items and responses directly into the structured protocol embedded in the *DFA* during test administration. Test examiners who best knew the students provided important evidence of student interaction and response modes for each pilot item and reviewed each task type linked to low, medium, or high levels of linguistic complexity. Additional evidence of the student characteristics was obtained from the student BIQ, which was completed prior to test administration by the test examiner. This study also gathered observational data of the test administration collected by two trained researchers on a structured protocol. The test examiner was interviewed immediately following the test administration to elicit more nuanced information about the testing experience. Follow-up questions from the *DFA* were used to elaborate on themes and clarify concepts. All interviews of test examiners who assented were tape-recorded to minimize loss of information.

The investigation of the interaction of task types used the following research questions to examine this topic:

* Do the task types gather appropriate evidence about the students’ English language skills?
* Are the task types appropriate for the target population?
* Do the task types measure the intended response process? If not, how can that be improved?
* What are the communication modes used to respond to the task types?

These four research questions will now each be discussed in turn, with data presented for each area of investigation.

#### Evidence of English Language Proficiency Skills

The first area of investigation explored test examiner and student interactions with the task types to determine whether the task types elicited the intended knowledge, skills, and abilities while minimizing construct irrelevant interference. This section provides the following explorations: the test examiner rating of each task type as it corresponds to the linguistic complexity levels (low, medium, high); themes related to task type investigations; student cognitive response data; test examiner ratings of students’ ELP; comparisons of student ELP ratings to overall ELP scores obtained on the pilot; and the knowledge, skills, and abilities of the participating students in the areas of listening, reading, speaking, and writing, separately and by ELP level.

##### Ratings of Task Types to Levels of Linguistic Complexity

The study investigated the task types by the assigned level of linguistic complexity. In accordance with the *Alternate ELPAC HLTD*, task types on the pilot test were organized in increasing levels of linguistic complexity (low, medium, and high) refer to table 1. Task types assigned to group 1 grades included kindergarten, grade two, and grade span six through eight, which piloted the following task types: *Recognize and Use Common Words*, *Understand a School Activity*, *Describe a Routine*, and *Interact with an Informational Text*. Task types assigned to group 2 included grade one, grade span three through five, and grade span nine through twelve, which piloted the following task types: *Communicate About Familiar Topics*, *Understand a School Exchange*, *Understand and Express an Opinion*, and *Interact with a Literary Text*. Task type descriptions are available for reference in [appendix A](#_Appendix_A_1). The findings are presented by the task type assignments following the design of the pilot blueprint for each grade and grade span and are presented in order of linguistic complexity.

During each test administration, after administration of the task type, test examiners rated each task type according to the corresponding level of linguistic complexity (low, medium, and high). Each test examiner was asked to respond to whether each task type represented an appropriate task for students functioning at the corresponding level of linguistic complexity (low, medium or high). A rating of “yes” represented test examiner agreement that the task type corresponded with the assigned level of linguistic complexity. If test examiner rating was no, test examiners disagreed that the linguistic complexity corresponded to the assigned level. Test examiners who selected “no” were asked follow-up questions to explain the disagreement.

Test examiner ratings of the task types were found to be generally appropriate for the intended population. All task types across all grade levels and grade spans ranged from 70 to 100 percent agreements, with the exception of two task types. Task types that received the highest levels of agreement from test examiners include: (1) *Recognize and Use Common Words* at 100 percent (kindergarten, grade two, and grade span six through eight) and (2) *Interact with an Informational Text* ranging from 83 percent (kindergarten) to 100 percent (grade two and grade span six through eight).

Two task types were associated with the lowest levels of agreement at a single grade level or grade span—*Communicate About Familiar Topics* at 54 percent (grade one) and *Understand a School Exchange* at 58 percent (grade one) and 50 percent (grade span nine through twelve). An examination of the data from test examiners for *Understand a School Exchange* in grade span nine through twelve and *Communicate About Familiar Topics* in grade one is associated with items that were flagged for students who are not yet verbal or were presymbolic users. Students who are presymbolic users have not yet demonstrated evidence of comprehending that a symbol, image, or picture represents a real object and need objects to access content. Some test examiners recommended that alternate response options be supported for expressive items. Test examiners flagged items for high linguistic text complexity, inappropriate topics for the population were cataloged for future item reviews. All remaining task types across all grade levels and grade spans had high levels of agreement ranging from 70 to 100 percent.

The following tables present the test examiner ratings for each task type for group 1 (kindergarten, grade two, and grade span six through eight) are presented first and followed by group 2 (grade one, grade span three through five, and grade span nine through twelve). The task type ratings are in order of arrangement on the blueprint.

**Group 1.** The first task type, *Recognize and Use Common Words*, at the low level of linguistic complexity, had the highest agreement across grade levels (100%; refer to table 9). Test examiners at early grade levels reported that this task type was really “appropriate for their students, especially those at the beginning stages of language development.” A test examiner in grade span six through eight reported that “this task type really helped the student warm up to the test.”

Table 9. Group 1, Task Type 1, *Recognize and Use Common Words* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **K** | **Gr 2** | **Gr 6–8** |
| Yes | 12 (100) | 11 (100) | 12 (100) |
| No | 0 (0) | 0 (0) | 0 (0) |

The ratings for *Understand a School Activity* are provided in table 10. Test examiners in three kindergarten administrations found some items were less appropriate for students at presymbolic levels of communication and recommended that students be given the opportunity to use objects. Several test examiners commented that early communicators may not have the “ability to respond to yes/no questions consistently.” Nearly all test examiners requested that more visuals be added to the items. A test examiner in grades six through eight recommended that pictures in the story should match the pictures in the answer choices.

Table 10. Group 1, Task Type 2, *Understand a School Activity* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **K** | **Gr 2** | **Gr 6–8** |
| Yes | 9 (75) | 11 (100) | 9 (82) |
| No | 3 (25) | 0 (0) | 2 (18) |

The ratings for the task type *Describe a Routine* indicated high levels of agreement that the task type was appropriate, as presented in table 11. Comments received by test examiners who reported that no, the task types were not appropriate for three kindergarten test administrations indicated the topics were unfamiliar, the length of the story was too long, and the vocabulary was too difficult for these three students.

Table 11. Group 1, Task Type 3, *Describe a Routine* (Medium Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **K** | **Gr 2** | **Gr 6–8** |
| Yes | 10 (83) | 11 (100) | 10 (100) |
| No | 2 (17) | 0 (0) | 0 (0) |

The ratings for the task type *Interact with an Informational Text* were positive and ranged from 83 to 100 percent agreement, as presented in table 12. Test examiners reported the linguistic complexity was high on some of these items. Conversely, at the higher grade levels, a test examiner commented, “Visuals and story are age appropriate and align with standards and curriculum administered in classroom.”

Table 12. Group 1, Task Type 4, *Interact with an Informational Text* (High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **K** | **Gr 2** | **Gr 6–8** |
| Yes | 10 (83) | 11 (100) | 10 (100) |
| No | 2 (17) | 0 (0) | 0 (0) |

**Group 2**. The following tables represent the data collected on the task types administered at the grade levels and grade spans for group 2 (grade one, grade span three through five, and grade span nine through twelve).

The ratings for *Communicate About Familiar Topics* produced high levels of agreement at grade spans three through five and nine through twelve (100%), while ratings for grade one were low (55%; refer to table 13). Test examiner ratings and comments were primarily in response to expressive items. Expressive skills are difficult for young kindergarten students and for students who are presymbolic communicators who may need concrete objects. Test examiners reported that the younger students do not have the ability to consistently respond to yes or no questions. One test examiner further recommended that choice making between two objects or pictures would have been better for that test examiner’s student. Another test examiner recommended, “Replace the yes/no questions; some kids just say yes to everything.”

Table 13. Group 2, Task Type 1, *Communicate About Familiar Topics* (Low Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **Gr 1** | **Gr 3–5** | **Gr 9–12** |
| Yes | 6 (55) | 10 (100) | 12 (100) |
| No | 5 (46) | 0 (0) | 0 (0) |

The ratings for *Understand a School Exchange* produced mixed results (refer to table 14). Test examiner ratings at grades three to five were positive at 80 percent agreement. However, at grade one, the expressive items were noted as being difficult for students who have no way to communicate responses because the students are “nonverbal.” Test examiner’s ratings at grade span nine through twelve indicated that all non-responses were associated with one item set that was too difficult, and the topic was not appropriate for the population.

Table 14. Group 2, Task Type 2, *Understand a School Exchange* (Low-to-Medium Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **Gr 1** | **Gr 3–5** | **Gr 9–12** |
| Yes | 7 (58) | 8 (80) | 5 (50) |
| No | 5 (42) | 2 (20) | 5 (50) |

The ratings for *Understand and Express and Opinion* are presented in table 15. The ratings for this task type were mostly positive. Test examiner ratings for three administrations at grade span nine through twelve did include comments for linguistic complexity and topic appropriateness which were flagged for review.

Table 15. Group 2, Task Type 3, *Understand and Express an Opinion* (Medium-to-High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **Gr 1** | **Gr 3–5** | **Gr 9–12** |
| Yes | 11 (92) | 8 (89) | 7 (70) |
| No | 1 (8) | 1 (11) | 3 (30) |

The ratings for *Interact with a Literary Text* represented high agreement (refer to table 16). Comments included a recommendation that “more visual support would improve student access” to stories.

Table 16. Group 2, Task Type 4, *Interact with a Literary Text* (Medium-to-High Level of Linguistic Complexity) Frequency and Percentage of Test Examiners’ Ratings of Correspondence from Each Test Administration

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **Gr 1** | **Gr 3–5** | **Gr 9–12** |
| Yes | 11 (92) | 7 (88) | 10 (91) |
| No | 1 (8) | 1 (13) | 1 (9) |

##### Themes to Improve Task Types

The interview protocol included questions to allow test examiners to elaborate on areas of improving task types and ways to address linguistic complexity. Collectively there was high agreement that task types corresponded appropriately to the assigned linguistic complexity level. Test examiners did have specific recommendations for improving the task types. Two themes of improvement emerged from the review of test examiner ratings and interviews.

The first theme to improve the task types is related to improving access to the content. A review found low ratings of task type appropriateness due to the linguistic complexity related to access to content. “Access is the interaction between construct-irrelevant item features and person characteristics that either permits or inhibits student response to the target measurement content of the item” (Winter et al, 2006, p. 276). Considerations of access for expressive items is critical for students who are presymbolic, use AAC devices, or have sensory disabilities.

The second theme to improve the task types are associated with construct irrelevance. Cho & So’s (2014) study of young English learners have similar findings and recognized the importance of minimizing construct irrelevance. Specific considerations to minimize construct irrelevance for this study are considerations of language complexity, text length, topic appropriateness, and ambiguity in pictures. Test examiners in this study indicated that more pictures would improve the accessibility of items for the students, particularly for expressive items. However, performance on items improved only when the visuals contained information essential for understanding and responding to the item. Visuals were found to be a source of error and distraction for students when the visuals were not purposeful to supporting a story or a response. Visuals depicting content can also be impacted by picture quality or picture choice. An example of ambiguity was observed from a young English learner who responded to a stimulus of a cup using the word “milk” in Spanish. The representation was a white cup, which the test examiner said is similar to an image being used to represent milk on the student’s communication board.

##### Student Response Data

This cognitive lab gathered evidence of student response processes to evaluate whether item response demands were introducing construct-irrelevant variance. Evidence of student response process can help test developers understand student-item interactions. Assertions that knowledge and skills demonstrated on an assessment reflect students’ true abilities require that assessment items must “elicit cognitive processes associated with the underlying cognitive model so that observed item responses can lead to valid inferences about the construct under investigation” (Ketterlin-Geller, 2008, p. 10). While this can be challenging for students with the most significant cognitive disabilities, is important in understanding how students make sense of assessment tasks.

A concurrent probing approach during test administration was used. Data was collected by the test examiner, and researchers observed the student during test administration. For the investigation of the student’s interaction of items, test examiners asked students the following probing questions:

“Was this question easy for you?”

“Did the pictures help you with your answer?”

“Did the story help you with your answer?” (for items with passages)

The test examiners could more fully engage with students to elicit more responses dependent on the students’ level of engagement. The data collected reflects that some test examiners did not administer all of the probes to the students. This was based on the individual student’s ability to respond. If no responses were provided by the student, the test examiner moved on to the next item. The recommendation was to abandon probing if no response was achieved after at least 6 of the 13 or 14 items. Student probes included using a yes-or-no response option.

Compiling data on the student responses across the three probes, only about half the students across grade levels and grade spans (*n* = 37) answered one or more of these questions. Other students either provided no responses or were consistently not administered. Chi-square analyses showed significant associations (p < .05) between whether there were student responses at all or not. For example, grade level and grade span showed nonresponse greater than expected in kindergarten and grade one; but students responses more often than expected in grade span nine through twelve (Cramer’s *V* = .56) and considered relatively strong. Analyses of verbal ability was associated with nonresponse greater than expected among nonverbal students (Goodman and Kruskal tau = .27), but not based on English language ability (p = .18; Cramer’s *V* = .23 and considered moderate) or EL and English only (EO) status (p = .89; Goodman and Kruskal tau = .00).

On average, students who responded to the probes with at least one or more responses earned more of the available points on the assessment (64.7%) than students who did not respond at all (45.0%)\*. Those who responded were predominantly verbal students and were predominantly EL students (both cases, 30 of 37; 81%). Additionally, the verbal group and the EO group were more likely to say yes to each of the three questions (> 90%) than their counterpart groups. \*For more information on percentage of points obtained on the test overall by ELP level, refer to table B1 and table B2. For the percentage of points obtained by verbal or nonverbal participants refer to table B7 and table B8, in [appendix B](#_Appendix_B_1).

Students were given the opportunity to respond at the end of the test session. Students were asked, “How did you feel about the test?” Students who responded reported by giving it a thumbs up; selecting the happy face; or verbally responding by saying “I feel good,” “I liked doing this,” or “Thank you for giving me the test.” Students were given an opportunity to respond to the question, “What else do you want to say about the test?” Some students could not respond to this question; however, a few comments were obtained. Comments from students included the following:

“It was beautiful!”

“It was great.”

“I liked the stories.”

“It was good to have pictures to help me read.”

Test examiners explicitly mentioned that “[they] appreciated the opportunity to have students to have a voice on their test too.”

##### Test Examiner Ratings of English Language Proficiency

This study gathered evidence of each student’s English language proficiency as rated by the test examiner. Prior to test administration, test examiners completed a student BIQ, and rated each individual student’s ELP. Test examiners were asked, “What is the student’s level of English language ability?” and ratings were on a scale of low, medium, high, or not sure. In consideration that test examiners may or may not have an understanding of English language proficiency for the students given this is a newly required state assessment, the intent was to determine whether test examiners were able to rate the students’ ELP. ELP ratings were missing from some student BIQs, and this is reflected in the tables in this section, where totals do not always equal 71. English only students were also rated.

The largest percentage of ELP ratings were at the medium level (41%) with equal proportions of low and high proficient students (25% respectively). The findings indicate that test examiners were able to assign a rating of English language ability over 90 percent of the time. The sample of students selected for observation appeared to fit the test design’s purpose for inclusion of the full range of ELP. The results are reported in table 17.

Table 17. Test Examiner Ratings of Students’ English Language Proficiency

|  |  |  |
| --- | --- | --- |
| **Rating** | **Number of Students** | **Percent** |
| Low | 17 | 25.0 |
| Medium | 28 | 41.2 |
| High | 17 | 25.0 |
| Not sure | 6 | 8.8 |
| **Total** | 68 | 100.0 |

##### Pilot Score Summaries and Distribution by ELP Ratings

The pilot scores were examined for the overall distribution of obtained scores. It is important for test design to support students across the full range of ELP. While the pilot was not a full test form, the test design is linear with task types that represent low, medium, and high levels of ELP.

The range of score performance across test takers is presented in table 18. The score summary data reflects that, overall, a range of performance was achieved. Students in grade two had scores ranging from a minimum of 1 point to a maximum of 15 points, and had a mean score of 10.42, which represents 57.9 percent of the available score points that were obtained by the 12 study participants. Kindergarten scores ranged from a minimum of 1 point to a maximum of 14 score points, and had a mean score of 5.25, which represents 29.2 percent of the available score points that were obtained by the twelve study participants.

Table 18. Score Summary for All Grade Levels and Grade Spans

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Overall Scores | N | Number of Items | Available Points | Minimum Score | Maximum Score | Mean Score | Std. Deviation | Mean % of Available Points |
| Kindergarten | 12 | 13 | 18 | 1 | 14 | 5.25 | 4.31 | 29.2 |
| Grade 1 | 12 | 14 | 20 | 0 | 17 | 7.58 | 4.62 | 37.9 |
| Grade 2 | 12 | 13 | 18 | 1 | 15 | 10.42 | 4.17 | 57.9 |
| Grade span 3–5 | 10 | 14 | 20 | 0 | 18 | 8.90 | 5.40 | 44.5 |
| Grade span 6–8 | 12 | 13 | 18 | 1 | 17 | 10.17 | 4.67 | 56.5 |
| Grade span 9–12 | 12 | 14 | 20 | 3 | 19 | 11.42 | 5.74 | 57.1 |

Score summaries are impacted by the student’s level of ELP. For example, kindergarten students rated at low levels of ELP impacted the overall mean of available points (see table 19). It is also worth noting that test examiners acknowledged some concern around alternate assessment decision making and the identification of students in the “early grades.”

The scores obtained were also investigated by comparison to teacher ratings of ELP. Test examiners assigned ELP ratings of low, medium, high, and not sure. The pilot scores obtained for each grade level and grade span have been disaggregated by the ELP levels for each grade and contain the number of students for each level of ELP, an average of the points obtained, the standard deviation and percent of available points obtained for each level of ELP, and are presented in table 19 and table 20. The total number of students include two students that completed the test administration with missing levels of ELP. One student refused to test.

Expected patterns include that students with medium-to-high levels of perceived ELP tended to attain higher scores; conversely, students rated with low ELP tended to obtain fewer points. The scoring data suggests that there is a progression of mean scores across the three levels of ELP (low, medium, and high) with the exceptions of grade one and grade span six through eight, which may be impacted by low numbers of participants at those grades.

Table 19. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Proficiency | KN | K Mean | K SD | K % Pts | Gr 2 N | Gr 2 Mean | Gr 2 SD | Gr 2 % Pts | Gr 6–‍8 N | Gr 6–‍8 Mean | Gr 6–‍8 SD | Gr 6–‍8 % Pts |
| Low | 5 | 2.00 | 1.73 | 11.11 | 5 | 8.20 | 5.36 | 45.56 | 1 | 11.00 | N/A | 64.71 |
| Medium | 5 | 8.00 | 4.64 | 44.44 | 5 | 12.00 | 2.55 | 66.67 | 5 | 9.00 | 2.24 | 52.94 |
| High | 1 | 9.00 | N/A | 50.00 | 1 | 14.00 | N/A | 77.78 | 5 | 12.20 | 6.42 | 71.76 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 10.00 | N/A | 55.56 | 1 | 5.00 | N/A | 29.41 |
| Missing | 1 | 4.00 | N/A | 22.22 | 0 | N/A | N/A | N/A | 0 | N/A | N/A | N/A |
| **Total** | 12 | 5.25 | 4.31 | 29.17 | 12 | 10.42 | 4.17 | 57.87 | 12 | 10.17 | 4.67 | 59.80 |

Table 20. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **Gr 1 N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–‍5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Low | 1 | 6.00 | N/A | 30.00 | 5 | 6.80 | 5.89 | 34.00 | 0 | N/A | N/A | N/A |
| Medium | 4 | 4.25 | 3.10 | 21.25 | 2 | 14.00 | 5.66 | 70.00 | 7 | 12.43 | 6.02 | 62.14 |
| High | 7 | 9.71 | 4.61 | 48.57 | 1 | 12.00 | N/A | 60.00 | 2 | 13.00 | 0.00 | 65.00 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 9.00 | N/A | 45.00 | 3 | 8.00 | 7.00 | 40.00 |
| Missing | 0 | N/A | N/A | N/A | 1 | 6.00 | N/A | 30.00 | 0 | N/A | N/A | N/A |
| **Total** | 12 | 7.58 | 4.62 | 37.92 | 10 | 8.90 | 5.40 | 44.50 | 12 | 11.42 | 5.74 | 57.08 |

Additional analyses examined the obtained receptive and expressive scores by ratings of ELP (see [appendix B](#_Appendix_B_1), table B1 and table B2 for receptive, and table B3 and table B4 for expressive). Expected patterns were obtained. Students accessed a higher percentage of the overall receptive points than expressive points. In kindergarten, for example, an overall of 38.10 percent of receptive points was obtained versus 25.83 percent of the expressive points. Students with high ELP ratings tended to obtain more of the available expressive score points overall. This is similar to findings of Karvonen & Clark (2019) that expressive communication is a strong predictor of students’ overall performance. Students at low ELP ratings received lower expressive scores overall.

Test administration observations provided insight into the obtained scores, the ratings of students’ ELP, and the task type levels. Students identified at low levels of ELP were not able to correctly answer items at the medium-to-high linguistic complexity levels. Conversely, students who were EOs or ELs with ratings of high levels of ELP, as categorized by those students’ teachers, did quite well.

The student performance on the task types was found by test examiners to be commensurate with the ELP skills demonstrated by the same students during instruction. For example, during a test observation, one student, Maria (a pseudonym), an EL who had enrolled 13 months prior, was assigned as having a low level of English ability by her teacher. During the test session, Maria was able to speak in one-to-three-word phrases in Spanish. The paraprofessional’s opinion was that Maria was better able to communicate in Spanish, but that she did understand some English. This was a consistent observation during the test session. Maria was not able to demonstrate English on items for task types at the medium and high levels of linguistic complexity. However, Maria was able to point to answers for items on the task type *Recognize and Use Common Words*, designed for low levels of ELP.

##### Knowledge, Skills and Abilities

A primary component of ECD is to understand the knowledge, skills, and abilities of test takers in the test population (Mislevy, Almond & Lukas, 2003, p. 20). At the heart of assessment validity is the importance of cognition, which is defined as a theory or set of beliefs about how students represent knowledge and develop competence, which is supported by describing students and patterns of learning characteristics (Marion & Pellegrino, 2006). This study examined the knowledge, skills, and abilities of participants from data collected from the student BIQ to learn about the knowledge, skills, and abilities of individual students in the areas of receptive and expressive language skills. The importance of collecting useful information on the knowledge, skills, and abilities of the test takers is useful evidence that provides understanding of the language skills of the test takers and provides a way to inform whether the linguistic complexity of the tasks support the range of test takers and those students’ ability to use English.

The section provides information on the knowledge, skills, and abilities of study participants. The student BIQ was completed by the test examiner for each student. The aggregate data obtained on students’ knowledge, skills, and abilities were rated and are presented according to receptive (listening, reading) and expressive (speaking, writing) skills. The data is presented in table 21 through table 24. Due to missing data, not all totals are equal to 71.

The student BIQ completed by the test examiner rated the receptive communication and listening skills for 68 students. Students were rated at what best describes the students’ listening skills. The highest area of listening skills for study participants is an ability to follow one-step directions (35.3%) and two-step directions (30.9%). Results are presented in table 21.

Table 21. Students’ Listening Skills

|  |  |  |
| --- | --- | --- |
| **Listening Skills** | **Frequency** | **Percent** |
| Follows 2 step directions | 21 | 30.9 |
| Follows 1 step directions | 24 | 35.3 |
| Attends and responds to simple commands | 9 | 13.2 |
| Performs a simple action upon request | 4 | 5.9 |
| Indicates a choice when offered an array of items | 2 | 2.9 |
| Points to or touches objects upon request | 6 | 8.8 |
| Does not attend to sound | 2 | 2.9 |
| Other | 0 | 0 |
| **Total** | 68 | 100.0 |

The student BIQ completed by the test examiner rated the reading skills for 68 students. Students were rated at what best describes the students’ reading skills. The students rated highest in the reading of words, phrases, and sentences when pictures or symbols are provided (15%). Few students were able read text without pictures or symbols. Nearly 33 percent of the students recognized letter or letter sounds. Refer to the receptive reading skills data presented in table 22.

Table 22. Students’ Reading Skills

|  |  |  |
| --- | --- | --- |
| **Reading Skills** | **Frequency** | **Percent** |
| Reads text without any picture or symbol support with comprehension | 3 | 4.4 |
| Reads text without picture or symbol support but without comprehension | 4 | 5.9 |
| Identifies individual words without picture support | 3 | 4.4 |
| Reads words, phrases, or sentences when pictures or symbols are provided with unfamiliar words | 10 | 14.7 |
| Recognizes letter sounds | 8 | 11.8 |
| Recognizes letters | 14 | 20.6 |
| Matches objects to pictures | 15 | 22.1 |
| Identifies and names objects | 5 | 7.4 |
| Does not yet have an understanding of print or text | 6 | 8.8 |
| Other | 0 | 0 |
| **Total** | 68 | 100.0 |

The student BIQ completed by the test examiner rated the speaking skills for 68 students. Students were rated on the basis of what best describes the students’ expressive skills. Speaking skills were highest at verbally speaking two or more words (25%). Overall, 62 percent of the students were able to express themselves verbally speaking one or more words (refer to table 23). All categories indicated communicative intent, with the exception of two students who were not yet intentionally communicative.

Table 23. Students’ Speaking Skills

|  |  |  |
| --- | --- | --- |
| **Speaking Skills** | **Frequency** | **Percent** |
| Verbally speaks 3 or more words in complete sentences using grammatical rules | 8 | 11.8 |
| Verbally speaks 2 or more words in sentences or phrases without consistently following grammatical rules | 17 | 25.0 |
| Verbally speaks 2-word phrases | 13 | 19.1 |
| Verbally speaks 1 word at a time | 4 | 5.9 |
| Uses touch and gestures by pointing and head nodding | 8 | 11.8 |
| Uses vocalizations, gestures, and facial expressions to communicate intentionally | 13 | 18.6 |
| Uses eye gaze with intentionality | 1 | 1.5 |
| Uses AAC | 2 | 2.9 |
| Not intentionally communicative | 2 | 2.9 |
| Other | 0 | 0 |
| **Total** | 68 | 100.0 |

The student BIQ was completed by the test examiner who rated the writing skills for 68 students. Students were rated at what best describes the students’ expressive (writing skills). The ratings for writing skills were the lowest skill ratings overall. Few students could write words and sentences (15%). Most students were able to copy letters and words but were not producing independent writing (26%; refer to table 24).

Table 24. Students’ Writing Skills

|  |  |  |
| --- | --- | --- |
| **Writing Skills** | **Frequency** | **Percent** |
| Writes 1–3 sentences (spelling not always correct) | 1 | 1.5 |
| Writes a simple sentence or phrase (spelling not always correct) | 2 | 2.9 |
| Writes words, spelling not always correct | 7 | 10.3 |
| Writes using word banks | 3 | 4.4 |
| Selects letters or symbols to express meaning | 1 | 1.5 |
| Copies letters and words but does not produce independent writing | 18 | 26.5 |
| Randomly selects letters or symbols when asked to write | 3 | 4.4 |
| Makes random marks or scribbles | 21 | 30.9 |
| Does not yet demonstrate expressive writing skills | 11 | 16.2 |
| Other | 1 | 1.5 |
| **Total** | 68 | 100.0 |

The overall ratings of listening, reading, speaking and writing skills were compared by test examiner ratings of English language proficiency levels (low, medium, and high). The comparison provides a way to determine the range of skills for each level of English proficiency (refer to [appendix C](#_Appendix_C)).

#### Appropriateness for the Population

The second research question with respect to interaction with task types was whether the task types were appropriate for the population. Overall, all the task types across grade levels and grade spans were found to be appropriate. During the interviews, teachers spoke positively about the task types and the progression of task types (e.g., “The task types are great!” and “This is the best alternate assessment I’ve ever given, very accessible.”). One test examiner noted that “many students had no experience testing as they have never made it through for the CAAs because of the response check and have not tested previously.”

A few test examiners did note, however, that the progression of language complexity was too rapid for a number of students as those students moved from low to medium task types. Test examiners found that the task type *Recognize and Use Common Words*, which was included in the pilot test but was not (at that time) planned for inclusion in the operational test blueprint, provided students a valuable opportunity to demonstrate ELP at the lowest level of linguistic complexity. Test examiners believed this task type could improve test participation on the assessment, especially for students who were performing at low levels of linguistic complexity, students who were presymbolic learners, and students with sensory deficits.

The appropriateness of the task types for students is presented in table 25. While some of the grade levels reflected high levels of agreement (e.g., grade two) others reflect more mixed ratings (particularly kindergarten and grade span three through five). A review of test examiner comments suggests reducing the linguistic complexity and considering adding objects and pictures for some items in kindergarten. Topic appropriateness on an item set was also mentioned.

Table 25. Were the Tasks Appropriate for Your Students? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **G 9–12** |
| Yes | 8 (67) | 10 (83) | 12 (100) | 6 (60) | 11 (92) | 10 (83) |
| No | 4 (33) | 2 (17) | 0 (0) | 4 (40) | 1 (8) | 2 (17) |

The ratings of stories and passages, as compiled from the Test Examiner Protocol, are presented in table 26. A review of these data indicates that the stories were appropriate in linguistic complexity. Yet, topic appropriateness was mentioned for item sets at kindergarten, grade one, and grade spans three through five and six through eight. The length of passages was a concern in early grades.

Table 26. Were the Stories or Passages Appropriate in Length, Complexity, and Topic? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 7 (58) | 6 (50) | 10 (83) | 6 (60) | 6 (50) | 11 (83) |
| No | 5 (42) | 6 (50) | 2 (17) | 4 (40) | 6 (50) | 1 (17) |

Ratings were obtained from test examiners regarding the use of the types of questions used instructionally, as presented in table 27, as compiled from the Test Examiner Protocol. The ratings of agreements obtained ranged from 75 to 100 percent. These ratings support the idea that the questions in the task types are similar to those used instructionally, which supports appropriateness.

Table 27. Do You Ask Students These Types of Questions During Instruction? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 11 (92) | 9 (75) | 11 (92) | 9 (90) | 12 (100) | 11 (92) |
| No | 1 (8) | 3 (25) | 1 (8) | 1 (10) | 0 (0) | 1 (8) |

#### Measurement of Intended Response

Overall, test examiners found that the task types measured the intended response processes of receptive or expressive skills appropriately across the grade levels and grade spans. Test examiners indicated that items accurately represented the measurement targets and were strongly aligned with both receptive skills (listening and reading) and expressive skills (speaking and writing).

A few test examiners indicated uncertainty whether writing was specifically included as an expressive measurement target. Expressive items did have some limitations for students who were nonverbal and presymbolic communicators. Test examiners requested more response options to include realia, manipulatives, or objects for presymbolic communicators and to provide additional picture support for AAC users.

In the *DFA,* test examiners were asked to rate whether each item measured receptive or expressive skills (yes, no, not sure). Rating agreements were high for both sets but somewhat higher for receptive skills. Please refer to data presented for agreement on measurement of the intended skills by items for receptive skills in table 28.

Table 28. Does This Item Measure Receptive Skills (Listening and Reading)? Average Percentages of All Items by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 100% | 100% | 98% | 96% | 100% | 98% |
| No | 0% | 0% | 2% | 4% | 0% | 2% |

The same data is presented for expressive skills in table 29.

Table 29. Does This Item Measure Expressive Skills (Speaking and Writing)? Average Percentages of All Items by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 96% | 90% | 89% | 93% | 96% | 98% |
| No | 4% | 10% | 11% | 7% | 4% | 2% |

#### Test Examiner Ratings of Item Difficulty

Ratings of item difficulty were collected to provide evidence that the items were appropriate for the full range of students and to confirm whether the task types targeted the intended response processes. Each test examiner rated each of the items for difficulty in the embedded research questions in the *DFA*. The ratings for item difficulty included whether the item was “easy”, “about right,” or “hard” for the individual student. If the test examiner indicated “hard,” a checklist was provided with the following options: “vocabulary too high,” “culturally biased,” “too difficult at this grade,” and “other, fill in the blank.” Ratings were useful in identifying whether an item was just viewed as too difficult for a particular student or whether the item was too difficult at the grade or grade span. Items were rated for difficulty and averaged at each grade level or grade span.

Some comments were received about the “test” with test examiners believing that the pilot was meant to approximate an operational test form rather than an item tryout. Vertical alignment concerns was an issue shared by a few test examiners. For example, one test examiner stated that the “test for grade span three through five was harder than the test for grade span six through eight.” While a different test examiner stated that the grade span “three through five test works well for grades four through five but may be too hard for grade three.”

The proportion of items rated as “hard” at the highest levels was in kindergarten and grade one, where test examiners indicated that the items were difficult. Ratings of “hard” received comments that items were difficult and too linguistically complex for those grade levels. Items rated as “hard” were found across all grade levels and grade spans predominantly for expressive items where only text and no visual support was provided. Students were observed trying to find a picture to point to on the screen. Test examiners found expressive items were not accessible and too difficult for students who could not provide a verbal response. Test examiners requested that the test include more items with objects and pictures.

The test examiner ratings for the items are presented by percent of average of frequency in each response category in table 30.

Table 30. Was the Item for This Student (Easy, About Right, Hard)? Percent Average of Frequency of Selection Across All Items by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Easy | 9.5 | 7.7 | 13.0 | 11.0 | 18.3 | 23.1 |
| About Right | 32.5 | 40.1 | 55.6 | 31.3 | 47.3 | 34.1 |
| Hard | 49.1 | 44.0 | 19.5 | 28.0 | 24.3 | 26.4 |
| Did Not Administer | 8.3 | 0.5 | 2.4 | 4.4 | 0.6 | 1.1 |

**Note:** Average percentages in table 30 include only the valid count of response option selections. Due to rounding, row percentages may not add up to exactly 100 percent.

Data was collected on whether the students performed as expected or responded in a way the students would be expected to respond. The data in table 31 shows that across all items by grade levels and grade spans, the large majority of students (minimum = 87 percent) performed as expected by the test examiner.

Table 31. Did This Student Perform as Expected? Percent Average of Frequency of Selection by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 92 | 95 | 89 | 90 | 87 | 89 |
| No | 8 | 5 | 11 | 10 | 13 | 11 |

#### Test Administration Time

The information on test administration time was collected to determine the time spent on specific task types, and the time spent for each intended response. Response-time data was collected to provide additional evidence for whether students responded as intended to examine if lengthier response times may have indicated challenges in the task types or accessing the content. The study collected test administration times for each task type, for grade level or grade span. The times are presented in ranges and average times in minutes, refer to table 32 and table 33.

Timing trends indicate that times for total test administration was reduced for students at higher grade levels. Expressive items tended to take slightly more time than receptive items for group 2, the opposite was true for group 1. Test examiner perceptions of test length is addressed later in this report.

The following abbreviations are used in table 32 and table 33.

*RaUC = Recognize and Use Common Words*

*USA = Understand a School Activity*

*DaR = Describe a Routine*

*IIT = Interact with an Informational Text*

K = kindergarten

Table 32. Test Administration Times in Minutes (Group 1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Items** | **K N** | **K M (SD)** | **K Range** | **Gr 2 N** | **Gr 2 M (SD)** | **Gr 2 Range** | **Gr 6–‍8 N** | **Gr 6–8 M (SD)** | **Gr 6–8 Range** |
| **Overall:** Total | 13 | 13 | 24.6 (13.0) | 11.0–60.0 | 12 | 28.2 (11.8) | 16.0–53.0 | 12 | 20.9 (8.8) | 9.0–42.0 |
| **Overall:** Items | 13 | 12 | 8.6 (3.7) | 3.5–15.6 | 12 | 7.9 (3.0) | 4.2–13.2 | 12 | 7.0 (2.9) | 3.0–11.9 |
| **Overall:** Diff. | 13 | 12 | 13.1 (6.1) | 5.1–22.4 | 12 | 20.3 (10.6) | 10.8–42.4 | 12 | 13.9 (7.7) | 2.0–30.4 |
| **Task Type:** *RaUC* | 2 | 12 | 1.9 (1.5) | 0.1–4.2 | 11 | 1.0 (1.0) | 0.2–3.4 | 12 | 0.8 (1.0) | 0.1–3.7 |
| **Task Type:** *USA* | 4 | 12 | 2.7 (1.7) | 0.9–6.3 | 12 | 1.7 (1.0) | 0.6–4.2 | 12 | 2.0 (1.4) | 0.8–4.9 |
| **Task Type:** *DaR* | 3 | 12 | 1.7 (0.8) | 0.6–3.9 | 12 | 2.2 (1.3) | 0.9–6.0 | 12 | 1.7 (0.8) | 0.5–3.0 |
| **Task Type:** *IIT* | 4 | 12 | 2.3 (1.2) | 0.8–5.1 | 12 | 3.1 (1.4) | 1.4–5.4 | 12 | 2.5 (1.2) | 0.7–5.6 |
| **Skill Domain:** Receptive | 7 | 12 | 4.6 (2.4) | 1.7–9.0 | 12 | 4.6 (2.7) | 1.6–9.7 | 12 | 3.5 (1.9) | 1.3–6.8 |
| **Skill Domain:** Expressive | 6 | 12 | 4.0 (1.4) | 1.8–6.6 | 12 | 3.3 (1.0) | 1.8–5.5 | 12 | 3.5 (1.7) | 1.2–7.2 |

Table 33. Test Administration Times in Minutes (Group 2)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Items** | **G1 N** | **G1 M (SD)** | **G1 Range** | **GS 3–5 N** | **GS 3–5 M (SD)** | **GS 3–5 Range** | **GS 9–‍12 N** | **GS 9–12 M (SD)** | **GS 9–12 Range** |
| **Overall:** Total | 14 | 12 | 24.1 (7.6) | 12.0–35.0 | 10 | 29.0 (10.4) | 17.0–47.0 | 12 | 33.3 (17.6) | 11.0–64.0 |
| **Overall:** Items | 14 | 12 | 8.6 (4.5) | 4.0–16.6 | 10 | 11.0 (5.6) | 3.9–23.0 | 12 | 9.8 (4.4) | 3.3–18.9 |
| **Overall:** Diff. | 14 | 12 | 15.5 (5.4) | 6.9–26.9 | 10 | 18.1 (6.4) | 8.8–29.6 | 12 | 23.6 (17.7) | 2.2–57.7 |
| **Task Type:** *CAFT* | 2 | 11 | 0.8 (0.8) | 0.3–3.0 | 10 | 1.5 (1.1) | 0.1–4.0 | 12 | 0.8 (0.5) | 0.2–1.5 |
| **Task Type:** *UaSE* | 4 | 12 | 2.7 (1.8) | 1.0–6.3 | 10 | 3.9 (2.3) | 1.3–7.4 | 12 | 3.1 (2.3) | 0.5–7.6 |
| **Task Type:** *UEO* | 4 | 12 | 2.5 (1.5) | 1.0–5.8 | 9 | 3.5 (1.8) | 1.6–6.8 | 11 | 3.3 (1.1) | 1.6–6.1 |
| **Task Type:** *ILT* | 4 | 12 | 2.7 (1.4) | 0.4–5.7 | 9 | 2.7 (1.9) | 0.4–7.0 | 11 | 3.1 (2.0) | 0.9–7.2 |
| **Skill Domain:** Receptive | 6 | 12 | 3.9 (2.1) | 1.7–7.4 | 10 | 5.1 (2.4) | 2.0–10.6 | 12 | 4.9 (2.6) | 1.1–8.1 |
| **Skill Domain:** Expressive | 8 | 12 | 4.7 (2.7) | 0.5–9.3 | 10 | 5.9 (3.3) | 1.9–12.4 | 12 | 4.9 (2.6) | 1.1–10.7 |

#### Communication Mode Usage

The study investigated the communication mode each student used to respond to the tasks. Students in the pilot used the full range of communication modes—verbal responses, vocalizations, gestures, AAC devices, eye gaze, and sign language. Data for the communication modes was collected from the student BIQ; due to missing data, the totals will not add up to 71 participants. Approximately half of the students participating in the pilot were able to use oral speech to communicate. Students who were nonverbal communicators had limited or no access in responding to many expressive items. For example, students who were presymbolic communicators had difficulty accessing text-only expressive items. Test examiners noted that open-ended expressive items could be made accessible to students if objects, picture cards, or comparable alternate response options were provided. In this study, no students responded in braille or in writing.

The following communication modes were used in this study (listed in order of frequency):Verbal, using three or more words including complete sentences; Verbal, two-word responses, phrases; Verbal, one-word responses; Gestures (e.g., pointing, nodding, touching, arranging); Vocalizations; AAC devices; eye gaze; and American Sign Language (ASL) or other signed response (e.g., Signed Exact English). Table 34 provides information on the communication mode usage across all students. Totals do not add up to 71 due to missing information.

Table 34. Students’ Primary Communication Mode

|  |  |  |
| --- | --- | --- |
| **Communication Mode** | **Frequency** | **Percent** |
| Verbal: 3 words or complete sentences | 21 | 30.9 |
| Verbal: 2-word responses | 13 | 19.1 |
| Verbal: 1-word responses | 9 | 13.2 |
| Gesture | 12 | 17.6 |
| Vocalizations | 5 | 7.4 |
| AAC | 4 | 5.9 |
| Eye Gaze | 1 | 1.5 |
| No established mode of communication | 3 | 4.4 |
| **Total** | 68 | 100.0 |

EO participants were slightly higher in verbal skills than the participating ELs in this study. Data collected for 11 participants who were EO is reflected in table 35.

Table 35. English Only Students’ Primary Communication Mode

|  |  |  |
| --- | --- | --- |
| **Communication Mode** | **Frequency** | **Percent** |
| Verbal: 3 words or complete sentences | 5 | 45.5 |
| Verbal: 2-word responses | 3 | 27.3 |
| Verbal: 1-word responses | 1 | 6.8 |
| Vocalizations | 1 | 6.8 |
| AAC | 1 | 6.8 |
| Eye Gaze | 1 | 6.8 |
| **Total** | 11 | 100.0 |

The pilot scores obtained for each grade and grade span were examined by verbal and nonverbal scores, and contains the number of students, an average of the points obtained, standard deviation and percent of available points obtained are presented in table 36 and table 37. The communication modes were collapsed into two categories to examine whether students’ communication mode influenced receptive or expressive scores. The verbal category included the following communication modes: verbal, three words or complete sentences; verbal, two-word responses, phrases; verbal, one-word responses. The nonverbal category included all other communication modes (vocalizations, gestures [e.g., pointing, nodding, touching, arranging], AAC devices, eye gaze, and ASL or other signed response). No student used braille or writing.

The expected pattern clearly indicates that students able to communicate verbally outperformed students who used other forms of communication. This emphasizes the importance in establishing a form of communication for students who are nonverbal. The investigation of AAC use was limited due to the small numbers of study participants who used a communication device. Data on receptive and expressive scores obtained for each grade and grade span by verbal and nonverbal scores confirmed similar patterns for expressive scores and are available in [appendix B](#_Appendix_B_1).

Table 36. Total Scores Compared by Verbal and Nonverbal Students (Group 1)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–‍8 Mean** | **Gr 6–‍8 SD** | **Gr 6–‍8 % Pts** |
| Nonverbal | 8 | 2.88 | 1.81 | 15.97 | 2 | 4.00 | 4.24 | 22.22 | 3 | 5.33 | 4.51 | 31.37 |
| Verbal | 4 | 10.00 | 3.92 | 55.56 | 10 | 11.70 | 2.87 | 65.00 | 9 | 11.78 | 3.63 | 69.28 |
| **Total** | 12 | 5.25 | 4.31 | 29.17 | 12 | 10.42 | 4.17 | 57.87 | 12 | 10.17 | 4.67 | 59.80 |

Table 37. Total Scores Compared by Verbal and Nonverbal Students (Group 2)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **Gr 1 N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–‍5 Mean** | **Gr 3–‍5 SD** | **Gr 3–‍5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Nonverbal | 7 | 5.43 | 3.05 | 27.14 | 6 | 5.50 | 3.15 | 27.50 | 5 | 7.40 | 4.83 | 37.00 |
| Verbal | 5 | 10.60 | 5.03 | 53.00 | 4 | 14.00 | 3.65 | 70.00 | 7 | 14.29 | 4.68 | 71.43 |
| **Total** | 12 | 7.58 | 4.62 | 37.92 | 10 | 8.90 | 5.40 | 44.50 | 12 | 11.42 | 5.74 | 57.08 |

### Accessibility

The second area of investigation examined accessibility on the Alternate ELPAC. Accessibility is a critical component of the test development process and helps to ensure the testing experience results in valid and meaningful interpretations of a student’s ELP. The *Standards for Educational and Psychological Testin*g (AERA, APA, & NCME, 2014) recommended that accessibility and universal design considerations are essentially intermeshed in the efforts to create accessible assessments (AERA, APA, & NCME, 2014). As applied to the Alternate ELPAC, accessibility is a set of comprehensive approaches to improve access for ELs with the most significant cognitive disabilities, so students have the opportunity to demonstrate language skills through fair, valid, and equitable testing opportunities.

The need to minimize bias and increase fairness is a requirement for valid and meaningful interpretations to ensure that eligible students’ performance on the assessment was reflective of those students’ ELP and English-based communication strategies rather than the students’ disability status. These considerations for validity and fairness are especially critical for the development of the Alternate ELPAC as they relate to access needs for domains, test administration considerations, and accessibility resources.

Accessibility resources include embedded and non-embedded universal tools, designated supports, and accommodations and any other AAC devices used instructionally by the student. These resources may be available to all students or they may be selectively available depending on a student’s IEP. This study investigated the accessibility resources used during test administration. To that end, the following research questions were investigated:

* Which accessibility resources from the *English Language Proficiency Assessments for California (ELPAC) Accessibility Resources for Operational Testing* were used during administration?
* To what extent do accessibility resources support access for the individual students as the students interact with and respond to each task type?
* What other materials or resources were used to provide access to support student responses to the task types?
* For students who independently use a computer or tablet, to what extent is the interface accessible?
* How do teachers perceive the effectiveness of task types and administration for students who participate using a range of communication modes?

The findings are presented in the following section.

#### Accessibility Resources

The Alternate ELPAC adopted the multitiered accessibility resources model used for the CAASPP System of assessments. Students had access to both embedded and non‑embedded resources that were appropriate for the Alternate ELPAC construct. The study investigated which accessibility resources were used during test administration. The accessibility resources included the following categories:

* Universal tools, which are available for all eligible students
* Designated supports, which are available for some eligible students with teacher or parent/guardian judgment
* Accommodations, which are available to eligible students with appropriate documentation such as an IEP

The overall aggregated usage of accessibility resources is presented in table 38 through table 40. Overall usage of accessibility resources was obtained by adding all resources used across individual students at all grade levels and grade spans. No other resources were used on the pilot.

The standardized administration procedures for the Alternate ELPAC provide a one-on-one‑‑ test administration where the test examiner usually enters the student responses. Students were encouraged to navigate the test session if the students were able to enter responses independently; however, few students were able to navigate the test session independently, and many universal tools were not accessed or needed. The Alternate ELPAC allows test examiners to provide students with individual breaks whenever necessary. Breaks over more than one day are permitted. The pilot at each grade level and grade span included 13–14 items that the majority of students completed in one test session, although a few students had to have short breaks.

Table 38 includes the usage of universal tools across all grade levels and grade spans.

The universal tools that were used by test examiners supported access and improved the attention and focus of students. Many test examiners had to be shown by a research team member how to navigate and use these tools to support the best presentation of the test items. Data in table 38 was taken from the Student Observation Protocol.

Table 38. Accessibility—Universal Tools: Average Percentage of Use Across All Items and Grade Levels and Grade Spans

|  |  |
| --- | --- |
| **Universal Tool** | **Average Percentage** |
| Expandable items | 32 |
| Expandable passages | 16 |
| Zoom (in or out) | 15 |
| Breaks | < 1 |

The following universal tools were available but not used by any students: digital notepad, highlighter, keyboard navigation, line reader (grades three through twelve), mark for review, oral clarifications of test directions by the test examiner in English, scratch paper, strikethrough (grades three through twelve); and writing tools (grades three through twelve).

Designated supports were available to all students when determined for use by an educator or a team of educators (with parent/guardian and student input, as appropriate) or specified in the student's IEP.

Separate setting involves changing the test location so the student is tested in a setting different from that made available for most students (ETS, 2019). Separate setting also may include the most beneficial time of day, special lighting or acoustics, and adaptive furniture. During test administration, students were observed to particularly enjoy the opportunity to have one-on-one time with the test examiner when in a separate setting.

Print on demand (POD) was observed to be used for optional purposes. For example, test examiners used POD to produce hard copy versions of the images to support students using AAC devices. POD was also observed to be substituted as a paper form of the test. POD does not capture every stimulus that is presented on the screen. In this instance the test examiner read the stimuli and pointed to the POD. Test examiners recommended having access to a larger version of the POD feature to retain image quality. Table 39 includes the usage of designated supports across all grade levels and grade spans.

Table 39. Accessibility—Designated Supports: Average Percentage of Use Across All Items and Grade Levels and Grade Spans

|  |  |
| --- | --- |
| Designated Support | Average Percentage |
| Separate setting | 66 |
| Print on demand | 8 |
| Streamline | 2 |
| Masking | < 1 |
| Translated test directions | < 1 |
| ASL or Manually Coded English (MCE) for test directions | < 1 |
| Simplified test directions | < 1 |
| Magnification | < 1 |

The following designated supports were available but not used by any students: amplification, color contrast, color overlay, designated interface assistant, magnification, medical supports, mouse pointer, noise buffer, and turn off any universal tools.

Accommodations on the Alternate ELPAC included the usage of alternate response options, breaks, scribe, and ASL or MCE for content and responses. No students in this study used braille or speech-to-text. Alternate response options included, but were not limited to, the following: adapted keyboards, large keyboards, sticky Keys, mouse Keys, filter Keys, adapted mouse, touch screen, head wand, and switches.

Test administration observations found high variability in the delivery of the provision of alternate response options. The *DFA* did not contain specific guidance on the use of alternate response options. The provision was variable and not standardized. Test examiners commented that more guidance is needed on when and how to provide for students who use alternate response options that include forms of AAC. For example, some test examiners tried to personally create materials for students who used nondynamic devices (e.g., communication boards, Picture Exchange Communication Systems). The provision included test examiners who personally created pictures or icons, some who copied from images contained in the POD version, and others who did not provide any additional support not specifically stated inside the *DFA*. For those students who needed alternate response options, no guidance was provided to consistently support access for expressive items. One test examiner stated that “assistive technology helps make the students’ ability to answer more accessible.”

Low technology communication aids provide communication access to students with speech, hearing, or cognitive impairments. Low technology communication aids are not electrical or battery operated and are displayed in a book of pictures or symbols that are used for communication used to access vocabulary.

High-technology AAC refers to computerized devices such as dedicated communication devices (e.g., Tobii Dynavox). High technology usually includes displays that are fixed or dynamic. Fixed displays are static and do not change when a picture or symbol is selected. A fixed display can be used with no-tech, low-tech, mid-tech, or high-tech systems. A dynamic display changes when an icon is selected. A dynamic display is used with high-tech systems, computerized devices, and touch screens. Dynamic displays allow for programming of extensive amounts of vocabulary because each page can be embedded into another. The need for categorization is important on such devices and creates the need for more time for responding.

Test examiners sought guidance from the research team on the use of low and high technology, including the use of fixed, dynamic, or hybrid devices. One concerned test examiner commented, “The AAC device was not programmed with folders or icons needed for options for my student to respond to these kind of questions.” Another test examiner noted that “the lack of material availability to make or adapt the device was not possible given the amount of time that I had, so the consideration of AAC was not able to be provided, which made it impossible for my student to respond to some of the items.”

Varying approaches to addressing the provision of alternate response options were concerning to test examiners who wanted to support the students but were not sure what was allowed and if the actions being taken would be a violation of test administration protocols. Several test examiners asked for “picture cards,” in particular, for accessing the expressive items.

Breaks are allowed on the Alternate ELPAC whenever a student needs one, throughout the day, and over several days as appropriate for the student. Test examiners did pause the test to provide brief breaks for the students. Only one student required the test to be paused and resumed the next day, but that only because school ended early that day.

No students on the pilot used braille or the speech-to-text accommodation. Accommodations used are presented in table 40.

Table 40. Accessibility—Accommodations: Average Percentage of Use Across All Items and Grade Levels and Grade Spans

|  |  |
| --- | --- |
| Accommodation | Average Percentage |
| Alternate response options | 11 |
| Breaks | 4 |
| Scribe | < 1 |
| ASL/MCE for content and responses | < 1 |

Issues with assistive technology varied across test administration. While some test examiners provided picture support for students using alternate response options; many test examiners provided no support for access and commented, “It must not be allowed.” While the number of students impacted by issues with assistive technology are low, these students represent an important part of the population where access is critical. Test examiners want guidance around what is and is not allowed for alternate response options, and test examiners commented that the *DFA* did not specifically state what is allowed or not allowed. Refer to table 41 for data on grades impacted by issues with assistive technology use.

Table 41. Issues with Assistive Technologies? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Span** | **Yes N (%)** | **No N (%)** |
| K | 2 (15) | 11 (85) |
| Grade 1 | 0 (0) | 12 (100) |
| Grade 2 | 0 (0) | 12 (100) |
| Grade span 3–5 | 2 (20) | 8 (80) |
| Grade span 6–8 | 1 (8) | 11 (92) |
| Grade span 9–12 | 1 (8) | 11 (92) |

Test examiners were asked to rate their level of familiarity with the accessibility resources for the Alternate ELPAC. Test examiner ratings of familiarity indicated no test examiner was very familiar and that the majority were not at all familiar. The ratings are presented in table 42. The test examiners reported this for each student so the numbers reflect test administrations. The test examiners received training and *DFAs* just before the pilot. Test examiners were not asked if they had used the guidance on accessibility resources in previous test administrations.

Table 42. How Familiar Are Test Examiners with the Accessibility Resources for the Alternate ELPAC? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Response | K | Gr 1 | Gr 2 | Gr 3–5 | Gr 6–8 | Gr 9–12 |
| Very Familiar | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Somewhat Familiar | 2 (17) | 4 (36) | 2 (17) | 3 (30) | 7 (58) | 3 (25) |
| Not At All Familiar | 10 (83) | 7 (64) | 10 (83) | 7 (70) | 5 (42) | 9 (75) |

#### Accessibility Resources and Accessing the Test

The study investigated how the accessibility resources supported access for students as the students interacted with the test. Considerations for access to content must include principles of universal design that consider engagement, representation, and expression (CAST, 2018).

The test examiners identified issues that impacted accessibility, such as text-based items without images, expressive items that did not provide a way for students to respond, and linguistically complex items. Other accessibility concerns included the quality of the technology available for testing and how to use the technology appropriately. Many test examiners expressed that test administration occurring on larger screen sizes would be better for the students. Test examiners commented, “Large touchscreens or smart boards are best for students, especially students with fine motor impairments.” However, for many test examiners, these were not available for use. Conversely, several test examiners said the ability to administer the test on an iPad would be better for the students and that the students are more familiar with using iPads.

The positioning of the testing device was a challenge for young students. For example, one test examiner struggled with trying to figure out what to do with the keyboard on the Chromebook, which was getting in the way of a kindergarten student’s “little arm,” and the student could not reach to touch the screen, although it was obvious the student wanted to touch the screen.

The ratings obtained for whether students had difficulties accessing the test are presented in table 43.

Table 43. Overall, Did Your Student Have Any Difficulties Accessing the Test? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Span** | **Yes N (%)** | **No N (%)** |
| K | 4 (31) | 9 (69) |
| Grade 1 | 0 (0) | 12 (100) |
| Grade 2 | 1 (8) | 11 (92) |
| Grade span 3–5 | 1 (10) | 9 (90) |
| Grade span 6–8 | 0 (0) | 12 (100) |
| Grade span 9–12 | 4 (33) | 8 (67) |

#### Test Interface Accessibility

Information was collected during the study on how the interface was accessible for students who independently used a computer or tablet. While the Alternate ELPAC is a one-on-one test administration, independent interaction was recommended if appropriate for the student, and at least one student was able to navigate the test independently. However, the majority of students were not able to independently take the assessment, although many students did interact with the test examiner leading them. Some students at early grade levels did demonstrate an inability to access the computer meaningfully.

When considering devices, the goal was usually to support independent interaction when possible. Few students could use a computer independently (15%). Most students could use a computer with assistance (49%). A large number of students did not yet use a computer (35%). Table 44 presents the students’ ability to interact with a computer. The data in table 44 was obtained from student BIQs, and missing data may impact the total numbers reflected for tables in this section.

Table 44. What Best Describes Your Student’s Computer Use?

|  |  |  |
| --- | --- | --- |
| **Computer Use** | **Frequency** | **Percent** |
| Can use a computer independently | 10 | 14.9 |
| Can use a computer with assistance | 33 | 49.3 |
| Does not yet use a computer | 24 | 35.8 |
| **Total** | 67 | 100 |

The data in table 45 reflects that more students were independent using a tablet than with a computer. Relatively few students did not yet use a tablet.

Table 45. What Best Describes Your Student’s Tablet Use?

|  |  |  |
| --- | --- | --- |
| **Student Tablet Use** | **Frequency** | **Percent** |
| Can use a tablet independently | 18 | 26.9 |
| Can use a table with assistance | 33 | 49.3 |
| Does not yet use a tablet | 16 | 23.9 |
| **Total** | 67 | 100 |

Test examiners were asked to indicate all the ways students needed access for technology use, refer to table 46. Some students can use a mouse (28%). Nearly all participants can use some kind of technology.

Table 46. What Best Describes Your Student’s Need for Access When Using Technology?

|  |  |  |
| --- | --- | --- |
| **Student Mode of Access Using Technology** | **Frequency** | **Percent** |
| Touchscreen | 39 | 60.9 |
| Uses a mouse | 18 | 28.1 |
| Uses a keyboard | 2 | 3.1 |
| Does not yet use technology | 5 | 7.8 |
| **Total** | 64 | 100 |

Information was collected during the study on internet connectivity during test administration. There were minor issues with connectivity for three test examiners. At least two test examiners experienced test shutdowns due to notification of a security threat and could not access Wi-Fi. One test examiner who had difficulty reported, “Generally we do not have issues with state testing.” The data on connectivity is provided in table 47 by numbers and percentages in parentheses.

Table 47. Any Issues with Internet Connectivity? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Status | K | Gr 1 | Gr 2 | Gr 3–5 | Gr 6–8 | Gr 9–12 |
| Yes | 1 (8) | 1 (8) | 0 (0) | 0 (0) | 0 (0) | 1 (8) |
| No | 12 (92) | 11 (92) | 12 (100) | 10 (100) | 12 (100) | 11 (92) |

**Note:** Test examiners were able to log back on, but it was distracting to students.

##### Test Delivery System Issues

Many test examiners were successful in using the test delivery system. One test examiner was excited and commented, “Doing the test on the computer is awesome.” However, there were issues with the test delivery system that occurred during test administration. The display of test content was a significant concern. Three different grade levels and grade spans reported issues where test items did not display all three answer choices on the screen. Despite all attempts to enlarge, magnify, or scroll, these items did not reflect the third answer choice. Some test examiners used the POD option for those items, others proceeded as though only two choices were offered. Test examiners commented that a horizontal layout would be preferable to the current vertical. First, a horizontal layout would render items to need less scrolling and would be organized in a left to right presentation that is consistent with reading.

Another issue occurred when a test examiner attempted to enter a “no response” for an item that was not reflected in the review before submission. The test examiner used the drop-down list to mark “no response” and the screen froze. The test examiner had to exit the test and log on again. Despite two attempts, the “no response” would not display in the system. This took valuable time and the student lost some attention during this event.

Test examiners provided feedback regarding the accessibility resources of the test delivery system for the students. Several test examiners noted that the [**NEXT**] button was too small for the students to use and that by enlarging the [**NEXT**] button the students could be more independent.

Test examiners also expressed concern that the images on the screen were not large enough, despite using zoom and magnification. At least two test examiners commented that the arrows for expanding of items and passages were confusing. The amount of scrolling needed during test administration was an area of concern. Test examiners noted difficulty with the amount of scrolling for themselves and that a mouse is needed for test administration.

#### Access for a Range of Communication Modes

The last area of investigation regarding accessibility was related to the test examiners’ perception of how effective the task types were across a wide range of communication modes. Test examiners provided ratings on the range of communication modes supported for accessing the test. The ratings associated with the items with low agreement were primarily related to access related to comments received on expressive items. For example, in kindergarten, test examiners commented the need for more pictures, icons, manipulatives, and objects to respond, particularly for expressive items. Test examiners noted that the expressive items were not accessible to students who were nonverbal and presymbolic.

The data on task type appropriateness across a wide range of communication modes is presented in table 48.

Table 48. Were the Task Types Appropriate for Students Across a Wide Range of Communication Modes? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Status** | **K** | **Gr 1** | **Gr 2** | **Gr 3–5** | **Gr 6–8** | **Gr 9–12** |
| Yes | 3 (25) | 5 (46) | 11 (92) | 7 (70) | 7 (58) | 9 (75) |
| No | 9 (75) | 6 (54) | 1 (8) | 3 (30) | 5 (42) | 3 (25) |

### Test Administration

The third area of investigation was to evaluate the clarity of test administration materials, *DFA*, and scoring rubrics to refine materials for future test administration. Test examiners were observed using the materials and interviewed about the test administration experiences using these materials. Test examiners’ use of the ancillary materials for the Alternate ELPAC administration, such as the *DFA* and scoring rubrics, were investigated. Specific research questions for this investigation were as follows:

* To what extent are the *DFA* and scoring rubrics clear for test examiners? What recommendations do test examiners have for the *DFA* and scoring rubrics?
* How do test examiners use the *DFA* during administration? What improvements could be made to the directions for clarity and ease of use?
* How do test examiners use the scoring rubrics? What improvements could be recommended?

The review of test administration materials is presented in the next section.

#### Test Material Clarity

This area of investigation sought to answer the research question about the extent to which the *DFA* and scoring rubrics were clear for test examiners and to determine if test examiners had recommendations for improving the clarity of the *DFA* and scoring rubrics. Test examiners provided overall ratings that the *DFAs* were clear and easy to use. Ratings ranged from 67 to 83 percent agreement across grade levels and grade spans. However, some test examiner comments included that the *DFA* was “wordy” and hard to read through quickly during test administration. One test examiner stated, “It was difficult navigating, and I kept getting lost between the *DFA* and the computer while I was trying to focus on the student and keep them engaged.” Test examiners recommended providing more space and visual cues in the *DFA*. Test examiners also recommended an opportunity to practice with the test materials and platform to role play a test administration session.

The ratings for *DFA* clarity are provided in table 49.

Table 49. Were the *Directions for Administration* Clear and Easy to Use? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Span** | **Yes** | **No** |
| TK/K | 9 (75) | 3 (25) |
| Grade 1 | 8 (67) | 4 (33) |
| Grade 2 | 8 (67) | 4 (33) |
| Grade span 3–5 | 7 (70) | 3 (30) |
| Grade span 6–8 | 9 (75) | 3 (25) |
| Grade span 9–12 | 10 (83) | 2 (17) |

Test examiners provided ratings on the ease of understanding of the scoring rubrics. Ratings of the clarity of the rubrics ranged from 67 to 100 percent agreement. Test examiners who felt the rubrics were difficult felt the rubric length was difficult to review during test administration. A few test examiners felt that there were too many rubrics. Test examiners expressed the need to have more time to become familiar with the rubrics. The findings of the clarity of the rubrics is presented in table 50.

Table 50. Were the Rubrics Easy to Understand? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Span** | **Yes** | **No** |
| TK/K | 8 (67) | 4 (33) |
| Grade 1 | 8 (67) | 4 (33) |
| Grade 2 | 11 (100) | 0 (0) |
| Grade span 3–5 | 9 (90) | 1 (10) |
| Grade span 6–8 | 9 (75) | 3 (25) |
| Grade span 9–12 | 9 (91) | 3 (9) |

#### *Directions for Administration*

Information was obtained about how test examiners used the *DFA* during test administration and provided an opportunity for making improvements. The *DFA* provided a lot of information for test examiners. Test examiners had a variety of experience in working in the test delivery system or had previous test administration experience. Test administration was difficult for some test examiners who had no experience with standardized assessment to kindergarten through grade two in general.

Some test examiners had never tested any students on the CAAs before this pilot administration and were unfamiliar with the testing platform. Test examiners commented that the training needed to navigate the system was not the same as the information the test examiners needed to know to test students. Some examples of the concern were test administration issues due to logistical and logon issues such as not having the student’s Statewide Student Identifier (SSID), resetting passwords, and updating browsers. Some test examiners were unfamiliar with the navigation through the system and the *DFA* did not contain a quick guide for troubleshooting technology issues. The research team often worked to resolve these issues. Other issues included browsers not being downloaded, computers that needed updating, test examiners not knowing needed passwords, and not having the correct SSID number. One test examiner stated that training on the test platform at a different time than the training for the Alternate ELPAC would have been helpful.

Test examiners requested more test administration practice and training. A test examiner recommended that a mock test administration with a peer during training would have been a big help and reduced anxiety over test administration preparation. Test examiners indicated a need for more time to read the *DFAs* and prepare materials for testing. Test examiners indicated a preference to have logged on and had access to *DFAs* and POD one month in advance. Some test examiners wanted more practice items to help become familiar with the tasks and rubrics and to prepare the students for taking the test.

#### Scoring Rubrics

The investigation of how test examiners used the scoring rubrics was part of the study. Test examiners were observed using the rubrics for scoring by the research team. The data collected was on issues related to scoring rubrics. The test examiners had no issues using the scoring rubrics across grade levels and grade spans from 60 to 91 percent of the time. Some test examiners commented on a need for more time to understand the rubrics and asked for reassurance from the research team regarding whether the items had been scored correctly. Rubrics at grade span three through five had the most issues with the scoring rubrics (40% of the time). One test examiner stated, “I need to understand when I can model. It’s not clear when to do it or not.” Researchers also noted that test examiners needed more clarity on when modeling was permissible. The percentages of agreement for issues with scoring are reflected in table 51 and are followed by correct score assignment in table 52. The test examiners did score the items correctly across grade levels and grade spans are 100 percent. However, the test examiners were uncertain about scoring correctly and often asked the research team to provide a level of assurance.

Table 51 presents the percentages of agreement with scoring.

Table 51. Test Examiner Self-Reported Issues with Scoring: Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |
| --- | --- | --- |
| **Grade level or Grade Span** | **Yes** | **No** |
| K | 3 (23) | 10 (77) |
| Grade 1 | 3 (25) | 9 (75) |
| Grade 2 | 1 (8) | 11 (92) |
| Grade span 3–5 | 4 (40) | 6 (60) |
| Grade span 6–8 | 2 (17) | 10 (83) |
| Grade span 9–12 | 1 (9) | 10 (91) |

Table 52 presents the correct score agreement.

Table 52. Did the Test Examiner Assign the Correct Score?

|  |  |  |
| --- | --- | --- |
| **Grade Level or Grade Span** | **Yes** | **No** |
| K | 12 (100) | 0 (0) |
| Grade 1 | 12 (100) | 0 (0) |
| Grade 2 | 12 (100) | 0 (0) |
| Grade span 3–5 | 10 (100) | 0 (0) |
| Grade span 6–8 | 12 (100) | 0 (0) |
| Grade span 9–12 | 12 (100) | 0 (0) |

Test examiners were asked to rate the amount of time needed for taking the test, ratings are included in table 53.

The general trend for the amount of time was about what the examiners expected; ratings ranged from 67 to 83 percent.

Table 53. Overall, How Would You Rate the Amount of Time Needed for Taking the Test, Not Including the Probing Questions? Frequencies and Percentages by Grade Level or Grade Span

|  |  |  |  |
| --- | --- | --- | --- |
| **Grade Level or Grade Span** | **About What I Expected** | **Shorter Than I Expected** | **Longer Than I Expected** |
| K | 9 (75) | 1 (8) | 2 (17) |
| Grade 1 | 10 (83) | 0 (0) | 2 (17) |
| Grade 2 | 8 (67) | 2 (17) | 2 (17) |
| Grade span 3–5 | 7 (70) | 2 (20) | 1 (10) |
| Grade span 6–8 | 9 (75) | 2 (17) | 1 (8) |
| Grade span 9–12 | 10 (83) | 2 (17) | 0 (0) |

## Limitations of the Study

The purpose of the Alternate ELPAC pilot cognitive study was to gather evidence about ELs with significant cognitive disabilities and test examiners’ interactions with task types and to investigate accessibility and use of test administration materials. There were 71 students who participated in the study, which is a relatively large number for a study of this type. Ultimately, the strength of using cognitive lab methodology to gather in-depth information is particularly relevant for low‑incidence disability populations that are highly individualized. While the study provided useful findings to inform development of the Alternate ELPAC operational field test, ETS cautions against overgeneralizing the findings from this study for other purposes.

This study used a purposeful sampling approach to stratify and capture major variations of students in a small and diverse population (Patton, 2015, p. 265). The study included as many home languages as possible but acknowledges not all home languages are represented in the study. The student BIQ was created to identify as many characteristics as possible to include the full range of students who would take the Alternate ELPAC. However, some survey questions from the student BIQ did not generate enough data to support inclusion for analyses. For example, test examiners did not consistently answer the following questions:

* Does the student receive English Language Development (ELD) Services? If yes, what type of ELD program?
* Who provides English language support?
* How does the student receive services to support English language development?

While the study exceeded the target sample in number of student recruitment, only three recently arrived ELs with significant cognitive disabilities participated, which means that the study does not provide direct evidence of the suitability of the test for new arrivals, those who are the intended population for the Initial Alternate ELPAC. No opportunities to observe students who were deaf or deaf-blind occurred, which will be important to consider for future investigations. This study was voluntary and required parental consent, which may have impacted the test examiners and students recruited for the study. It was particularly challenging to obtain parental permission for students in early grades.

The identification of students eligible for the Alternate ELPAC for the study was not readily available in the California Longitudinal Pupil Achievement Data System (CALPADS) and was a challenge for this study, particularly when identifying students in kindergarten through grade two. The CALPADs data system had missing data for several study participants, for example, primary disability and home language information. The researchers made attempts to use local information when possible.

## Recommendations

The learning characteristics of the assessed population have significant implications for an assessment’s validity (Pellegrino et al., 2001). Specifically, the validity evaluation of an assessment should consider two questions. First, it is necessary to know whether the assessment is appropriate for the intended population. This study provides evidence that answers this question. Secondly, it is important to ensure that the appropriate population is in fact the population being assessed (Pellegrino et al., 2001). This study found some challenges in identifying English learners with the most significant cognitive disabilities who are eligible for this test specifically. Consider gathering evidence to establish that only students who are eligible for the Alternate ELPAC are taking the assessment, including the validation processes used to assure only the target population is taking the assessment. Because the test is new to the field, it is important to determine if there are students who should be taking an alternate assessment who are not currently being identified as eligible, especially in early grades (Christensen et al., 2018). Equally important is the identification of students who may be over identified for an alternate assessment.

This study serves as part of the ECD approach to the development of the Alternate ELPAC by investigating and examining three principles presented in *A Brief Introduction to Evidence-Centered Design* (Mislevy, Almond & Lukas, 2003, p. 20), which are that

1. important knowledge in the domain of interest and an understanding of how that knowledge is acquired and put to use;
2. the chain of reasoning from what participants say and do in assessments to inferences about what those participants know, can do, or should do next, must be based on the principles of evidentiary reasoning; and
3. [the] purpose [of assessment] must be the driving force behind design decisions, which reflect constraints, resources, and conditions of use.

In particular, the *Preliminary Findings on the Development and Evaluation of Task Types for the Alternate English Language Proficiency Assessments for California* (2020), supported decisions and recommendations based on the study findings that impacted the test blueprint. These were presented to the California SBE in May 2020. The specific recommendations from that preliminary report are also included here as part of the final recommendations of the study.

It is worth noting that the pilot study, while led by ETS research staff, was conducted with the active involvement and partnership of ETS assessment development staff who are responsible for designing task types and developing test items and test forms, both for the pilot study and for the subsequent operational field test. Those assessment development staff participated in the identification of the following recommendations in consultation with the wider research team and have also taken action to implement these recommendations in the development of test items and test forms for the operational field test. Similarly, forthcoming revisions to guidance on accessibility resources, directions for administration, test familiarization materials, and training and practice procedures for test examiners (on administration and scoring of rubric-based items) will address many of the following recommendations.

The findings of this study are appropriate to support decisions on test design and assessment development in the areas of task design and development, accessibility resources, and administration. This study adds to the literature in the emerging field of alternate English proficiency assessments. This study provides an opportunity to add findings to an emerging field where studies on this target population are lacking. This study will provide valuable information about the characteristics of ELs with significant cognitive disabilities and improve the assessment of this important and vulnerable population.

### Recommendations on Interactions with Task Types

The first area of investigation examined student and test examiner interactions during the pilot administration of the Alternate ELPAC to determine whether the task types elicit the intended knowledge and skills without construct-irrelevant interference. Preliminary recommendations were provided to the CDE in a report titled *Preliminary Findings on the Development and Evaluation of Task Types for the Alternate English Language Proficiency Assessments for California* (2020). Thatreport described the following elements of the process:

* How the task types were developed on the basis of an analysis of the ELD Connectors.
* How the task types were reviewed by a range of stakeholders, including California educators with appropriate expertise.
* How the task types were evaluated during the Alternate ELPAC pilot, using cognitive lab methodology, leading to the appropriateness of the task types included in the proposed test blueprint to be submitted to the California SBE for approval.

The report included a set of recommendations to the CDE as the basis to adjust the *Alternate ELPAC Test Blueprints* presented to, and approved by, the California SBE in May 2020. The preliminary findings on task types in that report were obtained from data from the test examiner ratings provided during the test administration, which are embedded in the *Directions for Administration (DFAs),* and research observational data. The complete findings for the study are found in [section 7](#_6._Findings). The recommendations from that report are presented along with this set of recommendations, include any actions that have occurred, and are documented in the following section.

#### Recommendations and Actions

The set of recommendations for the first area of investigation of interactions with task types are presented in this section.

##### Blueprint Considerations

The inclusion of *Recognize and Use Common Words* as an initial task type across all grade levels and grade spans would create a more gradual progression for students and provide more information about the student's ability to take the assessment. This is an important consideration for providing access to students who are performing at low levels of linguistic complexity and developing language.

**Recommendation:** Adjust the draft test blueprint by adding *Recognize and Use Common Words* as a new first task type. To keep test length as proposed, reduce the number of *Communicate About Familiar Topics* task types on the blueprint from two sets to one set.

**Action:** This recommendation has been implemented. The revised test blueprint, including the *Recognize and Use Common Words* task type and a reduced number of *Communicate About Familiar Topics* sets, was submitted to the SBE and approved in May 2020.

##### Item Type Format Considerations

Students had difficulty understanding the directions for the multiple-selection multiple-choice item type within the *Understand and Express an Opinion* task type. This item type provided students with three options and asked them to select the two that were correct.

**Recommendation:** Remove the multiple-selection multiple-choice item type and replace it with an item in the more familiar single-selection multiple-choice format.

**Action:** This recommendation has been implemented. The items revised from multiple-selection multiple-choice to single-selection multiple-choice are aligned to the ELD Connectors as noted in the test blueprint. The change involved replacing a 2-point item with a 1-point item, the overall number of items on the blueprint remained the same, while the number of score points have decreased by one. The number of items was not increased because sufficient score points remain to support intended score reporting, and there is a compelling interest from the field not to lengthen the test.

##### Task Type Considerations

In the *Interact with an Informational Text* task type, test examiners had a mixed reaction to an item type designed to assess connectors aligned with Part III: Using Foundational Literacy Skills, which was planned for use at kindergarten, grade one, grade two, and grade span three through five. Some test examiners found that this item type assessed the foundational literacy skills in a manner similar to how such skills were being taught in the classroom; however, other test examiners reported that the approach in this item type did not correspond to classroom practice. Of particular concern was the fact that some kindergarten and grade one students in the Alternate ELPAC population may not have had an opportunity to learn the skills assessed by this item type.

**Recommendation:** For kindergarten and grade one, replace the *Interact with an Informational Text* item aligned with the Part III connector with an item aligned with another connector assessed in the same task type, such as Part I.B.5–Part I.B.6, Part II.A.1, or Part II.A.2. For grade two and grade span three through five, maintain the item aligned to the Part III connector, but look for possible adaptations to the item that will make it correspond to as broad a range of instructional styles as possible.

**Action:** This recommendation has been implemented. The item aligned to Part III in grade two and grade span three through five has been revised based on feedback received from the pilot study, reviewed and approved by the CDE, and reviewed and approved by California educators at the June 2020 IRM.

This change involves (for the grade levels and grade spans involved) replacing a 2-point item with a 1-point item; as a result, the overall number of items on the blueprint will remain the same, while the number of score points will decrease by one. As with Action 2, the number of items was not increased because sufficient score points remain to support intended score reporting, and there is a compelling interest from the field not to lengthen the test. (After the implementation of both Action 2 and Action 3, the score points for each test form will be 28–32, as opposed to a range of 30–34 score points for the previous draft of the test blueprint.)

##### Task Type Adjustments for Expressive Items

**Recommendation:** Consider some minor adjustments to the task types, including addressing alternate response options for expressive items to be more inclusive for students who are nonverbal and students who use AAC devices.

**Action:** Optional Individualization has been added to rubric-scored expressive items that either allow for the use of real objects or provide picture card responses that could be used in isolation, with a static level AAC device, or can be imaged into a dynamic level AAC device to provide the student with access to expressive options in the communication mode with which the student would be most familiar and comfortable. This allows access for nonverbal students, students who use low- and high-tech AAC and picture exchange to access expressive items scored with a rubric. Directions regarding Optional Individualization and the picture cards themselves have been included in the *DFA*.

##### Optional Individualization

**Recommendation:** Consider including guidance for increasing the number of objects and manipulatives for students who are presymbolic communicators.

**Action:** Optional Individualization has been added to stories, passages, and items in which real objects or manipulatives could be provided to the student. Directions regarding Optional Individualization have been included in the *DFA*. A list of additional materials, including any objects that can be used on the assessment, has been included in the *DFA*.

##### Minimize Construct Irrelevance

Findings of construct-irrelevant factors included complexity of language in questions, ambiguity in pictorial options, and passage length.

**Recommendation:** Consider revisiting the task type templates to make sure task descriptions are kept simple and clear.

**Action:** Revisions to the task type specifications have been scheduled and will reflect this consideration.

**Recommendation:** Consider supporting item writers by providing specific examples of construct-irrelevant factors to support future item writer training.

**Action:** The February 2021 Item Writer Workshop will be designed to address this consideration. Examples will be shared with educators at all future Item Writer Workshops as well.

**Recommendation:** Recruit the expertise of special educators based on certifications and experience with students on alternate assessment. Consider recruiting the expertise of specialists in the areas of vision, speech, language, and assistive technology for item reviews, in addition to ELD specialists and special educators.

**Action:** The criteria for educator recruitment for item writing and item review committees will reflect this consideration.

**Recommendation:** Consider shortening passages, reducing text complexity, and reviewing topic appropriateness.

**Action:** Text complexity guidelines have been revised to address this consideration. Items developed for the Operational Field Test reflect the consideration as well; for example, total word counts for stories and passages have been reduced where appropriate; any multicategorical options have been replaced with options of a single category; character names in stories and passages have been simplified; and culled text from stories and passages has been included with the test question.

**Recommendation:** Consider adding visuals and developing a guide for item developers; for example, visuals should only contain information essential for understanding and responding to the item. Use pictures and objects that reflect those used by students with the most significant cognitive disabilities. Refrain from adding visuals to make items more interesting. Review visuals for ambiguity and accessibility.

**Action:** Visuals have been added to items developed for the Operational Field Test, as appropriate. Visuals have been revised to minimize or remove any nonessential information. Guidelines for visuals will be added to the task type specifications.

##### Face Validity

Face validity means the test should have certain characteristics that reflect what informed stakeholders expect about the test. A test that does not have such evidence may encounter resistance from the field, and the governing body may bear criticism. Test examiners specifically shared an expectation that some tasks would require students to produce writing. The Alternate ELPAC test design uses the combinations of receptive (listening and reading) and expressive (speaking and writing) skills. Educators within the community of ELD specialists may expect the Alternate ELPAC to have four traditional domains (Listening, Speaking, Reading, and Writing).

**Recommendations:** Identify ways to communicate effectively with stakeholders how the Alternate ELPAC test design combines the listening, speaking, reading, and writing skills into the domains of receptive and expressive.

* Consider supporting the field with specific examples of receptive and expressive skills using a variety of communication modes for test administration training.
* Consider reviewing items for traditional domain skills representation.

##### Future Study Recommendations

Conduct future studies to support evidence for validity of the Alternate ELPAC.

**Recommendations:** Consider future studies of ELs with the most significant cognitive disabilities including investigations of opportunity to learn (OTL).Studies of OTL include considerations on the use of the ELD connectors and how students receive instruction on those connectors. Other topics of investigation to support validity are presented here.

* Consider conducting studies of OTL that makes explicit the use of the ELD standards via the connectors for the target population including how students will receive instruction or support for learning English. It is important to make explicit the opportunity to learn, as it relates to the range of service delivery options and the provision of English language development services for Alternate ELPAC.
* Consider future analyses and investigations of ELs with recent arrival status.
* Conduct studies to investigate how young ELs in kindergarten through grade two are assigned to alternate assessments.
* Consider future studies or processes for validation of the target population.
* Consider future investigations of the relationship between the three levels of ELP and the results from the CAA content assessment.
* Consider future studies to support the consequential basis of validity, how scores will be used, how will the assessment will be monitored and reported, including the intended and unintended consequences of the assessment.

### Recommendations for Accessibility

The set of recommendations for the second area of investigation of accessibility are presented in this section.

#### Accessibility Resources

The Alternate ELPAC identifies the universal tools, designated supports and accommodations in the Accessibility Resources manual to support student access for the test.

##### Test Examiner Guidance and Accessibility Resources

Test examiners were observed to need assistance to use the accessibility resources for the Alternate ELPAC.

**Recommendations:** The following possible actions are recommended to support test examiners and the field in better understanding how to support access for students taking Alternate ELPAC.

* Consider multiple ways to support individual decision making for test examiners in the selection and use of the accessibility resources. For example, provide examples of universal tools, designated supports and accommodations that are likely to be encountered on the assessment.
* Consider ways to assure the documentation of the provision, monitoring, and use of accessibility resources.
* Consider providing resources or guidance for test examiners and students who use braille.
* Consider training for optimal test settings, for example, student’s familiarity with test settings, noise, and test security.
* Consider providing guidance on the definition of separate setting for test administration; including examples of separate setting for the Alternate ELPAC.

##### **Accessibility: Presentation of Test Content**

Test examiners had to be supported to use the online universal tools of zoom and expanding of items and passages to provide access when presenting test content to students.

**Recommendations:** Access to test content was enhanced for the students when test examiners knew how to use the universal tools.

* Consider providing multiple opportunities for students and test examiners to practice the use of the following universal tools: expandable items, passages, and zoom features.
* Consider test administration training for the test delivery system separately.

#### Accessibility Resources and Standardization

Test administration procedures were not standardized for some test administrations for specific accessibility supports and accommodations.

**Recommendations:** Test examiners need explicit guidance in the provision of accessibility resources that impact a standardized test administration.

* Consider multiple ways to train test examiners to support a standardized test administration. Provide materials to support the correct usage of accessibility resources during test administration training. Specifically provide examples of allowed and nonallowed uses of accessibility resources. For example, materials, such as a video to model the correct usage of POD and a variety of AAC uses.
* Consider providing specific examples of allowed and nonallowed test administration practices in test administration training. Include clarification and specific guidance around the use and misuse of POD and alternate response options to support and ensure standardized test administration while supporting the individual needs of students.
* Consider providing training, guidance, and modeling for the use of alternate response options. Include support for test examiners who support students who use assistive technology, both low and high technology, and who use fixed, dynamic, or hybrid devices.

#### Test Delivery Accessibility

The study found two test accessibility concerns related to the test delivery platform.

Recommendations: Consider improving the accessibility of the test delivery system.

* Consider the horizontal layout of items so that all answer choices are on one page.

**Action:** This consideration has been implemented.

* Consider enlarging the [NEXT] button to promote more student independence.

### Recommendations for Test Administration

The following set of recommendations are from the third area of investigation of test administration.

#### Test Examiner Training

Test examiners expressed the need for support and more time to become familiar with the test materials.

Recommendations: Provide test examiners with training to provide effective test administration.

* Consider providing test administration training in multiple ways (e.g., online, face-to-face, video) including a mock test administration.
* Consider a walk-through of test materials and resources prior to test administration.
* Consider the timely delivery of the *DFA* and other test materials to test examiners prior to administration to allow ample time to review and practice using the training materials.

#### Practice Items

Test examiners wanted more practice items to be available earlier so that the students would be more familiar with the tasks.

##### Practice Test Items

**Recommendations:** Consider practice opportunities for students and test examiners prior to test administration.

* Consider providing practice tests as early as possible, so that the students and test examiners may practice using the test interface.

**Action:** The Alternate ELPAC training tests (which are relatively short and focused on interactions with the test delivery system) will be released in September 2020, four months prior to the launch of the operational field test. Practice tests (which follow the test blueprint and therefore provide a realistic example of what students will experience on an operational test) will be released in November 2020, two months prior to the launch of the operational field test.

* Consider providing examples of task types for test examiners to practice instructionally with students.

**Action:** A practice test will be developed for each grade level and grade span. Each practice test will follow the operational test blueprint. (That is, each practice test will contain the same number and distribution of items found on an operational form of items, including a full set of each of the seven task types.)

* Consider providing a “sandbox” to enable students and test examiners to practice a session.

#### Scoring Rubrics

Test examiners expressed uncertainty about scoring and using the scoring rubrics and requested clarity on how and when modeling is permissible.

**Recommendations:**

* Consider including practice with the scoring rubrics and demonstrating the modeling rubric flow chart as part of the test administration training and materials.
* Consider materials, such as a short video or resource guide, to demonstrate both correct and incorrect provision of modeling support.

**Action:** ETS will collaborate with SCOE to develop videos or vignettes (or both) of test examiners administering and scoring items with a rubric, including the modeling rubric, and students responding to such items for training purposes.

* Consider developing a video for training on the scoring rubrics.

**Action:** ETS will collaborate with SCOE to develop videos or vignettes (or both) of test examiners administering and scoring items with a rubric and students responding to such items for training purposes. Additionally, the scoring rubrics in the *DFA* have been simplified to help ensure reliable and accurate in-the-moment scoring of expressive items with a rubric. For example, all rubrics are now 3-point rubrics; the 4-point rubric presented in the study was revised.

* Consider providing sample items for scoring practice prior to test administration.

**Action:** Both the training tests and practice tests will include expressive items with a rubric for test examiners to review and practice.

## References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, *71(2)*, 287–311.

Blair, J., & Conrad, F. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly, 75(4),* 636–59.

California Department of Education. 2014. *California English Language Development Standards: Kindergarten through grade 12.* Adopted by the State Board of Education November 2012. Sacramento, CA: California Department of Education. Retrieved from <http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf>.

CAST (2018). *Universal design for learning guidelines version 2.2*. Retrieved from <http://udlguidelines.cast.org/binaries/content/assets/udlguidelines/udlg-v2-2/udlg_graphicorganizer_v2-2_numbers-yes.pdf>

Charmaz, K. (2002). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & J. S. Holstein (Eds.). *Handbook of interview research: Context and method.* Thousand Oaks, CA: SAGE Publications Inc.

Cho, Y., & So, Y. (2014). Construct-irrelevant factors influencing young English as a foreign language (EFL) learners’ perceptions of test task difficulty (Research Memorandum No. RM-14-04). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/research/pdf/RM-14-04.pdf>

Christensen, L. L., Gholson, M. L., & Shyyan, V. V. (2018, April). Establishing a definition of English learners with significant cognitive disabilities (ALTELLA Brief No. 1). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research, Alternate English Language Learning Assessment project: [https://altella.wceruw.org//pubs/ALTELLA\_Brief-01\_Definition\_070218.pdf](https://altella.wceruw.org/pubs/ALTELLA_Brief-01_Definition_070218.pdf)

Educational Testing Service. (2019a). *Alternate ELPAC task type specifications* (Unpublished manuscript). Sacramento, CA: Educational Testing Service.

Educational Testing Service. (2019b). *California English Language Development Connectors for the Alternate English Language Proficiency Assessments for California* (Unpublished manuscript). Sacramento, CA: Educational Testing Service.

Educational Testing Service. (2019c). *English Language Proficiency Assessments for California accessibility resources for operational testing*. Unpublished document. Sacramento, CA: California Department of Education.

Educational Testing Service. (2019d). *Proposed high-level test design for the Alternate English Language Proficiency Assessments for California.* Sacramento, CA: California Department of Education. Retrieved from <https://www.cde.ca.gov/ta/tg/ep/documents/proposedhltdaltelpac.pdf>.

Educational Testing Service. (2019e). *Working with minors procedural document.* Unpublished internal document.

Educational Testing Service. (2020a). *Preliminary findings on the development and evaluation of task types for the Alternate English Language Proficiency Assessments for California* (Unpublished manuscript). Sacramento, CA: Educational Testing Service.

Educational Testing Service. (2020b). *Alternate English Language Proficiency Assessments for California proposed test blueprint* (Prepared for submission to the California State Board of Education. Sacramento, CA: Educational Testing Service.

Every Student Succeeds Act of 2015, Pub. L. No. 114‐95 § 114 Stat. 1177 (2015). Retrieved from <https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>

Fuchs, L., & Fuchs, D. (1989). Effects of examiner familiarity on Black, Caucasian, and Hispanic children: A meta-analysis, *Exceptional Children,* 55, 303–08.

Karvonen, M. (2019). *Cognitive lab methods for alternate assessments.* Presented at the Assessing Special Education Students (ASES) State Collaborative on Assessment and Student Standards (SCASS) meeting, February 21 in Los Angeles, CA.

Karvonen, M. & Clark, A.K. (2019). Students with the most significant cognitive disabilities who are also English learners. *Research and Practice for Persons with Severe Disabilities*, 44(2), 71-86. Sage Publication, CA. <https://doi.org/10.1177%2F1540796919835169>

Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practice*, 27, 3–16.

Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product and process.* New York, NY: Wiley.

Leighton, J. P. (2017). Using think-aloud interviews and cognitive labs in educational research. Oxford, UK: Oxford University Press.

Marion, S. F., & Pellegrino, J. W. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, d*25* (4), 47. <https://doi.org/10.1111/j.1745-3992.2006.00078.x>

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: SAGE Publications Inc.

Miller, K., Willson, S., Chepp, V., & Ryan, J. M. (2014). Chapter 4 Analysis. In K. Miller, S. Willson, V. Chepp, & J. L. Padilla (Eds.), *Cognitive interviewing methodology* (p. 35­50). John Wiley & Sons. doi: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118838860.ch4>

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design (Research Report 03-16). Princeton, NJ: Educational Testing Service.

Nash, B., Clark, A. K., & Karvonen, M. (2016). *First contact: A census report on the characteristics of students eligible to take alternate assessments* (Technical Report No. 16-01). Lawrence: University of Kansas, Center forEducational Testing and Evaluation. Retrieved from <https://dynamiclearningmaps.org/sites/default/files/documents/publication/First_Contact_Census_2016.pdf>

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development, 50*(4), 217–23.

Saldaña, J. (2014). Coding and analysis strategies. In P. Leavy (Ed.), Oxford handbook of qualitative research. doi:10.1093/oxfordhb/9780199811755.013.001

Steinberg, J., Brenneman, M., Castellano, K., Lin, P., & Miller, S. (2014). A Comparison of Achievement Gaps and Test-Taker Characteristics on Computer-Delivered and Paper-Delivered *Praxis I*® Tests. (ETS Research Report No. RR-14-35). Princeton, NJ: Educational Testing Service.

Thurlow, M. L., Wu, Y., Quenemoen, R. F., & Towles, E. (2016). *Characteristics of students with significant cognitive disabilities* (NCSC Brief No. 8). Minneapolis: University of Minnesota, National Center and State Collaborative. Retrieved from <http://www.ncscpartners.org/Media/Default/PDFs/Resources/NCSCBrief8.pdf>

U.S. Department of Education. (2018). *A state’s guide to the U.S. Department of Education’s assessment peer review process*. Retrieved from <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Beverly Hills, CA: Sage.

Winke, P., Lee, S., Ahn, J. I., Choi, I., Cui, Y., & Yoon, H. (2018). The cognitive validity of child English language tests: What young language learners and their native-speaking peers can reveal. *TESOL Quarterly, 52*(2), 274–303. <https://doi.org/10.1002/tesq.396>

Winter, P. C., Karvonen, M., & Christensen, L. L. (2018, August). *Developing item templates for alternate assessments of English language proficiency.* Madison, WI: University of Wisconsin–Madison, Alternate English Language Learning Assessment (ALTELLA). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research: [https://altella.wceruw.org//pubs/ALTELLA\_Report\_Item‌\_083108.pdf](https://altella.wceruw.org/pubs/ALTELLA_Report_Item_083108.pdf)

Winter, P. C., Kopriva, R. J., Chen, C-S., & Emick, J. E. (2006). Exploring individual and item factors that affect assessment validity for diverse learners: Results from a large-scale cognitive lab. *Learning and Individual Differences, 16*(4), 267–76. <https://doi.org/10.1016/j.lindif.2007.01.001>

Zieky, M. (2014). An introduction to the use of evidence-centered design in test development. *Psicologia Educativa. (20)*2, 79–87. <https://doi.org/10.1016/j.pse.2014.11.003>

## Appendices

### Appendix A

Table A1. Alternate ELPAC Task Types Included in Pilot

|  |  |
| --- | --- |
| Task Type Name | Description |
| *Recognize and Use Common Words* | The student observes one to three photos of common nouns or objects. (If a student has a visual impairment, the test examiner can use accessibility resources or give the student real objects or manipulatives.) The student responds to one receptive item and one expressive item about the objects. |
| *Communicate About Familiar Topics* | The student is presented with a brief story of one to three sentences about a familiar topic. A photo or illustration is included for context and support. The student responds to one receptive item and one expressive item about the familiar topic. |
| *Understand a School Activity* | The student is presented with a story related to a familiar classroom- or school-based activity, event, or procedure. The student responds to two receptive and two expressive items about the stimulus. |
| *Understand a School Exchange* | The student is presented with a story related to a school or classroom activity with at least two characters. The student responds to one receptive item and three expressive items about the stimulus. |
| *Describe a Routine* | The student is presented with a story related to a familiar school or classroom routine. The student responds to two receptive items and one expressive item about the stimulus. |
| *Understand and Express an Opinion* | The student is presented with a story that describes a situation in which two or more characters make a choice and give one or more reasons or facts for the choice. The student responds to three receptive items and one expressive item about the stimulus. |
| *Interact with a Literary Text* | The student is presented with a literary story on a familiar topic. The student responds to one receptive item and three expressive items about the stimulus. |
| *Interact with an Informational Text* | The student is presented with an informational passage on a familiar or unfamiliar topic. The student responds to two receptive and two expressive items about the stimulus. |

### Appendix B

**Note:** The following abbreviations are used in these tables:

K = kindergarten

N/A indicates that no standard deviation calculation was provided for less than two students.

Table B1. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **K N** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Low | 5 | 2.00 | 1.73 | 11.11 | 5 | 8.20 | 5.36 | 45.56 | 1 | 11.00 | N/A | 64.71 |
| Medium | 5 | 8.00 | 4.64 | 44.44 | 5 | 12.00 | 2.55 | 66.67 | 5 | 9.00 | 2.24 | 52.94 |
| High | 1 | 9.00 | N/A | 50.00 | 1 | 14.00 | N/A | 77.78 | 5 | 12.20 | 6.42 | 71.76 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 10.00 | N/A | 55.56 | 1 | 5.00 | N/A | 29.41 |
| Missing | 1 | 4.00 | N/A | 22.22 | 0 | N/A | N/A | N/A | 0 | N/A | N/A | N/A |
| **Total** | 12 | 5.25 | 4.31 | 29.17 | 12 | 10.42 | 4.17 | 57.87 | 12 | 10.17 | 4.67 | 59.80 |

Table B2. Total Score by Grade Level or Grade Span and Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **Gr 1N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–‍5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Low | 1 | 6.00 | N/A | 30.00 | 5 | 6.80 | 5.89 | 34.00 | 0 | N/A | N/A | N/A |
| Medium | 4 | 4.25 | 3.10 | 21.25 | 2 | 14.00 | 5.66 | 70.00 | 7 | 12.43 | 6.02 | 62.14 |
| High | 7 | 9.71 | 4.61 | 48.57 | 1 | 12.00 | N/A | 60.00 | 2 | 13.00 | 0.00 | 65.00 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 9.00 | N/A | 45.00 | 3 | 8.00 | 7.00 | 40.00 |
| Missing | 0 | N/A | N/A | N/A | 1 | 6.00 | N/A | 30.00 | 0 | N/A | N/A | N/A |
| **Total** | 12 | 7.58 | 4.62 | 37.92 | 10 | 8.90 | 5.40 | 44.50 | 12 | 11.42 | 5.74 | 57.08 |

Table B3. Receptive Scores by Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Low | 5 | 1.20 | 0.45 | 17.14 | 5 | 2.60 | 1.67 | 37.14 | 1 | 5.00 | N/A | 62.50 |
| Medium | 5 | 3.60 | 1.14 | 51.43 | 5 | 4.40 | 0.89 | 62.86 | 5 | 4.00 | 1.73 | 50.00 |
| High | 1 | 5.00 | N/A | 71.43 | 1 | 5.00 | N/A | 71.43 | 5 | 5.80 | 2.68 | 72.50 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 4.00 | N/A | 57.14 | 1 | 2.00 | N/A | 25.00 |
| Missing | 1 | 3.00 | N/A | 42.86 | 0 | N/A | N/A | N/A | 0 | N/A | N/A | N/A |
| **Total** | 12 | 2.67 | 1.56 | 38.10 | 12 | 3.67 | 1.50 | 52.38 | 12 | 4.67 | 2.27 | 58.33 |

Table B4. Receptive Scores by Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **Gr 1N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Low | 1 | 4.00 | N/A | 66.67 | 5 | 2.40 | 1.82 | 34.29 | 0 | N/A | N/A | N/A |
| Medium | 4 | 2.50 | 2.65 | 41.67 | 2 | 4.50 | 0.71 | 64.29 | 7 | 4.71 | 1.60 | 67.35 |
| High | 7 | 2.57 | 1.40 | 42.86 | 1 | 3.00 | N/A | 42.86 | 2 | 5.50 | 0.71 | 78.57 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 4.00 | N/A | 57.14 | 3 | 2.00 | 1.73 | 28.57 |
| Missing | 0 | N/A | N/A | N/A | 1 | 3.00 | N/A | 42.86 | 0 | N/A | N/A | N/A |
| **Total** | 12 | 2.67 | 1.78 | 44.44 | 10 | 3.10 | 1.52 | 44.29 | 12 | 4.17 | 1.95 | 59.52 |

Table B5. Expressive Scores by Ratings of English Language Proficiency (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Low | 5 | 0.80 | 1.30 | 8.00 | 5 | 5.60 | 4.22 | 50.91 | 1 | 5.00 | N/A | 50.00 |
| Medium | 5 | 4.40 | 3.85 | 44.00 | 5 | 7.60 | 2.30 | 69.09 | 5 | 5.00 | 1.22 | 50.00 |
| High | 1 | 4.00 | 4.00 | 40.00 | 1 | 9.00 | N/A | 81.82 | 5 | 6.40 | 3.85 | 64.00 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 6.00 | N/A | 54.55 | 1 | 3.00 | N/A | 30.00 |
| Missing | 1 | 1.00 | N/A | 10.00 | 0 | N/A | N/A | N/A | 0 | N/A | N/A | N/A |
| **Total** | 12 | 2.58 | 3.06 | 25.83 | 12 | 6.75 | 3.14 | 61.36 | 12 | 5.50 | 2.65 | 55.00 |

Table B6. Expressive Scores by Ratings of English Language Proficiency (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proficiency** | **Gr 1N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Low | 1 | 2.00 | N/A | 14.29 | 5 | 4.40 | 4.10 | 33.85 | 0 | N/A | N/A | N/A |
| Medium | 4 | 1.75 | 1.50 | 12.50 | 2 | 9.50 | 4.95 | 73.08 | 7 | 7.57 | 4.58 | 58.24 |
| High | 7 | 7.14 | 4.06 | 51.02 | 1 | 9.00 | N/A | 69.23 | 2 | 7.50 | 0.71 | 57.69 |
| Not sure | 0 | N/A | N/A | N/A | 1 | 5.00 | N/A | 38.46 | 3 | 6.00 | 5.29 | 46.15 |
| Missing | 0 | N/A | N/A | N/A | 1 | 3.00 | N/A | 23.08 | 0 | N/A | N/A | N/A |
| **Total** | 12 | 4.92 | 4.14 | 35.12 | 10 | 5.80 | 4.05 | 44.62 | 12 | 7.17 | 4.13 | 55.13 |

Table B7. Total Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Nonverbal | 8 | 2.88 | 1.81 | 15.97 | 2 | 4.00 | 4.24 | 22.22 | 3 | 5.33 | 4.51 | 31.37 |
| Verbal | 4 | 10.00 | 3.92 | 55.56 | 10 | 11.70 | 2.87 | 65.00 | 9 | 11.78 | 3.63 | 69.28 |
| **Total** | 12 | 5.25 | 4.31 | 29.17 | 12 | 10.42 | 4.17 | 57.87 | 12 | 10.17 | 4.67 | 59.80 |

Table B8. Total Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **Gr 1N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Nonverbal | 7 | 5.43 | 3.05 | 27.14 | 6 | 5.50 | 3.15 | 27.50 | 5 | 7.40 | 4.83 | 37.00 |
| Verbal | 5 | 10.60 | 5.03 | 53.00 | 4 | 14.00 | 3.65 | 70.00 | 7 | 14.29 | 4.68 | 71.43 |
| **Total** | 12 | 7.58 | 4.62 | 37.92 | 10 | 8.90 | 5.40 | 44.50 | 12 | 11.42 | 5.74 | 57.08 |

Table B9. Receptive Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Nonverbal | 8 | 2.00 | 1.20 | 28.57 | 2 | 2.00 | 2.83 | 28.57 | 3 | 2.33 | 1.53 | 29.17 |
| Verbal | 4 | 4.00 | 1.41 | 57.14 | 10 | 4.00 | 1.05 | 57.14 | 9 | 5.44 | 1.94 | 68.06 |
| **Total** | 12 | 2.67 | 1.56 | 38.10 | 12 | 3.67 | 1.50 | 52.38 | 12 | 4.67 | 2.27 | 58.33 |

Table B10. Receptive Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **Gr 1 N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Nonverbal | 7 | 2.86 | 2.04 | 47.62 | 6 | 2.33 | 1.37 | 33.33 | 5 | 3.00 | 2.35 | 42.86 |
| Verbal | 5 | 2.40 | 1.52 | 40.00 | 4 | 4.25 | 0.96 | 60.71 | 7 | 5.00 | 1.15 | 71.43 |
| **Total** | 12 | 2.67 | 1.78 | 44.44 | 10 | 3.10 | 1.52 | 44.29 | 12 | 4.17 | 1.95 | 59.52 |

Table B11. Expressive Scores Compared for Verbal and Nonverbal Students (Group 1 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **KN** | **K Mean** | **K SD** | **K % Pts** | **Gr 2 N** | **Gr 2 Mean** | **Gr 2 SD** | **Gr 2 % Pts** | **Gr 6–‍8 N** | **Gr 6–8 Mean** | **Gr 6–8 SD** | **Gr 6–8 % Pts** |
| Nonverbal | 8 | 0.88 | 0.99 | 8.75 | 2 | 2.00 | 1.41 | 18.18 | 3 | 3.00 | 3.00 | 30.00 |
| Verbal | 4 | 6.00 | 2.94 | 60.00 | 10 | 7.70 | 2.41 | 70.00 | 9 | 6.33 | 2.06 | 63.33 |
| **Total** | 12 | 2.58 | 3.06 | 25.83 | 12 | 6.75 | 3.14 | 61.36 | 12 | 5.50 | 2.65 | 55.00 |

Table B12. Expressive Scores Compared for Verbal and Nonverbal Students (Group 2 Grade Levels and Grade Spans)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mode of Communication** | **Gr 1N** | **Gr 1 Mean** | **Gr 1 SD** | **Gr 1 % Pts** | **Gr 3–‍5 N** | **Gr 3–5 Mean** | **Gr 3–5 SD** | **Gr 3–5 % Pts** | **Gr 9–‍12 N** | **Gr 9–‍12 Mean** | **Gr 9–‍12 SD** | **Gr 9–‍12 % Pts** |
| Nonverbal | 7 | 2.57 | 2.64 | 18.37 | 6 | 3.17 | 1.83 | 24.36 | 5 | 4.20 | 2.28 | 32.31 |
| Verbal | 5 | 8.20 | 3.70 | 58.57 | 4 | 9.75 | 2.99 | 75.00 | 7 | 9.29 | 3.90 | 71.43 |
| **Total** | 12 | 4.92 | 4.14 | 35.12 | 10 | 5.80 | 4.05 | 44.62 | 12 | 7.17 | 4.13 | 55.13 |

### Appendix C

#### Knowledge Skills and Abilities by ELP Ratings

Table C1. Frequencies and Percentages of Students’ Listening Skills by English Language Proficiency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Skills** | **Low N (%)** | **Medium N (%)** | **High N (%)** | **Not Sure N (%)** |
| Follows 2 step directions | 2 (12) | 8 (29) | 10 (59) | 1 (17) |
| Follows 1 step directions | 6 (35) | 10 (36) | 5 (29) | 3 (50) |
| Attends and responds to simple commands | 3 (18) | 4 (14) | 1 (6) | 1 (17) |
| Performs a simple action upon request | 0 (0) | 3 (11) | 1 (6) | 0 (0) |
| Indicates a choice when offered an array of items | 0 (0) | 2 (7) | 0 (0) | 0 (0) |
| Points to or touches objects upon request | 4 (24) | 1 (4) | 0 (0) | 1 (17) |
| Does not attend to sound | 2 (12) | 0 (0) | 0 (0) | 0 (0) |

Table C2. Frequencies and Percentages of Students’ Reading Skills by English Language Proficiency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Skills** | **Low N (%)** | **Medium N (%)** | **High N (%)** | **Not Sure N (%)** |
| Reads text without any picture or symbol support with comprehension | 0 (0) | 0 (0) | 3 (18) | 0 (0) |
| Reads text without picture or symbol support but without comprehension | 1 (6) | 2 (7) | 1 (6) | 0 (0) |
| Identifies individual words without picture support | 1 (6) | 2 (7) | 0 (0) | 0 (0) |
| Reads words, phrases, or sentences when pictures or symbols are provided with unfamiliar words | 1 (6) | 6 (21) | 3 (18) | 0 (0) |
| Recognizes letter sounds | 1 (6) | 4 (14) | 3 (18) | 0 (0) |
| Recognizes letters | 2 (12) | 6 (21) | 3 (18) | 3 (50) |
| Matches objects to pictures | 5 (29) | 5 (18) | 2 (12) | 3 (50) |
| Identifies and names objects | 2 (12) | 1 (4) | 2 (12) | 0 (0) |
| Does not yet have an understanding of print or text | 4 (24) | 2 (7) | 0 (0) | 0 (0) |

Table C3. Frequencies and Percentages of Students’ Speaking Skills by English Language Proficiency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Skills** | **Low N (%)** | **Medium N (%)** | **High N (%)** | **Not Sure N (%)** |
| Verbally speaks 3 or more words in complete sentences using grammatical rules | 0 (0) | 2 (7) | 6 (35) | 0 (0) |
| Verbally speaks 2 or more words in sentences or phrases without consistently following grammatical rules | 1 (6) | 8 (29) | 7 (41) | 1 (17) |
| Verbally speaks 2-word phrases | 2 (12) | 9 (32) | 1 (6) | 1 (17) |
| Verbally speaks 1 word at a time | 3 (18) | 1 (4) | 0 (0) | 0 (0) |
| Uses touch and gestures by pointing and head nodding | 5 (29) | 2 (7) | 0 (0) | 1 (17) |
| Uses vocalizations, gestures, and facial expressions to communicate intentionally | 4 (24) | 5 (18) | 2 (12) | 2 (33) |
| Uses eye gaze with intentionality | 0 (0) | 1 (4) | 0 (0) | 0 (0) |
| Uses AAC | 0 (0) | 0 (0) | 1 (6) | 1 (17) |
| Not intentionally communicative | 2 (12) | 0 (0) | 0 (0) | 0 (0) |

Table C4. Frequencies and Percentages of Students’ Writing Skills by English Language Proficiency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Skills** | **Low N (%)** | **Medium N (%)** | **High N (%)** | **Not Sure N (%)** |
| Writes 1–3 sentences (spelling not always correct) | 0 (0) | 0 (0) | 1 (6) | 0 (0) |
| Writes a simple sentence or phrase (spelling not correct) | 0 (0) | 1 (4) | 1 (6) | 0 (0) |
| Writes words, spelling not always correct | 1 (6) | 3 (11) | 3 (18) | 0 (0) |
| Writes using word banks | 1 (6) | 1 (4) | 1 (6) | 0 (0) |
| Selects letters or symbols to express meaning | 0 (0) | 1 (4) | 0 (0) | 0 (0) |
| Copies letters and words but does not produce independent writing | 3 (18) | 8 (29) | 5 (29) | 2 (33) |
| Randomly selects letters or symbols when asked to write | 1 (6) | 1 (4) | 1 (6) | 0 (0) |
| Makes random marks or scribbles | 7 (41) | 9 (32) | 3 (18) | 2 (33) |
| Is not yet able to demonstrate | 4 (24) | 4 (14) | 1 (6) | 2 (33) |