

# English Language Proficiency Assessments for California Usability Pilot: A Final Report

**Contract #CN140284**

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

**Presented September 19, 2019**

****

**Educational Testing Service**

**660 Rosedale Road**

**Princeton, NJ 08541**

Table of Contents

[Executive Summary 1](#_Toc19791566)

[1. Background 4](#_Toc19791567)

[2. Study Purposes 5](#_Toc19791568)

[3. Areas of Investigation and Research Questions 6](#_Toc19791569)

[3.1. Students’ Interaction with the Platform and Computer-Based ELPAC Items 6](#_Toc19791570)

[3.2. Accessibility Resources at the Time of the Pilot 7](#_Toc19791571)

[3.3. Usability of Test Examiner Materials and Scoring Guidelines 7](#_Toc19791572)

[4. Development of Pilot Test Materials 8](#_Toc19791573)

[4.1. Conversion of Items 8](#_Toc19791574)

[4.2. Development of Pilot Test Forms 13](#_Toc19791575)

[4.3. Accessibility Resources 13](#_Toc19791576)

[4.4. Development of Test Examiner Materials 14](#_Toc19791577)

[5. Methodology 15](#_Toc19791578)

[5.1. Sampling and Recruitment 15](#_Toc19791579)

[5.2. Participants 16](#_Toc19791580)

[5.3. Study Instruments 18](#_Toc19791581)

[5.3.1. Student Background Questionnaire 18](#_Toc19791582)

[5.3.2. Cognitive Interview Protocol Including a Rating Form 18](#_Toc19791583)

[5.3.3. Observation Protocol Including a Rating Form 19](#_Toc19791584)

[5.3.4. Semi-structured Focus Group Interview Protocol 19](#_Toc19791585)

[5.3.5. Test Examiner Survey 19](#_Toc19791586)

[5.3.6. Tutorial Videos 19](#_Toc19791587)

[5.4. Procedure 20](#_Toc19791588)

[5.5. Data Analysis 21](#_Toc19791589)

[6. Findings 22](#_Toc19791590)

[6.1. Students’ Interaction with the Platform and Computer-Based ELPAC Items 22](#_Toc19791591)

[6.1.1. Understanding Directions 22](#_Toc19791592)

[6.1.2. Navigating the Platform and Items Independently 23](#_Toc19791593)

[6.1.3. Need for Technological Assistance 25](#_Toc19791594)

[6.1.4. Usability of the Audio Delivery Model 26](#_Toc19791595)

[6.1.5. Usability of Speaking Recording Functions 27](#_Toc19791596)

[6.1.6. Typing on the Computer to Respond to Writing Items 28](#_Toc19791597)

[6.1.7. Usability of the Writing Tools 29](#_Toc19791598)

[6.2. Accessibility Resources 29](#_Toc19791599)

[6.2.1. Universal Tool Use 30](#_Toc19791600)

[6.2.2. Universal Tools’ Perceived Helpfulness 31](#_Toc19791601)

[6.2.3. Other Universal Tools Needed 31](#_Toc19791602)

[6.3. Usability of Test Examiner Materials and Scoring Guidelines 32](#_Toc19791603)

[6.3.1. Clarity of the *DFAs* 32](#_Toc19791604)

[6.3.2. Scoring the Speaking Domain Responses 34](#_Toc19791605)

[7. Limitations of the Study 37](#_Toc19791606)

[8. Recommendations 38](#_Toc19791607)

[8.1 Recommendations on Usability by Students 38](#_Toc19791608)

[Recommendation 1 38](#_Toc19791609)

[Recommendation 2 39](#_Toc19791610)

[Recommendation 3 40](#_Toc19791611)

[Recommendation 4 40](#_Toc19791612)

[8.2 Recommendations on Accessibility Resources 41](#_Toc19791613)

[Recommendation 5 41](#_Toc19791614)

[Recommendation 6 41](#_Toc19791615)

[8.3 Recommendations on Test Examiner Materials and Support 42](#_Toc19791616)

[Recommendation 7 42](#_Toc19791617)

[Recommendation 8 43](#_Toc19791618)

[Recommendation 9 43](#_Toc19791619)

[References 45](#_Toc19791620)

List of Tables

[Table 1. Total Number of Items by Domain and Grade or Grade Span 13](#_Toc19791621)

[Table 2. Number of Participating Students by Grade Level and Grade-Span 17](#_Toc19791622)

[Table 3. EL Students with Disabilities by Grade Level or Grade-Span and Primary Disability Type 18](#_Toc19791623)

[Table 4. Navigating Independently in Listening: Frequency and Percentage of Students Observed in Each Rating Category 24](#_Toc19791624)

[Table 5. Navigating Independently in Reading: Frequency and Percentage of Students Observed in Each Rating Category 24](#_Toc19791625)

[Table 6. Navigating Independently in Writing: Frequency and Percentage of Students Observed in Each Rating Category 24](#_Toc19791626)

[Table 7. Using the Scroll Bar Independently: Frequency and Percentage of Students Observed in Each Rating Category 25](#_Toc19791627)

[Table 8. Need of Technological Assistance: Frequency and Percentage of Students in Each Rating Category 26](#_Toc19791628)

[Table 9. Using Audio Buttons Independently in Listening: Frequency and Percentage of Students in Each Rating Category 27](#_Toc19791629)

[Table 10. Listening to a Stimulus More than Once: Frequency and Percentage of Students in Each Rating Category 27](#_Toc19791630)

[Table 11. Hesitance to Record Speaking Responses: Frequency and Percentage of Students Observed in Each Rating Category 28](#_Toc19791631)

[Table 12. Easiness of Typing on the Computer: Frequency and Percentage of Students in Each Rating Category 29](#_Toc19791632)

[Table 13. Number of EL Students Who Used Each Universal Tool by Grade or Grade Span 30](#_Toc19791633)

[Table 14. Clarity of Directions in the *DFAs* for Test Examiners: Frequency and Percentage of Student Administrations Observed in Each Rating Category 34](#_Toc19791634)

[Table 15. Test Examiner Survey Response: Ease of Use with the *DFAs* 34](#_Toc19791635)

[Table 16. Ease of Scoring Speaking Tasks: Frequency and Percentage of Student Administration Observations Across Each Rating Category 35](#_Toc19791636)

## Executive Summary

The English Language Proficiency Assessments for California (ELPAC) is the required state test for English language proficiency (ELP) that is given to students whose primary language is other than English. The ELPAC is aligned with California’s 2012 English Language Development Standards and assesses students’ ELP skills in Listening, Speaking, Reading, and Writing. With the approval of the State Board of Education, the California Department of Education (CDE) began planning the transition of the current paper-based ELPAC to a computer-based assessment with the goal of enhancing the state’s assessment system. The state has been implementing the statewide content-area assessments (English language arts/‌literacy, mathematics, and science) on the computer since 2015. Transitioning the paper-based ELPAC to the computer-delivery format is anticipated to (a) create consistency with other California assessments by utilizing the same online test delivery platform and (b) increase the range of available accessibility resources.

As part of the transition work, Educational Testing Service (ETS), in collaboration with the CDE and the Sacramento County Office of Education (SCOE), conducted a small-scale usability pilot employing cognitive laboratory methodology (henceforth called “the study”) on the ELPAC task types in an online environment. The study was conducted at the earliest stage of the development of the computer-based ELPAC prior to the large-scale conversion of paper-based ELPAC items to a computer-based format. Based on the study findings, a set of practical recommendations was made to inform the areas of improvements for the future field test and operational administration.

The overall goal of the study was to ensure that the ELPAC paper-based items can be appropriately converted to a computer-delivered format and that the Initial and Summative computer-based ELPAC are being developed based on empirical evidence about the usability of various computer-delivered features. Collecting such evidence is expected to support the validity of the computer-based ELPAC in terms of assessing the intended construct (i.e., students’ ability to demonstrate their English language knowledge and skills without interference from their computer skills).

This study investigated three major areas for usability:

1. Students’ interaction with the platform and computer-based ELPAC items
2. Students’ use of the available accessibility resources
3. Usability of test examiner materials (e.g., *Directions for Administration*) and scoring guidelines

The study utilized the cognitive lab methodology to closely examine test users’ interactions with computer-based features. Each pilot test administration was a one-on-one session between one test examiner and one student, and each session was observed by a researcher. As a small-scale usability testing study, a deliberate sampling plan was used to recruit a relatively small number of students. Specifically, recruitment efforts were made to ensure that students who had little experience in computer-based assessment (e.g., transitional kindergarten through grade two students and recently arrived ELs) were included in the study. A small number of non-EL students who would be able to perform their grade-appropriate tasks were also included in the sampling criteria to allow researchers to identify any EL-specific difficulties in interacting with the computer-based assessment features. Participating students represented diverse background characteristics in terms of their grade level, ELP level, home language, recently arrived EL status, computer familiarity, and disability status. A total of 19 test examiners and 107 (89 English learner [EL] and 18 non-EL) students from transitional kindergarten to grade eight across six schools from two districts participated in the study. Of the 89 EL students, 13 were EL students with disabilities. The data sources included students’ background information, observers’ ratings on the usability of various features based on the study protocols, observation notes, focus-group interview notes, and survey responses from test examiners. The notes and test examiners’ comments were qualitatively analyzed to find recurring themes regarding the usability of the materials and areas for improvements. Descriptive statistics for the ratings and survey responses were also computed and summarized.

Regarding the first area of investigation, observation notes indicated that, in general, participating students were highly engaged in the computer-based ELPAC items during the cognitive lab sessions. Some usability difficulties were also observed for students. The most notable pattern was that students tended to have more difficulty in understanding directions on the first item; however, students’ understanding and familiarity increased as the test progressed within each language domain. This tendency was also observed for some non-EL students. These results suggest that although students learned quickly during the testing sessions, students’ familiarity and practice with the new test format is an important consideration for the computer-based ELPAC. Of particular interest, younger students in kindergarten to grade one participated with the assistance of their test examiner and demonstrated the ability to attend to the computer-delivered stimuli. An additional topic of interest pertaining to the Listening test was the number of times a student listened to an audio stimulus. The results indicate that most students listened to Listening stimuli once. Yet there was one student who played a stimuli excessively, raising some concern about score interpretations from the Listening test.

The second area of investigation focuses on students’ use of accessibility resources, including the universal tools (e.g., highlighter, line reader, zoom), accompanying the computer-based ELPAC. Observation notes indicated that usage varied as students were able to self-select tools built into the test platform. The ability to request assistance with the test navigation, scratch paper, and expandable passages were the most frequently used resources and were utilized by the largest number of students. ELs with disabilities also used the available resources, both the universal tools and a number of resources that were not part of the delivery system (e.g., sensory toys) to help them focus on the test administration. Test examiners also tried out the concentration assistance tools that enlarge and change the color of the mouse cursor to aid in the administration of the assessment, particularly for those administering the test to the younger students.

The third area of investigation focused on the usability of the test examiner materials and scoring procedures. The *Directions for Administration* (*DFAs*) and the Data Entry Interface (DEI) were used for each grade-level and grade-span administration. Overall, the data from the observations, surveys, and focus groups indicated that the clarity and organization of both the *DFAs* and the DEI for administering and scoring the Speaking domain should be improved. Additional findings related to the administration components resulted in recommendations to improve seating arrangements and technology placement, as well as additional directions to supplement the test administration and simultaneous management of the technology.

Based on the findings, the following recommendations are made to guide test developers in converting the paper-based ELPAC to the computer-delivery format appropriately in preparing for the field test and for future operational administration of the computer-based ELPAC. The recommendations are also intended to enhance the usability of the platform, computer-based ELPAC items, and their administration materials for test users. The recommendations include the following:

1. Improve test familiarity materials (tutorials, training tests, practice tests) to ensure students are prepared to take the computer-based ELPAC and test examiners are prepared to administer it
2. Create resource materials for educators and test examiners to help determine if students are ready to take the computer-based ELPAC under typical conditions
3. Allow students to listen only once to audio stimuli on the Listening test
4. Maintain recorded audio files for Listening stimuli on the kindergarten (K) and grade one Listening tests, similar to the grades two through eight Listening tests
5. Increase opportunity for familiarity and practice of accessibility resources for both test examiners and students
6. Provide appropriate supports to ensure students’ level of familiarity with technology does not impede their ability to take the computer-based ELPAC
7. Simplify the Speaking administration to make the administration of the test and scoring easier for the test examiner
8. Improve the organization of the *DFAs*
9. Enhance training on administration and scoring for test examiners

Detailed results and proposed action items for each recommendation are provided in this report.

## Background

The ELPAC is the required state test for ELP that is given to students whose primary language is a language other than English. State and federal laws require that local educational agencies (LEAs) administer a state test of ELP to eligible students in kindergarten through grade twelve. The ELPAC is aligned with California’s 2012 English Language Development Standards and assesses students’ ELP skills in Listening, Speaking, Reading, and Writing.

In November 2018, the State Board of Education approved Amendment 6 of the California Assessment of Student Performance and Progress (CAASPP) contract, which involves transitioning the current paper-based ELPAC to a computer-based assessment. In May 2019, the SBE approved the recommended plan for transitioning the ELPAC to a computer-based assessment in the *Proposed High-level Test Design for the Transition to Computer-Based ELPAC* (ETS, 2019). As part of the transition work, ETS, in collaboration with the CDE and SCOE, conducted a small-scale usability pilot employing cognitive lab methodology (heretofore called “the study”) on the ELPAC task types in an online environment. The study was conducted at the earliest stage of the development of the computer-based ELPAC prior to the large-scale conversion of paper-based ELPAC items to a computer-based format. This document reports the purposes, method, and findings of the study. In addition, this document includes a set of recommendations in preparing for the field test and future operational administration of the computer-based ELPAC, based on the study findings.

## Study Purposes

The overall goal of the ELPAC computer-based assessment study was to ensure that the ELPAC paper-based items can be appropriately converted to a computer-delivered format and that the Initial and Summative computer-based ELPAC are being developed based on empirical evidence about the usability of various computer-delivered features. Note that the content and psychometric quality of the paper-based ELPAC items have already been examined throughout the paper-based ELPAC development process. Thus, the study focused on the test users’ interaction with ELPAC task types presented on the computer and other computer-delivered features rather than the psychometric properties of the ELPAC items. The need to understand the usability of potential computer-based ELPAC task types and features warrants a close examination of the test-taking processes of students in the targeted population. The Usability Pilot study was therefore conducted utilizing the cognitive lab method to closely observe students’ test-taking processes as well as test examiners’ administration processes. This methodological approach allowed for collecting important evidence to support the validity of the computer-based ELPAC in terms of assessing the intended construct (e.g., students’ ability to demonstrate their English language knowledge and skills without interference from their computer skills).

The specific goals of the study were the following:

* To investigate the deliverability and functionality of ELPAC task types on the computer and prepare for the large-scale field test of the computer-based ELPAC items
* To examine the usability of test examiner guidelines, directions, and local scoring for the computer-based ELPAC to refine the materials prior to the field test and to support the validity of the ELPAC uses
* To examine students’ response processes to ensure that the computer-based ELPAC elicits the intended knowledge and skills without construct-irrelevant interference from sources related to technology use
* To examine the usability of the universal tools that ELs and EL students with disabilities may need when taking the computer-based ELPAC. (Note that designated supports and accommodations were in development at the time of the study; however, if ELs with disabilities were able to participate in the ELPAC with universal tools, those students were invited to participate in the study.)

Based on the study findings, the ETS team, in discussion with key members from the CDE and SCOE, aimed to develop recommendations for the ongoing computer-based ELPAC development process, particularly for refining the *Specifications for Conversion of ELPAC Task Types for Computer-Based Delivery* (ETS, 2019; to be reviewed separately by the CDE). It is anticipated that the study findings would also be used to fulfill some of the critical elements required in the U.S. Department of Education’s guidance on the peer-review process. In the recent peer-review guidance document (U.S. Department of Education, 2018), usability and cognitive lab studies are recommended to collect critical validity evidence for technology-based assessments.

## Areas of Investigation and Research Questions

This study entailed three major areas of investigation. The first area was students’ interaction with the platform and computer-based assessment items. This area of investigation focused on how well items render on the platform and how students navigate and complete items presented on the computer. For example, it is important to examine how students independently navigate computer-based assessment features, such as maneuvering the scroll bar and identifying where to select an answer, advancing to the next item, playing audio stimuli, and recording Speaking responses to complete items delivered on the computer. Students’ test-taking processes were closely examined to find any potential construct-irrelevant factors stemming from the computer-delivered format. Special attention was also devoted to recently-arrived students and students in kindergarten through grade two who might be unfamiliar with computer use.

The second area of investigation was students’ use of embedded universal tools. As universal tools were available to all test takers, the examination of the tools was conducted for both ELs and ELs with disabilities. Specifically, ELs with disabilities were invited to participate in the study in accordance with guidance from their individualized education program or Section 504 plan. Designated supports or accommodations were not included in the current study but were planned to be included in a study, the “Accessibility Pilot and Cognitive Labs,” during the field test window in fall 2019. The cognitive lab study that will complement the Accessibility Pilot in fall 2019 will focus on students with visual impairments or who are deaf or hard of hearing and need designated supports or accommodations to access the computer-based ELPAC.

The third area of investigation was the usability of test examiner materials, including *DFAs*—previously referred to as *Examiner’s Manuals* in the paper-based ELPAC—and scoring guidelines modified for the computer-based ELPAC. In this study, a sample of test examiners was closely observed as the test examiners used the materials. The test examiners were interviewed about their experiences administering the computer-based ELPAC using these materials.

The following three subsections describe the specific research questions for each of the areas of investigation.

### Students’ Interaction with the Platform and Computer-Based ELPAC Items

1. To what extent are the directions clear, allowing students to understand what they are asked to do? What difficulties, if any, do students at different levels of English proficiency have in understanding the directions?
2. To what extent can students navigate the features of the computer-based ELPAC items and tasks independently? What difficulties do students have in navigating items on the computer? How are these difficulties associated with students’ background characteristics (e.g., ELP levels, grade levels, recently arrived EL status, disability categories, computer familiarity)?
3. How consistently do students respond appropriately to items with audio stimuli under the present audio delivery model? What difficulties, if any, do students have with the audio stimuli?
4. To what extent do students use the Speaking recording functions (i.e., recording and playing spoken responses) independently without interventions from the test examiner? What difficulties, if any, do students have in recording their responses?
5. For grades administered Writing online, to what extent do students have difficulties typing their responses into Writing items? How do they use the writing tool? To what extent are students’ difficulties, if any, associated with computer keyboarding?

### Accessibility Resources at the Time of the Pilot

1. How do the students use embedded universal tools? What types of difficulties do students experience in accessing and interacting with the embedded universal tools?
2. To what extent do universal tools meet the needs of students with disabilities while they interact with items?
3. What other accessibility resources do ELs with disabilities report needing access to in order to take the ELPAC on the computer?

### Usability **of Test Examiner Materials and Scoring Guidelines**

1. To what extent are the directions in the *DFA*s and scoring guides clear for test examiners? What part of the directions are not clear?
2. How do test examiners use the *DFA* during the computer-based ELPAC administration? What types of difficulties do the test examiners have, if any?
3. How do test examiners use the local scoring materials during the computer-based ELPAC administration? What types of difficulties do the test examiners have, if any?

## Development of Pilot Test Materials

The present study utilized the *ELPAC Practice Tests* (CDE, 2018) as pilot test materials since they were nonsecure items and because they included representative items from all ELPAC task types. All items from the *ELPAC Practice Tests* previously underwent rigorous reviews through the Content Review Panel (CRP) and Bias and Sensitivity Review Panel (BSRP) meetings. After those meetings, members of the CDE approved the recommendations of the CRP and BSRP and verified that items were updated accurately.

The following subsections describe the procedure of converting the study test materials, including the test forms and test examiner materials based on the *ELPAC Practice Tests*.

### Conversion of Items

In preparation for the transition to the computer-based ELPAC, ETS developed the *Specifications for Conversion of ELPAC Task Types for Computer-Based Delivery* document (2019). Based on this document, ELPAC practice items were converted to computer format in collaboration with the American Institutes for Research.

All Listening and Reading items were multiple-choice (MC) items and all Speaking and Writing items were constructed-response (CR) items. As one of the computer-based assessment features, the directions, the Listening domain stimuli, and some Writing domain prompts are delivered via audio files. In these cases, an audio play bar with the [**Play/Pause**] toggle button is shown on the computer screen. Figure 1 and Figure 2 present a sample screen shot comparing a Listening item in the paper-based ELPAC to the computer-based version with the audio bar. At the time of the study, listening to stimuli multiple times for the Listening domain was allowed. For the operational administration, students can listen to the prompt only once, but can listen to the question and response choices multiple times.

Figure 1. A screenshot of a sample Listening item from the paper-based ELPAC

*See the* [*Alternate Text for Figure 1*](#AltText_Figure1) *for a description of this image.*

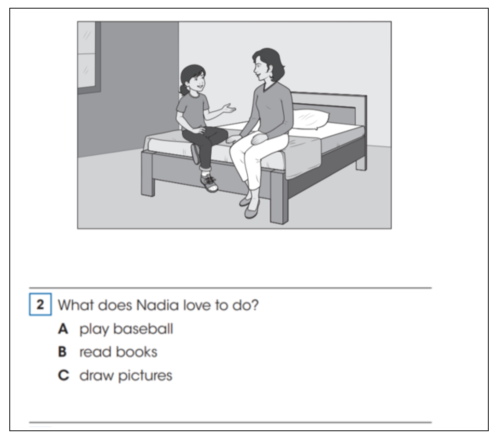
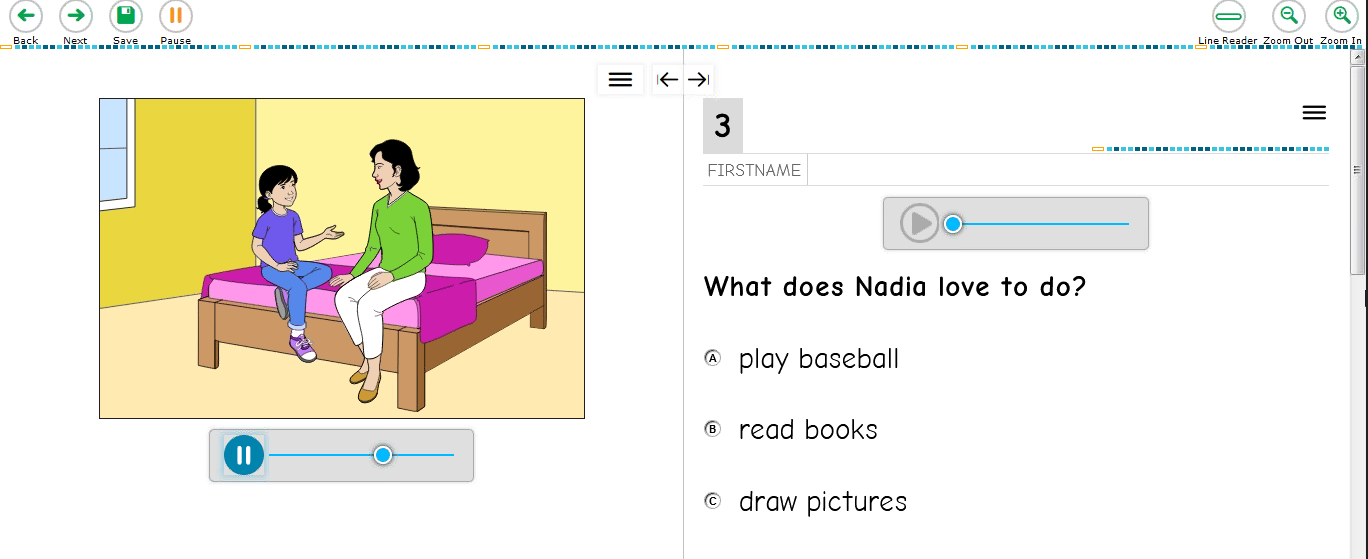


Figure 2. A screenshot of a sample Listening item from the computer-based ELPAC

*See the* [*Alternate Text for Figure 2*](#AltText_Figure2) *for a description of this image.*



Another notable difference from the paper-based ELPAC items compared to the computer-based assessment is that students’ responses for Speaking items can be recorded in case a second scoring other than the real-time, local scoring is needed. The [**Microphone**] and [**Play**] buttons are included in the Speaking domain, as shown in comparison with the paper-based test in Figure 3 and Figure 4. Considering these types of computer-based features, the administration model for kindergarten to grade one was to have the test examiner select the buttons during the one-on-one administration for the Listening, Reading, and Speaking domains. In the Speaking domain for kindergarten to grade one, test examiners also read aloud directions and prompts in person instead of using audio files. Test examiners’ scripts are included in the test examiners’ materials. For the operational administration, grade two will be included in the one-on-one administration model for Listening, Reading, Writing, and Speaking domains.

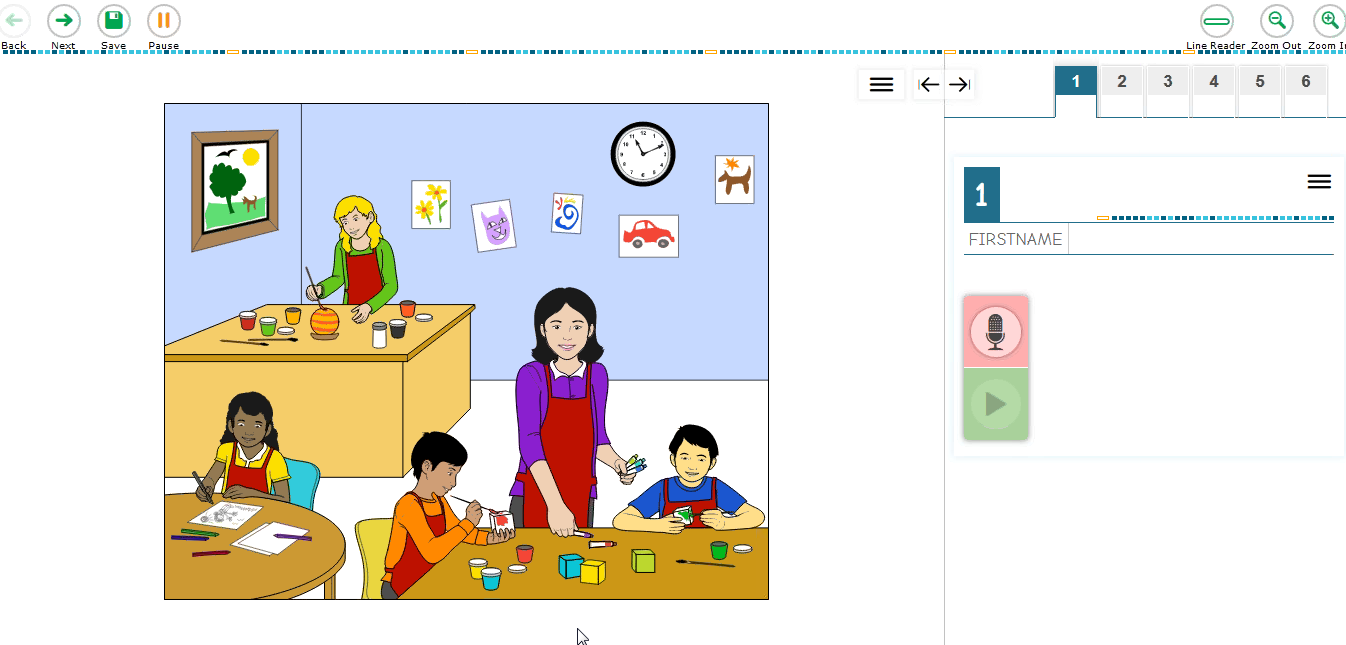
Figure 3. A screenshot of a sample Speaking item from the paper-based ELPAC

*See the* [*Alternate Text for Figure 3*](#AltText_Figure3) *for a description of this image.*



**Figure 4. A screenshot of a sample Speaking item from the computer-based ELPAC**

*See the* [*Alternate Text for Figure 4*](#AltText_Figure4) *for a description of this image.*



The Writing domain is administered on the computer starting at grade three. A writing tool bar is presented including some formatting features (e.g., font style and indentation) and, in some instances, an audio bar is present, so that the prompt can be read to the student. Figure 5 and Figure 6 compare a screenshot of a sample Writing item, from the original paper-based version, shown in Figure 5, to the new computer-based version, shown in Figure 6.

Figure 5. A screenshot of a sample Writing item from the paper-based ELPAC

*See the* [*Alternate Text for Figure 5*](#AltText_Figure5) *for a description of this image.*

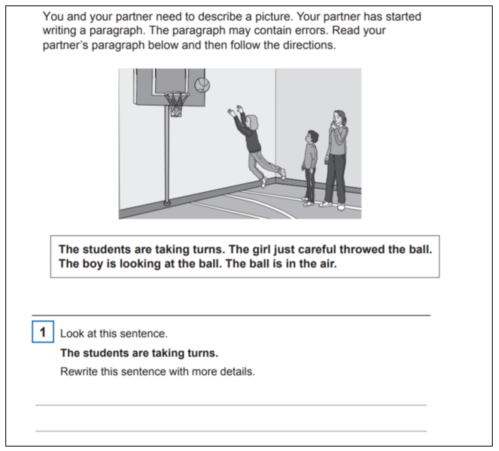
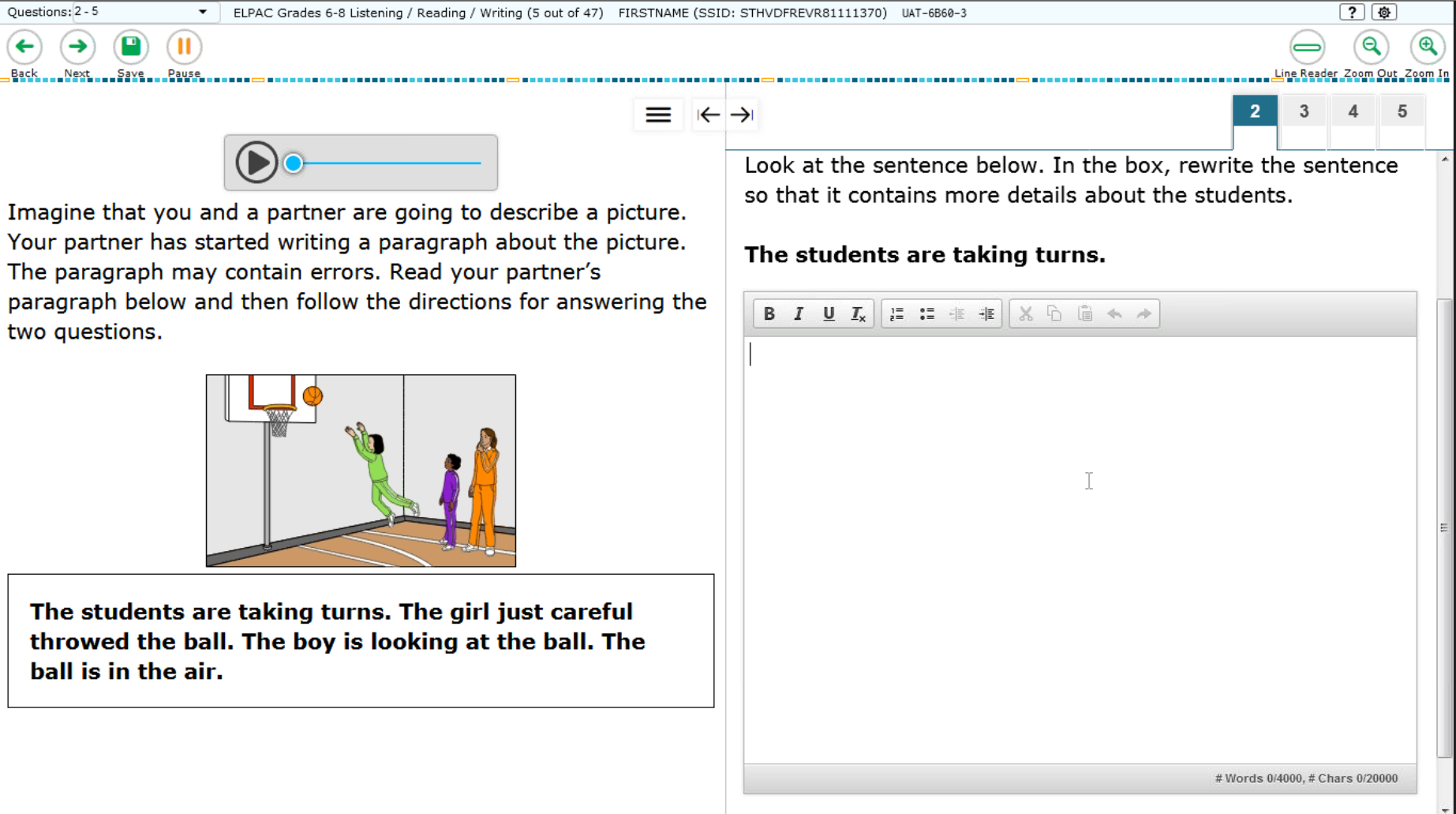


Figure 6. A screenshot of a sample Writing item from the computer-based ELPAC

*See the* [*Alternate Text for Figure 6*](#AltText_Figure6) *for a description of this image.*



Assessment specialists who were familiar with the ELPAC task types reviewed the converted items in the American Institutes for Research previewer within the Item Banking Information System to ensure that they appeared on the computer platform as described in the specifications. After assessment specialists had reviewed and approved the items, members of the CDE reviewed and approved the items with reference to the *Specifications for Conversion of ELPAC Task Types for Computer-Based Delivery* (ETS, 2019)document.

### Development of Pilot Test Forms

One test form of Listening, Speaking, Reading, and Writing items was assembled for each grade or grade span. To use items efficiently, some items were shared across adjacent grades or grade spans. Each grade or grade span had items from each of the task types administered in the Initial and Summative ELPAC. Table 1 shows the total number of items in each test form.

Table 1. Total Number of Items by Domain and Grade or Grade Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Domain (Item Type) | TK/K | Grade 1 | Grade 2 | Grade span  3–5 | Grade span  6–8 | Total |
| Listening (MC) | 7 | 7 | 7 | 11 | 12 | 44 |
| Speaking (CR) | 9 | 9 | 10 | 10 | 11 | 49 |
| Reading (MC) | 9 | 10 | 10 | 20 | 21 | 70 |
| Writing (CR) | 0 | 0 | 0 | 8 | 8 | 16 |
| Total | 25 | 26 | 27 | 49 | 52 | 179 |

Since the Writing test maintained the paper-based format for kindergarten to grade two, no Writing items were administered at these grades during the pilot.

### Accessibility Resources

For the present study, universal tools specified in the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing* (CDE, 2019) were made available and embedded in the computer-based ELPAC platform. Universal tools are those that are allowed for all students, regardless of EL or disability status, without needing teacher support or an individualized education program or Section 504 plan to gain access. Examples of allowed universal tools include accessibility resources such as the highlighter and digital notepad. Students’ use of those tools for the computer-based ELPAC was examined as one of the major areas of investigation for the study.

Among the existing universal tools included in Matrix One: California CAASPP Accessibility Resources (the table of allowable accessibility resources across the CAASPP assessments; CDE, 2018), those that may interact with the measurement of students’ ELP skills (e.g., an English glossary support) or those that are not appropriate for the construct (e.g., a calculator or the periodic table) were excluded in the present study. The specific universal tools and guidance for the use of those tools are also detailed in a separate document, the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing* (CDE, 2019), which includes the full range of accessibility resources and the review process of both Matrix One and the ELPAC paper and pencil test Matrix Four. At the time of the present study, the *English Language Proficiency Assessment for California Accessibility Resources for Operational Testing* document was under development. The findings from the study were anticipated to provide useful information for the refinement of the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing*.

As mentioned earlier, designated supports and accommodations were in development at the time of the study and will be examined at a later date (i.e., as part of the ELPAC Accessibility Pilot and Cognitive Labs conducted during the field test administration window).

### Development of Test Examiner Materials

A set of documents called *DFA*s was developed to guide test examiners in the administration of the pilot items and the local scoring of CR items. The content of the *DFAs* were limited to test-administration and scoring procedures that were particular to a given grade or grade span. To preserve security, directions and confidential item information were, to the greatest extent possible, displayed on the computer rather than appearing in the paper-based *DFAs*.

With respect to scoring, the *DFAs* contained instructions for scoring CR items (e.g., Speaking and Writing items) that were expected to be scored locally during some operational ELPAC administrations. The operational model for the computer-based ELPAC calls for Speaking items to be scored locally for both the Initial ELPAC and the Summative ELPAC in real-time. For Writing items, local scoring by test examiners is implemented for the Initial ELPAC and central scoring by a scoring vendor is planned for the Summative ELPAC. For the present study, the goal of providing scoring instructions was to examine the usability of the *DFAs* and gain feedback about the clarity of the directions and procedures for test examiners’ scoring of Speaking items.

Assessment specialists developed the *DFAs* as PDF documents that could be shared with test administrators online. *DFAs* went through multiple internal ETS reviews, as well as a review by CDE staff.

## Methodology

The study was conducted mainly utilizing the cognitive lab methodology. Given the primary purpose of the study to examine test users’ interactions with computer-based features, it was essential to closely observe each student’s behavior during the pilot administration. A sizable body of relevant literature has documented the benefits of the methodology not only for collecting validity evidence but for gathering practical information for developing innovative assessments for a target population with diverse backgrounds (e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Leighton, 2017; Winke et al., 2018; Winter, Kopriva, Chen, & Emick, 2006; Wolf, Lopez, Oh, & Tsutagawa, 2017; Wolf, Guzman-Orth, & Wain, 2015). Thus, for this study we employed the cognitive lab method which included close observation of test users’ test-taking processes followed by retrospective, cognitive interviews. To gather test examiners’ feedback on the computer-based ELPAC materials and their administration experiences, a survey and a focus-group interview with test examiners were also conducted.

Detailed information about the participants, study instruments, procedure, and data analysis is described in the following subsections.

### Sampling and Recruitment

In conducting cognitive labs for usability testing, previous literature places an emphasis on purposeful, targeted sampling rather than on sampling size (Beatty & Willis, 2007; Willis, 2005). Research suggests that samples of 5 to15 cases are common practice for cognitive labs that are intended to identify issues in usability testing (e.g., Blair & Conrad, 2006; 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). In this study, we made a deliberate sampling plan to include a variety of ELs in terms of key EL background characteristics. Such key background characteristics for consideration included grade level, ELP level, home language, length of U.S. residence, disability status, and computer familiarity. While efforts were made to recruit diverse students in terms of these background characteristics, students’ computer familiarity was of particular importance in the sampling criteria. Specifically, recruitment efforts were made to ensure that students who had little experience in computer-based assessment (e.g., transitional kindergarten through grade two students and recently arrived ELs) were included in the study. According to a federal document (i.e., Resource Guide: Accountability for English Learners under the ESEA, U.S. Department of Education, 2017, p. 21), “recently arrived ELs are defined by the statute as ELs who have been enrolled in schools in the U.S. or the District of Columbia (not including Puerto Rico and the outlying areas) for less than 12 months (ESEA sections 1111(b)(3)(A) and 8101(48)).” Given a wide range of formal educational background of these students, the inclusion of recently arrived ELs in the study sample was of critical importance. A small number of non-EL students who would be able to perform their grade-appropriate tasks were also included in the sampling criteria to allow researchers to identify any EL-specific difficulties in interacting with computer-based assessment features.

The targeted sample included students across transitional kindergarten through grade eight only. The rationale to exclude grade span nine through twelve was that ELPAC task types are the same in grade span six through eight and grade span nine through twelve, and that linguistic and cognitive processes of grade spans six through eight and nine through twelve would be similar. That is, findings about the usability of computer-based assessment features based on the sample for grade span six through eight would be applicable to the conversion of materials for grade span nine through twelve for the computer-based assessment format. Furthermore, targeting the sample in this way was expected to reduce strain on the LEA to provide participants.

In consultation with the CDE, ETS and SCOE selected two specific LEAs to conduct the study to meet the target sampling criteria. Specific attention was paid to LEAs with differing degrees of capacities in terms of technology use for students. In addition, details about the availability of accessibility resources were communicated with each district coordinator so that schools could identify appropriate students to participate in the study, particularly for the sample of ELs with disabilities who could take the general ELPAC. Study participation was on a voluntary basis, and parents/guardians’ consent forms were collected for participating students.

With respect to test examiners, educators who had expertise and experience in ELPAC training, administration, and scoring procedures were recruited across the state.

### Participants

A total of 107 (89 EL and 18 non-EL) students from transitional kindergarten to grade eight from six schools in two school districts participated in the study. Table 2 presents the number of students by grade level or grade span. Of those EL students, 24 recently arrived ELs (who had been in U.S. schools fewer than 12 months at the time of the study) and were enrolled in grades one through eight included students who were born outside of the continental U.S.: China, Colombia, Egypt, Mexico, Puerto Rico, and Ukraine. Of the EL students participating in the study (excluding recently arrived ELs), 28 students were born outside of the continental U.S.: China, Colombia, Haiti, Mexico, Puerto Rico, Ukraine, and Vietnam. The remaining students in the study were born in the United States. Of the participating students who took the kindergarten test form, six students were enrolled in transitional kindergarten and five of those six students were recently arrived ELs.

Table 2. Number of Participating Students by Grade Level and Grade-Span

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Status | TK/K | Grade 1 | Grade 2 | Grade span 3–5 | Grade span 6–8 | Total |
| ELs | 23 | 11 | 13 | 17 | 25 | 89 |
| Non-ELs | 3 | 4 | 4 | 3 | 4 | 18 |
| Total | 26 | 15 | 17 | 20 | 29 | 107 |

Participating students’ ELP levels were varied with 18 percent of the students being Level 1, 29 percent of the students being Level 2, 38 percent of the students being Level 3, and 16 percent of the students being Level 4, based on their previous ELPAC scores. The majority of the EL students’ home language was Spanish (*n* = 68). The remaining students’ home languages were diverse, including Arabic (*n* = 3), Cantonese (*n* = 1), French/Creole (*n* = 1), Hmong (*n* = 4), Mandarin (*n* = 4), Punjabi (*n* = 1), Russian (*n* = 1), Ukrainian (*n* = 3), and Vietnamese (*n* = 2).

Participating students were asked whether they had experience using a computer in school or at home. Both EL and non-EL groups in kindergarten to grade two included students who had no experience using a computer in school (ranging from 33 percent to 75 percent of participating students in kindergarten to grade two). Three students (18 percent) in grades three through five responded that they had no experience in using a computer in school while all participating students in grade span six through eight had used the computer in school. Across grades, there was a higher range of students in both EL and non-EL groups who did not use a computer at home (50 percent in kindergarten, 42 percent in grade one, 42 percent in grade two, 47 percent in grade span three through five, and 30 percent in grade span six through eight). In sum, the students that participated in the study had varying levels of computer familiarity.

Of 89 EL students, 13 students were ELs with disabilities. Table 3 summarizes the number of participating ELs with disabilities by grade levels and their primary disability type as reported by school personnel.

Table 3. EL Students with Disabilities by Grade Level or Grade-Span and Primary Disability Type

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Primary Disability Type | TK/K | Grade 1 | Grade 2 | Grade span 3–5 | Grade span 6–8 | Total |
| Attention Deficit Hyperactivity Disorder | 0 | 1 | 0 | 0 | 0 | 1 |
| Autism Spectrum Disorder | 1 | 0 | 3 | 0 | 0 | 4 |
| Intellectual Disability | 0 | 1 | 0 | 0 | 0 | 1 |
| Specific Learning Disability | 0 | 1 | 0 | 1 | 4 | 6 |
| Section 504 Plan | 0 | 0 | 1 | 0 | 0 | 1 |
| Total | 1 | 3 | 4 | 1 | 4 | 13 |

A total of 19 test examiners participated in the study. Test examiners ranged in their professional roles, such as an administrator (*n* = 4), EL support personnel (*n* = 4), teacher (*n* = 4), ELD coordinator (*n* = 3), ELD program specialist (*n* = 3), and ELD curriculum specialist (*n* = 1). On average, the test examiners had taught for more than 10 years (*n* = 14), with a majority of the test examiners having specifically taught ELs for more than 10 years (*n* = 8, and one having taught ELs for 10 years or fewer) in ELD settings (*n* = 9). The majority of test examiners also had experience administering the paper-based ELPAC (*n* = 12).

### Study Instruments

ETS developed a set of study instruments reviewed and approved by the CDE to collect the data for the study. A description of each instrument follows.

#### Student Background Questionnaire

The student background questionnaire was developed to gather relevant background information about participating students (e.g., grade, gender, age, length of stay in the United States, recently arrived EL status, country-of-origin status, ELP, home language, computer experience, and disability type). The questionnaire was completed by a teacher who was familiar with the participating student.

#### Cognitive Interview Protocol Including a Rating Form

The cognitive interview protocol was developed to facilitate the standardized observation of and interview with a student. For efficiency, test examiners conducted the cognitive interview based on this protocol immediately following the test administration of each domain. The protocol contained background questions about the student’s experiences using computers and taking technology-enhanced assessments, questions specific to each item type or online tool, and questions to collect overall feedback and comments on the items and the embedded universal tools. Additionally, the protocol included a rating form where test examiners entered their evaluation of the students’ interaction with the platform and the items based on their observation notes and interviews with students.

#### Observation Protocol Including a Rating Form

The observation protocol was developed for ETS staff to systematically observe the test examiners’ and students’ performance with the pilot test and to record their notes and ratings related to the targeted areas of the study. The protocol contained the same rating form included in the Cognitive Interview Protocol in the format of a Likert scale (e.g., Entirely, Partially, Not at All) to evaluate the degree to which a student was able to navigate the platform independently. In addition, the observation protocol included a rating form to evaluate the degree of usability of the test examiner’s interaction with the student, the platform, and the *DFAs*. Each rating form also included explicit spaces for observers to take detailed notes to justify their ratings.

#### Semi-structured Focus Group Interview Protocol

The semi-structured interview protocol was used for ETS staff to systematically conduct a focus-group interview with test examiners at the end of each pilot day. The protocol consisted of questions to collect test examiners’ feedback on the test administration materials and areas to be improved.

#### Test Examiner Survey

The test examiner survey was developed to further collect test examiners’ written feedback on the usability of the *DFAs*, any difficulties in administration and scoring, and areas of concern about students’ interaction with the pilot items and platform features.

#### Tutorial Videos

A tutorial video was developed as part of the study instruments (not for operational testing uses) to ensure that the students were introduced to the basic features of the test-delivery system, such as navigation and universal-tool functionalities. This tutorial video was played at the beginning of each cognitive lab, except for the kindergarten and grade one administrations. For kindergarten and grade one administrations, the test examiner mainly interacts with the computer features and enters students’ responses on the computer. Additionally, a separate, brief tutorial video for each domain was developed and shown to the student at the beginning of each domain. These tutorial videos described the audio buttons (e.g., [**Play**], [**Pause**], [**Stop**]) for Listening, the layout for Reading items, and the writing tools for the Writing domain.) The Writing domain tutorial was only administered to grades three through eight since the paper-based ELPAC was maintained for kindergarten through grade two students for the Writing domain.

To check students’ understanding of the features shown in the tutorial video prior to proceeding to the pilot items, the test examiner explicitly asked each student about their understanding of the tutorial videos. In the case that students expressed difficulties in understanding the tutorials due to their English language proficiency, school staff, ETS staff, or SCOE staff were available to describe the features in students’ home languages to the extent possible.

### Procedure

Prior to the pilot and cognitive lab administration, separate trainings for observers and test examiners were conducted. Observers consisting of 11 ETS staff members (researchers, assessment specialists, psychometricians, program managers) took part in a series of training sessions to have hands-on experience with the platform, pilot tests, and protocols. SCOE, in collaboration with ETS, conducted in-person training sessions for participating test examiners. The training sessions focused on the use of the test platform, scoring guidelines, cognitive interview protocol, and *DFAs*.

The pilot and cognitive lab administration (i.e., data collection) took place from April 1 to April 9, 2019, in two school districts described in the sampling section. Site coordinators were designated from each district and provided their assistance in collecting consent forms, coordinating daily testing sessions, assembling daily focus-group meetings, and completing participating students’ background questionnaires. Each school provided quiet rooms for the testing sessions. Each cognitive lab session was conducted one-on-one with a test examiner and a student while an observer was present.

In the beginning of the session, a test examiner described the intent of the testing to the student and encouraged the student to speak to the test examiner about any difficulties during the testing. The test examiner followed the cognitive lab protocol, starting by asking the student the background questions (e.g., language in which the student wanted to speak, the student’s computer familiarity, etc.). Then, the test examiner and student watched the tutorial video together, and the test examiner checked with the student about the student’s understanding of the tutorial video. During testing, items from duplicate item types were skipped (i.e., a targeted pool of items from each item type were selected for administration). Test examiners asked students retrospective questions from the cognitive interview protocol immediately following each domain test. Translation assistance for the interview was provided when resources were available (e.g., a Spanish-speaking translator was present in cases where a student spoke only Spanish).

During the testing session, an observer took detailed notes and completed the ratings forms embedded in the observer protocol. All sessions were audio-recorded; select sessions were video-recorded based on parent/guardian consent. On average, the sessions—including both the test administration and interview—took approximately 42 minutes for students taking the kindergarten form, 46 minutes for students taking the grade one form, 56 minutes for students taking the grade two form, 85 minutes for students taking the grade span three through five form, and 87 minutes for students taking the grade span six through eight form.

At the end of each day’s testing sessions, ETS staff facilitated a focus-group interview with the test examiners, focusing on gathering test examiners’ feedback about the test administration procedures, materials, and any areas for revision and improvement. The daily focus-group interviews were audio-recorded. Additional ETS staff took notes during the interviews.

At the end of their participation in the Usability Pilot, test examiners also completed the test examiner survey, providing their perception and feedback on the usability of test materials.

### Data Analysis

Collectively, the data sources were comprised of students’ background information, ratings from the protocols, observation notes, focus-group interview notes, and test-examiner-survey responses. Although students’ responses were collected, scores were not examined, as the focus of the study was to examine the usability of the computer-based assessment materials and test-taking processes on the computer.

The ETS team compiled all the data into Excel files to organize the data. The observation notes, test examiners’ comments, and focus-group discussions were qualitatively analyzed to find recurring themes regarding the usability of the materials and the areas for improvements. The coding categories included students’ understanding of directions, need of technological assistance, independence of the use of navigation features in each domain, and use of universal tools. Additional categories related to test-examiner materials included the clarity of the *DFAs*, ease of administration, and ease of scoring. In each category, the areas of difficulties that emerged in the data were summarized for further modification or improvement. Descriptive statistics for the ratings and survey responses were also computed and summarized. The data were summarized for patterns in response to three areas posited in research questions:

1. Students’ interaction with computer-based assessment platform and features
2. Use of accessibility resources
3. Test examiners’ usability for testing materials

## Findings

In this section, the results of the analyses are summarized by three areas of investigation that answer the research questions in each respective area. In addition to qualitative summaries of emerging patterns, the results of the frequency analysis of the observers’ ratings are presented wherever applicable. Due to the intended variation in administration of the grade level and grade-span forms, some areas may not include all grade levels and grade spans. Since some students skipped more items than others due to logistical issues (e.g., scheduling and time constraints or technology problems), the total number of students varied among the areas of analysis.

### Students’ Interaction with the Platform and Computer-Based ELPAC Items

The first area of investigation focused on students’ ability to use the testing platform and the computer-based ELPAC features. Research questions in this area largely concerned students’ understanding of directions and ability to navigate the computer-based assessment features, including the audio-delivery functions, the Speaking recording functions, and the writing tools. In general, observation notes indicated that participating students were highly engaged in the computer-based ELPAC items during cognitive lab sessions.

Some usability difficulties were observed for students. Specific usability areas are reported next. The findings are summarized for EL students by grade level. Although the analyses were also conducted by EL students’ background characteristics such as ELP levels, recently arrived ELs, and computer familiarity, no specific patterns emerged for these background characteristics. Some individual students’ characteristics are noted wherever relevant for the usability areas of interest that follow.

#### Understanding Directions

The observers took detailed notes and rated the extent to which students understood general directions in each domain, indicating whether the students knew what they were expected to do on the computer. The results of the frequency analysis of the ratings indicated that most students (over 80 percent) across grades and across domains understood the directions without any difficulty. Overall, summaries of the average frequency in each rating category across items for each domain (except Speaking) were calculated. Speaking was not included because the Speaking domain test was designed as a one-on-one administration where the test examiner explains the directions and guides the navigation of items for a student individually. On the same token, students in kindergarten to grade one were excluded for the frequency analysis for understanding directions about the computer interaction since test examiners controlled the computer navigation for these grades. In general, the findings suggest that students within and across the grade level/grade-span tests mostly understood the directions on the domain level tests (ranging from 60 percent to 91 percent of the students demonstrating they entirely understood the directions for their domain level tests).

A closer look at the ratings across items suggests that students tended to have more difficulty in understanding directions for the first item of the first domain (i.e., Listening was the first domain administered in this study). The observation notes indicate that some students did not know what to do when they saw the first screen of the test. However, once students became familiar with the [**Play**] button to listen to audio files and the [**Next**] button, all students eventually understood what they were asked to do. This pattern of increased familiarity with the directions within each domain test emerged across all grades. This pattern is also true of non-EL students, suggesting that the familiarity of the new computer-delivery format contributed to students’ behaviors and observers’ ratings.

For students who were rated as having “Partially” understood the directions or having understood them “Not at all,” the notes from observers and test examiners revealed some common themes: younger students (kindergarten to grade two) tended to confirm with the test examiner as to what to do next, such as advancing to the next item and selecting the [**Play**]button. In the Listening and Reading domains, some ratings of “Partially” or “Not at All” occurred when students did not realize that they needed to scroll down the screen to see multiple questions on one screen. In the Speaking domain, the “Partially” or “Not at All” ratings were mostly related to the cases where students started to answer the questions before the test examiner pressed the recording button. As described earlier, this case tended to occur more often for the first item of this domain. In the Writing domain, the “Partially” ratings were given in cases where students asked about or confirmed where to enter their responses. Generally, “Not at All” ratings were assigned when students could not read the directions due to their ELP. In these situations during the cognitive lab sessions, a school staff member explained directions in the student’s home language.

#### Navigating the Platform and Items Independently

The cognitive lab protocol explicitly asked test examiners and observers to rate and take detailed notes on students’ ability to navigate the platform and general computer-based ELPAC features independently in grades two through eight (e.g., understanding the layout to proceed to the next item and navigating the platform independently). The rating for this category in kindergarten and grade one was not applicable since test examiners navigated the platform and computer features for students in these grades. Table 4 through Table 9 present the summary of the average frequency in each rating category across items for each domain. As mentioned previously, the Speaking domain was excluded from the analysis for this category because test examiners guided students individually during the Speaking domain.

In Table 4 through Table 6, percentages appear in parentheses. Due to rounding, percentages do not always equal 100.

Table 4. Navigating Independently in Listening: Frequency and Percentage of Students Observed in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| Grade 2 | 2 (18) | 6 (54) | 5 (45) |
| Grade span 3–5 | 0 (0) | 4 (24) | 13 (76) |
| Grade span 6–8 | 0 (0) | 3 (13) | 21 (88) |

Table 5. Navigating Independently in Reading: Frequency and Percentage of Students Observed in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| Grade 2 | 3 (21) | 4 (29) | 7 (50) |
| Grade span 3–5 | 1 (6) | 5 (31) | 10 (63) |
| Grade span 6–8 | 0 (0) | 1 (3) | 22 (96) |

Table 6. Navigating Independently in Writing: Frequency and Percentage of Students Observed in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| Grade span 3–5 | 1 (7) | 4 (29) | 9 (64) |
| Grade span 6–8 | 2 (9) | 2 (9) | 19 (82) |

As shown in Table 4 through Table 6, the higher the grade level, the better that students were able to navigate the platform and computer-based assessment features independently. In terms of difficulties that students experienced, observation and focus-group notes indicate that some students did not realize that there were multiple items on one screen. This was particularly the case for some second and third graders in the Listening and Reading domains. In these instances, test examiners prompted students to scroll down on the screen. Examining the observation notes, the scroll bar was the most frequently reported comment associated with ratings of “Partially” and “Not at All” for grade two students. These students were observed to learn to scroll during the testing session. One of the observation protocol ratings specifically asked about students’ use of the scroll bar in the Reading domain, where reading passages may be longer than the size of the computer screen.

Table 7 displays the average frequency of each rating category across the items for the use of scrolling in the Reading domain. The results indicate that younger students (i.e., grades two and three) had more difficulty using the scroll bar than other students. One second grade student even asked about the meaning of “scrolling.” While some second grade students seemed to have little experience with scrolling, they were eventually able to scroll up and down the screen after test examiners’ assistance during testing. This pattern was also exhibited in non-EL students, suggesting that students in younger grades have more limited experience with computers than those in older grades in general. Although there were six recently arrived ELs in grade span six through eight, they had no difficulty in using the scroll bar. All of these students had previous experience with computers at school, while some of them did not use a computer at home. It was also observed that some students used the touch-screen function to scroll the screen up and down.

In Table 7, percentages appear in parentheses. Due to rounding, percentages do not always equal 100 percent.

Table 7. Using the Scroll Bar Independently: Frequency and Percentage of Students Observed in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| Grade 2 | 3 (30) | 1 (10) | 6 (60) |
| Grade span 3–5 | 2 (13) | 2 (13) | 12 (75) |
| Grade span 6–8 | 0 (0) | 0 (0) | 23 (100) |

Comments regarding “Partially” or “Not at All” ratings of students’ independent navigation included some third to eighth grade students who asked how to advance to the next item in the Writing domain. The Writing domain was designed in a way that students were able to advance to the next item in two ways (i.e., by selecting the [**Next**] button or by selecting the [**Item Number**] tab). Some students seemed confused by these two ways in the beginning of the Writing test, but the students grew familiar with these options as they progressed through the domain.

Another reason for a “Partially” rating across domains occurred when students asked test examiners whether they could move to the next item. This confirmation behavior occurred for some second and third grade students. Notably, this behavior also occurred for non-EL students, indicating a general test-familiarity issue, not an EL-specific issue.

#### Need for Technological Assistance

Students’ propensity to request technological assistance was intricately intertwined with their interactions with and usability of the computer-based ELPAC. Detailed further in [subsection 6.2](#_Accessibility_Resources), observers noted cases where students needed technological assistance from test examiners.

Table 8 summarizes the average frequency of cases when students needed technological assistance across items in each domain. Note that the Speaking domain is excluded since test examiners provided their assistance as prescribed in the DEI. As presented in Table 8, the pattern was that upper-grade students were observed to need fewer instances of technological assistance compared to lower-grade students.

In Table 8, percentages appear in parentheses. Due to rounding, percentages do not always equal 100.

Table 8. Need of Technological Assistance: Frequency and Percentage of Students in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Domain | Grade or Grade Span | No | Yes |
| Listening | Grade 2 | 7 (54) | 6 (46) |
| Listening | Grade span 3–5 | 13 (76) | 4 (24) |
| Listening | Grade span 6–8 | 21 (88) | 3 (13) |
| Reading | Grade 2 | 8 (67) | 4 (33) |
| Reading | Grade span 3–5 | 13 (76) | 5 (29) |
| Reading | Grade span 6–8 | 22 (96) | 1 (4) |
| Writing | Grade span 3–5 | 12 (80) | 3 (20) |
| Writing | Grade span 6–8 | 20 (91) | 2 (9) |

The types of assistance that test examiners provided across domains included selecting [**Play**] buttons to listen to audio, finding the [**Next**] button, and scrolling. Two students also asked about the use of universal tools (e.g., the use of highlighter and zoom). Test examiners’ assistance was also needed when technical issues arose, such as the appearance of a pop-up message or when the computer screen froze.

#### Usability of the Audio Delivery Model

The computer-based ELPAC was designed to deliver the Listening of stimuli and answer choices via audio files with play, pause, and stop functions. The observers took detailed notes and used the rating scales related to students’ behaviors during the Listening test. Observation notes indicated that students, in general, were engaged in listening to audio stimuli across all grade levels. However, it was also noted that test examiners’ verbal interaction with younger grade students in kindergarten and grade one was important in making students feel comfortable and in directing students’ attention to the test items.

As students, starting in grade two, were expected to use the audio features independently, the observers rated and took notes about students’ independence to use the audio buttons (i.e., selecting the [**Play**], [**Pause**], or [**Stop**] buttons). Table 9 presents the average frequency of each rating category across Listening items for the independent use of audio buttons. While students in grade span six through eight had no difficulty in using the audio buttons independently, students in grades two and three were found to have more difficulty. As described earlier, some students did not realize that they were expected to select the [**Play**] button to listen to the stimuli when they first saw the item. This pattern was also found in non-EL students. These students all learned that selecting the [**Play**] button activated the audio files so that they could listen to stimuli during testing. In addition, the protocol rating form asked observers to indicate when students used the audio scrubber to select a specific place to listen in the audio file, instead of listening to the entire stimuli. It was found that only one sixth grade student used the scrubber.

In Table 9, percentages appear in parentheses. Due to rounding, percentages do not always equal 100.

Table 9. Using Audio Buttons Independently in Listening: Frequency and Percentage of Students in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | No | Partially | Entirely |
| Grade 2 | 3 (23) | 1 (8) | 9 (69) |
| Grade span 3–5 | 1 (6) | 3 (17) | 14 (78) |
| Grade span 6–8 | 0 (0) | 0 (0) | 23 (100) |

The computer-based ELPAC Listening domain also contained the function where students could listen to stimuli multiple times using the audio buttons. The number of times a student is able to listen to a stimulus was one area of interest in making a final decision for the Listening domain design. The observers’ rating data demonstrated that the majority of students listened to the stimuli once, as shown in Table 10. One student was found to play the stimuli more than three times, raising some concern about score interpretations. While this student was an EL with disabilities, the general functionality of replaying multiple times does raise some concern about score interpretations for non- EL students with disabilities.

In Table 10, percentages appear in parentheses. Due to rounding, percentages do not always equal 100.

Table 10. Listening to a Stimulus More than Once: Frequency and Percentage of Students in Each Rating Category

|  |  |  |
| --- | --- | --- |
| Grade or Grade Span | No | Yes |
| TK/K | 18 (86) | 3 (14) |
| Grade 1 | 9 (90) | 1 (10) |
| Grade 2 | 11 (85) | 2 (15) |
| Grade span 3–5 | 14 (82) | 3 (18) |
| Grade span 6–8 | 19 (83) | 4 (17) |

#### Usability of Speaking Recording Functions

The computer-based Speaking test is one of the significant changes compared to the paper-based ELPAC in that students are expected to speak to the computer while recording their responses. To mitigate effects from this substantial change from the paper-based ELPAC, a one-on-one administration was maintained for the Speaking domain for all grade levels. In doing so, test examiners provided technological individual assistance in recording responses for students. The observers’ ratings and notes were thus regarding the extent to which students were hesitant to speak on the computer and record their responses.

Table 11 presents the summary of the average frequency across Speaking items for each rating category. Almost all students were observed to have no hesitation to speak to the computer and record their responses. One or two students were rated as “Partially” or “Entirely” hesitant in each grade or grade span. Observation notes suggest that the students were too shy or did not respond to the item due to its content difficulty. The shyness was also observed from non-EL students with “Partially” or “Not at All” ratings. Students with a “Partially” rating in kindergarten and grade one needed the test examiners’ prompting and encouragement to complete the given item.

In Table 11, percentages appear in parentheses. Due to rounding, percentages do not always equal 100 percent.

Table 11. Hesitance to Record Speaking Responses: Frequency and Percentage of Students Observed in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| TK/K | 21 (95) | 0 (0) | 1 (5) |
| Grade 1 | 8 (80) | 1 (10) | 1 (10) |
| Grade 2 | 11 (85) | 1 (8) | 1 (8) |
| Grade span 3–5 | 14 (93) | 0 (0) | 1 (7) |
| Grade span 6–8 | 19 (86) | 2 (9) | 1 (5) |

The Speaking domain was designed to be one-on-one with test examiners’ playing a significant role in using the [**Recording**] button. However, it was also observed that some students independently pressed the [**Recording**]and [**Stop**] buttons. Particularly, students in grade span six through eight were found to have little difficulty using those buttons, even as they learned how to use them during the testing.

#### Typing on the Computer to Respond to Writing Items

For the computer-based ELPAC, starting in grade three, students are asked to type their responses in the Writing domain. Students in grades three and above also take state content assessments in English language arts and mathematics on the computer. One of the areas of interest for the current study lay in students’ abilities to use the keyboard to type their responses for the Writing domain.

Table 12 summarizes the average frequency of students in each rating category across the Writing items regarding the easiness of typing a writing response. Students in the upper grades were found to be more skillful at typing than those in the lower grades in general (e.g., finding the keys easily with a relatively fast pace). Ratings of “Partially” or “Not at All” were mainly due to typing speed. The observers noted that some students typed very slowly using one or two fingers. A closer examination of the data indicated that the students in the ratings of “Partially” or “Not at All” were mostly third graders with little computer experience.

Some EL students in grades three were also found to need some help from test examiners on how to delete the typed letters using the backspace keyboard. The students who had difficulties typing in grade span six through eight had no clear patterns in terms of their background characteristics. These students seemed to need more practice to improve their typing skills. In all grades from three through eight, the students who were recently arrived ELs did not show any particular pattern in the ratings of typing ability. All non-EL students in the sample were not observed having any difficulty with typing.

In Table 12, percentages appear in parentheses. Due to rounding, percentages do not always equal 100.

Table 12. Easiness of Typing on the Computer: Frequency and Percentage of Students in Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| Grade span 3–5 | 1 (7) | 7 (50) | 6 (43) |
| Grade span 6–8 | 1 (4) | 4 (17) | 18 (78) |

#### Usability of the Writing Tools

As one of the computer-based ELPAC features, the writing tools consisted of various formatting functions (e.g., font style, indent) and were embedded in the writing-response window. The data on the writing tools focused on students’ ability to use the tools independently. Observers’ data indicated that most students did not use the writing tools. In fact, completing the writing tasks did not necessarily require the use of the writing tools. Among those who used the writing tools (five students in grade span three through five and eight students in grade span six through eight), one student in each grade span had the rating of “Partially” independent in their use of a tool due to difficulty in selecting the text to apply the formatting style (e.g., underlining).

### Accessibility **Resources**

The second area of investigation focused on students’ interaction with the accessibility resources available for the computer-based ELPAC, including the universal tools. Universal tools are available for all students to self-select, choosing whether, when, and how they would like to use the tools. At the time of this study, the following universal tools, including both embedded and non-embedded resources where applicable, were available:

Breaks

Digital notepad

Expandable screen for reading passages and items

Highlighter

Keyboard navigation

Line reader

Mark for review

Scratch paper

Simplified directions

Strikethrough

Test examiner assistance with test navigation

Writing tools

Zoom

Research questions guiding this area of investigation focused on how the participating ELs and ELs with disabilities used those universal tools and whether they had any difficulties in using them. Overall, evidence from the data suggests that students self-selected their tool usage, and use was variable. The following section describes the major patterns of students’ universal-tool usage on the computer-based ELPAC.

#### Universal Tool Use

A small number of students across grades two through eight was observed to use the universal tools. The types of tools that these students used were varied. A summary of the number of students in each grade or grade span who used each tool is provided in Table 13. Students from kindergarten and grade one were excluded from the analysis as they were not self-selecting the universal tools, but were guided by the test examiners’ use of the tools.

Table 13. Number of EL Students Who Used Each Universal Tool by Grade or Grade Span

|  |  |  |  |
| --- | --- | --- | --- |
| Universal Tool | Grade 2 | Grade span 3–5 | Grade span 6–8 |
| Breaks | 1 | 0 | 1 |
| Digital notepad | 0 | 0 | 1 |
| Expandable item | 1 | 3 | 2 |
| Expandable passage | 1 | 6 | 5 |
| Highlighter | 1 | 3 | 1 |
| Keyboard navigation | 0 | 1 | 0 |
| Line reader | 0 | 4 | 3 |
| Mark for review | 0 | 1 | 0 |
| Scratch paper | 0 | 4 | 8 |
| Simplified directions | 0 | 1 | 1 |
| Strikethrough | 0 | 0 | 1 |
| Test navigation assistant | 6 | 5 | 3 |
| Writing tools | 0 | 3 | 1 |
| Zoom | 0 | 2 | 3 |

Overall, across grades, assistance for test navigation was used by the most students (14 students) while expandable passages and scratch paper were used by 12 students each. Line reader was the next most-used tool (seven students), with the expandable items used by six students, and the highlighter and zoom both used by five students each. Four students were observed to use the writing tools, while only two students requested breaks or simplified directions, and only one student used the mark for review passage and strikethrough of the response options.

In grade two, three students were observed to request assistance from the test examiner to help them navigate the computer-based ELPAC. Eleven of 17 students in grade span three through five were observed to use the universal tools. For students in grade span three through five, the most frequently used tool was the expandable passages feature, which was used by six students. For the test navigation assistant tool, interestingly, four students were observed to request assistance with navigating the test platform 14 separate times, ranging from two to five times each.

Similar variation was noted for students taking the grade span six through eight test form. Across the grade span six through eight test form, 13 of 25 students were observed to use the universal tools. Unlike students in other grades and grade spans, students in grade span six through eight used the non-embedded scratch paper the most (used by eight students across 10 instances on the Writing and Speaking domains), followed by the expandable passages tool (used by five students). Expandable passages was used by five students seven times across two task types. Test navigation assistance was needed by three students on eight occasions across all four domains. Observation notes also indicated that observers gave the “Partially” ratings when students had some difficulty in using the universal tools (e.g., difficulty selecting the text to highlight or zooming).

For ELs with disabilities, students used specific tools or their test examiner selected tools to assist in the administration. One first grader had a test examiner who used the zoom and expand passage feature to help the student visually focus on specific content on the screen, rather than focus on the split-screen presentation. This same first-grade student, along with another first-grade student, needed additional resources from the test examiner above and beyond the allowed universal tools; these will be described more in the closing sections. One second grader, one third grader, and one sixth grader were observed using the universal tools. The second grader needed a break in the middle of the test administration to go with the school nurse for blood glucose monitoring. The third grader used the scratch paper in the Writing domain, and the sixth grader used the scratch paper in the Writing and Speaking domains.

In terms of any difficulties using the tools, students demonstrated few difficulties when they chose to use the tools. Some younger grade students (e.g., grade two or grade three) demonstrated some difficulties selecting the text to highlight or underline their text for their typed response on the Writing domain (grade three only).

#### Universal Tools’ Perceived Helpfulness

The universal tools were generally helpful for the students who used them, but as noted in the previous section, tool usage was not widespread across the student participants. Test examiners and observers also noted that some students seemed to forget that the tools were available during the testing session. Although students watched the tutorial video prior to interacting with the items, the initial exposure may not have been enough support for students to fully understand how and why they should use the tools. Additionally, some students (particularly those in grades three through eight) reported familiarity with the tools but chose not to use the tools because they thought the tools were not helpful or that they did not think they needed them.

#### Other Universal Tools Needed

As described previously, ELs in transitional kindergarten, kindergarten, and grade one were not expected to interact with the universal tools independently because their test examiners were intended to work one-on-one with the student, and the test examiner was expected to navigate the platform.

Two instances of test examiners administering the grade one test form to ELs with disabilities resulted in the test examiners needing to access additional non-embedded, unspecified resources to assist in the administration. Specifically, the two students were observed to be very sensory-seeking, which impacted the test examiners’ ability to administer the test on the computer. As a result, the test examiners used sensory toys (e.g., squishy toys) to provide tactile input for the students, and one test examiner used a thick, stiff pad to cover the keyboard and track pad when the student expressed continued interest in seeking sensory input on the face and would lay the head on the computer keypad to look at the screen.

Additionally, to assist in the ease of administration with the younger students (with and without disabilities), additional embedded resources were tried out during select cognitive labs. Specifically, these resources were intended to help students concentrate on the tasks; such tools as the enlarged mouse pointer and the option to change the color of the mouse pointer were used. Test examiners were instructed to layer these tools to create a large cursor that showed prominently on the screen and was a distinct color that would promote visual acuity. Test examiners who tried this approach reported that the change was favorable compared to the default option for a small mouse cursor that was the same color as the background of the test. One test examiner tried the extra-large-sized mouse cursor and reported that the size seemed too large against the layout of the content and thus would have preferred a large-sized cursor instead. Test examiners who tried this embedded resource expressed a desire to have access to a “quick start” guide to know how to set up this resource before they start testing each of their students. Similarly, test examiners administering the transitional kindergarten, kindergarten, and grade one test forms wanted more guidance on resources that were available, including guidance on how and when to use each resource to maximize the potential with the new computer-based delivery model for the ELPAC content.

Collectively, the results suggest that additional opportunities for practice and clarification on how to use the accessibility resources would be important for the students and test examiners to maximize the benefits of accessibility resources, including the universal tools.

### Usability of Test Examiner Materials and Scoring Guidelines

The third area of investigation examined the extent to which the materials that the test examiners used to administer the computer-based ELPAC were clear and easy to use. Generally, the research questions focused on clarity and use of the *DFAs* provided for all domains. Additionally, the scoring guidelines included in the *DFAs* and the DEI for the Speaking domain was of major interest given the new way to administer the Speaking domain on the computer (e.g., students talking on the computer to record their response under test examiners’ guidance). While the *DFAs* were mostly clear for the test examiners, common areas of usability challenges also emerged. These areas are detailed next.

#### Clarity of the *DFAs*

The observers were asked to rate and take detailed notes as to whether test examiners experienced any difficulty using the grade-level and grade-span-specific *DFAs* because of clarity issues. As shown in Table 14, excepting grade one, a majority of the test examiners were observed to clearly interpret the *DFAs* and administer the test accordingly across the Listening, Reading, Writing, and Speaking domains. In grade one, the data suggests that test examiners had a more difficult time with the clarity of the *DFAs*. Complementing the observation rating results, Table 15 summarizes test examiners’ survey responses on the ease of use of the *DFAs*.

Observation, survey response data, and focus-group notes indicate that test examiners wanted more explicit direction across all domains. They requested direction specific to whether or not to read aloud the general instructions, how to point to the screen to track when reading (particularly when the screen on the computer is touch-activated), and how to determine when a student should navigate the test versus when the test examiner should navigate the test. This direction seemed particularly important for the grade one test administrations since some students tended to try to self-navigate and self-pace their interaction with the content; sometimes the first graders’ desire for independence caused some challenges for the test examiners. For example, test examiners reported that it was confusing to know when to scroll the reading pane for the student to read all the content on the screen. This confusion was particularly evident when the students were reading silently in their heads and did not ask the test examiner to scroll. In these instances, test examiners were unclear about how to proceed and how and when to prompt the student. While this example is specific to test administrations for younger students, instances of general confusion about how to proceed and how and when to prompt were evident across all grade levels.

These administrative considerations may have influenced how test examiners felt about the *DFAs* and whether or not they were easy to use. Overall, across the grade level and grade-span test administrations, the test examiner did not think that the *DFAs* were easy to use. Rather, the majority of the test examiners reported some difficulties using the *DFAs* across the Listening, Speaking, Reading, and Writing domains.

In Table 14 and Table 15, percentages appear in parentheses. Due to rounding, percentages do not always equal 100 percent. The Writing test was not administered on the computer for students in transitional kindergarten and kindergarten or for grades one and two.

Table 14. Clarity of Directions in the *DFAs* for Test Examiners: Frequency and Percentage of Student Administrations Observed in Each Rating Category

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Domain | Grade or Grade Span | Not at All | Partially | Entirely |
| Listening | TK/K | 1 (5) | 2 (11) | 16 (84) |
| Listening | Grade 1 | 0 (0) | 4 (44) | 5 (56) |
| Listening | Grade 2 | 0 (0) | 2 (17) | 10 (83) |
| Listening | Grade span 3–5 | 0 (0) | 2 (13) | 13 (87) |
| Listening | Grade span 6–8 | 0 (0) | 3 (15) | 17 (85) |
| Reading | TK/K | 1 (5) | 4 (19) | 16 (76) |
| Reading | Grade 1 | 2 (17) | 3 (25) | 7 (58) |
| Reading | Grade 2 | 1 (7) | 2 (14) | 11 (79) |
| Reading | Grade span 3–5 | 0 (0) | 2 (13) | 13 (87) |
| Reading | Grade span 6–8 | 1 (5) | 5 (25) | 14 (70) |
| Writing | Grade span 3–5 | 0 (0) | 1 (7) | 14 (93) |
| Writing | Grade span 6–8 | 0 (0) | 1 (5) | 20 (95) |
| Speaking | TK/K | 1 (5) | 2 (10) | 18 (86) |
| Speaking | Grade 1 | 0 (0) | 4 (40) | 6 (60) |
| Speaking | Grade 2 | 0 (0) | 2 (17) | 10 (83) |
| Speaking | Grade span 3–5 | 0 (0) | 3 (25) | 9 (75) |
| Speaking | Grade span 6–8 | 0 (0) | 4 (22) | 14 (78) |

Table 15. Test Examiner Survey Response: Ease of Use with the *DFAs*

|  |  |  |
| --- | --- | --- |
| Statement | Agree (%) | Disagree (%) |
| The **Listening** test was easy to administer based on the *DFAs*. | 37 | 63 |
| The **Reading** test was easy to administer based on the *DFAs*. | 47 | 53 |
| The **Writing** test was easy to administer based on the *DFAs*. | 50 | 50 |
| The **Speaking** test was easy to administer based on the *DFAs*. | 21 | 79 |

#### Scoring the Speaking Domain Responses

As part of the procedures to administer the computer-based ELPAC, test examiners are also required to score the students’ spoken responses live “in the moment.” To do so, test examiners were asked to access the DEI on a separate device. The DEI contained the test examiner prompts to administer the Speaking items as well as the scoring rubrics and the rating options for the test examiner to select to assign a score for the spoken response. Observers were asked to record whether the test examiner had any difficulties in scoring the students’ response.

Overall, the results shown in Table 16 suggest that most test examiners had difficulties in scoring students’ responses during the Speaking administration. Observation notes, survey responses, and focus group notes suggest that test examiners had difficulty scoring the Speaking responses due to handling multiple concurrent tasks during the Speaking administration (e.g., reading *DFAs* in kindergarten to grade one, navigating tasks and recording functions for students, and maneuvering the DEI for scoring). The survey responses regarding the Speaking administration demonstrated that 79 percent of test examiners found the Speaking *DFAs* challenging to use and 84 percent of test examiners found the Speaking scoring challenging.

In Table 16, percentages appear in parentheses. Due to rounding, percentages do not always equal 100 percent.

Table 16. Ease of Scoring Speaking Tasks: Frequency and Percentage of Student Administration Observations Across Each Rating Category

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Not at All | Partially | Entirely |
| TK/K | 16 (76) | 5 (24) | 0 (0) |
| Grade 1 | 8 (73) | 2 (18) | 1 (9) |
| Grade 2 | 9 (75) | 2 (17) | 1 (8) |
| Grade span 3–5 | 10 (77) | 2 (15) | 1 (8) |
| Grade span 6–8 | 19 (90) | 1 (5) | 1 (5) |

Unpacking the Speaking domain data further, some patterns emerged in regard to difficulties associated with accessing the DEI. Additional patterns emerged regarding administering the task (i.e., orienting the DEI screen and the student-facing screen and moving back and forth seamlessly to maintain eye contact with the student; activating the recording button). Patterns also emerged regarding the usability of the DEI (i.e., locating and differentiating the components in the DEI such as the stimulus, the prompts, the directions, and the rubrics; scrolling to locate the rubrics after all of the prompts had been administered; selecting the rating and moving to the next item). Essentially, the DEI embodied multiple components and required multiple actions from the test examiners including the act of scoring the spoken responses. Observers noted that test examiners sometimes had to stop and double-check that the item that they were administering was the same item as was on the student-facing screen.

Test examiners generally noted that the act of double-checking was made more complicated by the seating arrangements for the Speaking administration. After trying out multiple seating arrangements, several test examiners reported preferring a 90-degree seating arrangement or a face-to-face seating arrangement where the student device is angled in such a way for the test examiners to see, compared to a side-by-side seating arrangement. Other test examiners managed to work a student self-checkpoint into their administration language (e.g., “I’m on task 4, make sure your screen is on task 4. Are you on task 4?”).

Observers also noted that sometimes when students were controlling the microphone button for recording, some test examiners also started giving students additional directions to help manage when the student turned off the microphone to ensure that the student did not turn off the recording before all of the follow-up prompting was administered (e.g., “I’m going to ask you some questions, make sure that you don’t stop the recording until I tell you to.”). Observation notes indicated that longer pauses ensued when test examiners were scrolling through the DEI to find the prompt, read through the rubric, and then enter the students’ score; test examiners confirmed that these longer pauses and administrative considerations made the Speaking domain challenging because the conversation that emerged with the new administrations was different from the conversational flow that they had experienced with the paper-and-pencil ELPAC administration.

Taken together, it is evident that while the test examiners were able to score the responses, they ultimately felt it was not an ideal process to employ during this small-scale usability pilot, and it would be even more challenging to employ scaling up for field testing or operational test administration when they are testing larger numbers of students. It should also be noted that at the time of the study, the materials were not yet fully developed and the training of test examiners using these materials was limited. Irrespective of limitations, the results of this study provided useful information to improve the test examiner materials for the Speaking administration for the future field testing and operational testing.

## Limitations of the Study

There are some limitations to the present study. Because the study was conducted at the initial stages of the development as part of the test development process, the platform, training, and *DFAs* were not yet fully developed. It is possible that additional training for students and test examiners may have ameliorated some of the initial usability challenges that were experienced and might have remedied themselves as users gained more familiarity with the platform.

Although a strength of cognitive lab methodology is the in-depth individualized data that is obtained through the observational and interview data, this method is limited to a relatively small sample size due to its one-on-one, intensive data collection. While the purposeful sampling used in the study targeted the key variables and student groups of interest, it is possible that other students with different characteristics not represented in the current sample may experience different outcomes when interacting with the computer-based ELPAC. Additionally, a convenience sample was used to collect the data. Although six schools across two LEAs participated in the study because they represented the key variables of interest, the sample was not representative of geographic diversity across the state of California. As a result, ETS cautions against wide-spread generalization of the results of the study. Despite these limitations, the study findings yielded critical information for the areas of improvements for the computer-based ELPAC.

## Recommendations

In this section, a set of practical recommendations are presented based on the findings in the three areas of investigation. As one of the goals of this study was to ensure that the computer-based ELPAC would be developed in a way that students demonstrate their ELP without interference by the usability of computer-based assessment features, action items are also proposed for each recommendation. These action items are intended to facilitate test developers in converting the paper-based ELPAC to the computer-delivery format appropriately in preparing for the field test and future operational administration of the computer-based ELPAC. Furthermore, recommendations and action items are related to enhancing the usability of the platform, computer-based ELPAC items, and their administration materials. Recommendations in this section follow the three main areas of investigation for this study. Each recommendation is followed by key contextual information and the recommended proposed actions to improve the development and administration of the computer-based ELPAC. Because of the interrelated nature of the data that was collected from the students and test examiners, the context for the recommendations may be not be unique and instead may be shared across recommendations.

Note that an addendum document to the present report (*Addendum: English Language Proficiency Assessments for California Usability Pilot Report,* ETS, 2019) is available to describe how the recommendations have been implemented.

### Recommendations on Usability by Students

#### Recommendation 1

Improve test-familiarity materials (tutorials, training tests, practice tests) to ensure students are prepared to take the computer-based ELPAC and test examiners are prepared to administer it.

##### Context:

This study was the first time that students saw the computer-based ELPAC, and observational data indicated that some students were unclear about the layout and item presentation, especially in the beginning of the test. Observations indicated that students became familiar with the test navigation as they proceeded. Following are proposed actions to implement to address students’ observed navigation issues. These proposed actions also address opportunities where student navigation can be improved by test examiners’ familiarity with the computer-based ELPAC and ancillary materials. See Section 6.1, Table 4 through Table 8 for more details.

##### Proposed Actions:

1. Ensure that the full suite of test-familiarity materials as planned (including tutorials, training tests, and practice tests) is developed and made available as scheduled in advance of the field test and the operational launch.
2. Clearly communicate to LEAs that test examiners and students participating in the field test should take the training tests prior to the field test.
3. Provide explicit training and direction for test examiners and students about the need to use the scroll bar and other embedded navigational tools that are built into the test delivery system to minimize risk of usability difficulties interfering with students’ ability to navigate items and enter responses.
4. Along with the training tests, provide test examiners the opportunity to practice using all platforms and tools that will be used to administer the Speaking domain, including the voice-capture tools, the DEI, and the *DFAs*.
5. Before the operational administration, make available full-length practice tests that represent the item types, task types, and content that students can expect to experience on an operational computer-based ELPAC.
6. Provide resources for educators, including test examiners, about key administration features of the computer-based ELPAC (i.e., technical specifications manual, test administration manual, etc.).
7. Communicate the changes and new features of the computer-based ELPAC prior to the field test, including, but not limited to, the following topics:
   * The number of times that a student can play the Listening stimuli
   * Reading on the screen for kindergarten to grade two
   * Recording, replaying, and re-recording Speaking responses
   * Available accessibility resources, including computer assistance for students who are in need of those resources
   * Test examiners’ roles in providing assistance in computer and interface navigation
8. Develop communication materials to inform educators, administrators, and parents about key features and changes made in the computer-based ELPAC.
9. For the Speaking domain, consider the inclusion of a practice item that requires test examiners and students to use the voice capture tools together.
10. Create an interactive tutorial that will orient students to the testing platform.
11. Consider making domain-specific tutorials in different languages.

#### Recommendation 2

Create resource materials for educators and test examiners to help determine if students are ready to take the computer-based ELPAC under typical conditions.

##### Context:

Observational data from the Usability Pilot indicated that students did not have proficient typing skills, evidenced by some students looking for the right letters on the keyboard. Scrolling was another skill that some students did not know. Given that students in the study, and across California, will demonstrate varying levels of computer familiarity, careful attention should be dedicated to ensuring students are appropriately prepared to take the computer-based ELPAC. To serve this purpose, additional resources may be created to help distinguish between students who have the necessary skills to take the computer-based ELPAC and those students who may need additional support. For more information, see the findings in Section 6.1, Table 8.

##### Proposed Action:

1. Create a checklist of students’ computer and technology skills to navigate essential and optional tools or features of the computer-based ELPAC for students and test examiners.

#### Recommendation 3

Allow students to listen only once to audio stimuli on the Listening test.

##### Context:

During the Usability Pilot, data indicated that students were able to listen and attend to the audio stimuli. Educators also favored the single-listen audio stimuli presentation because of the similarity to authentic contexts as well as the similarity to the original task design and administration of the paper-based ELPAC. For more information, see Section 6.1, Table 9 through Table 10.

##### Proposed Actions:

1. Before the field test, implement an item-level setting and configuration that will limit the number of times Listening stimuli are played to once. This setting would allow students to listen to Listening stimuli one time and listen to stems and options multiple times. Exceptions, where students are assigned the designated support that allows the students to listen to stimuli multiple times, would be managed via a test setting that would be entered in the Test Operations Management System (TOMS).
2. Another option to consider is posting the audio files to TOMS and have test examiners stream them, similar to what is done for the paper-based ELPAC.

#### Recommendation 4

Maintain recorded audio files for Listening stimuli on the kindergarten and grade one Listening tests, similar to the grades two through eight Listening tests.

##### Context:

Findings indicated that students in kindergarten through grade one were engaged in listening to the pre-recorded audio stimuli. Looking at the general practice in other ELP assessments, multistate assessment consortia such as the English Language Proficiency Assessments for the 21st Century has administered pre-recorded audio to both kindergarten and grade one students, and WIDA has administered pre-recorded audio to students in grade one. For more information, see Section 6.1.

##### Proposed Actions:

1. Maintain administering Listening with audio recordings at kindergarten to grade one.
2. Add directions to item-level audio files that prompt the student to point to the answer or state the answer.

### Recommendations on Accessibility Resources

#### Recommendation 5

Increase opportunity for familiarity and practice of accessibility resources for both test examiners and students.

##### Context:

Findings from the Usability Pilot suggested that students demonstrated variability in how they interacted with the universal tools. For students in the younger grades, use of the universal tools was demonstrated by the test examiner, while students in the upper grades appeared to be able to use the universal tools independently. For more information, see Table 8 in Section 6.1 and Table 13 in Section 6.2.

##### Proposed Actions:

1. Provide test examiners and students opportunities to learn and try out accessibility resources on ELPAC task types outside of the formal test-taking experience through the use of practice tests, training tests, and tutorials. These opportunities should be provided so that students ideally are not getting their initial exposure immediately preceding or during the test.
2. Provide explicit direction for test examiners in their training, as well as in their *DFAs*, to use universal tools. Aspects of training particularly relevant to accessibility resources include
   * providing test examiners of kindergarten, grade one, and grade two test administration training to use appropriate accessibility resources for their students (e.g., setting up streamline mode or using expand passages features, line reader to track text on the screen);
   * setting up any additional resources (e.g., large mouse cursor) to facilitate administration of the computer-based ELPAC; and
   * adding directions in training and *DFAs* to remind students to use the resources built into the platform, not affordances of the specific device (e.g., zoom using the test delivery system, not the track pad or touch screen) to minimize risk of unforeseen usability challenges (e.g., a student zooming beyond what is allowed by the built-in test delivery system constraints).

#### Recommendation 6

Provide appropriate supports to ensure a student’s level of familiarity with technology does not impede a student’s ability to take the computer-based ELPAC.

##### Context:

During the administration of the computer-based ELPAC, test examiners and students demonstrated varying degrees of familiarity with the technology and available resources designed to improve the test administration experience. For more information, see Table 8 in Section 6.1 and Table 13 in Section 6.2.

##### Proposed Actions:

1. Explore the possibility of making supports (i.e., print on demand or other solutions as specified in the *English Language Proficiency Assessments for California Accessibility Resources for Operational Testing*) available to help younger students who may not be comfortable reading stimuli in the Reading section on a computer screen.
2. Provide comprehensive communication around Matrix Four: Universal Tools, Designated Supports and Accommodations for the ELPAC and the Individual Student Assessment Accessibility Profile process. Test examiners should be made familiar with the enhanced accessibility resources allowed for the computer-based ELPAC during training and reinforced through reminders in the *DFAs*.

### Recommendations on Test Examiner Materials and Support

#### Recommendation 7

Simplify the Speaking administration to make the administration of the test and scoring easier for the test examiner.

##### Context:

Balancing the conversational administration and the need for in-the-moment scoring, a unique administration model was developed for the Usability Pilot. Findings indicated that this model was challenging for test examiners to manage. Particularly challenging was managing the multiple screens for the student and the data-entry interface while still maintaining a conversational flow with the student, listening to the response, reprompting (if necessary), scoring, entering the score, and managing the voice capture tools. For more information, see Table 11 in Section 6.1 and Tables 14 through 16 in Section 6.3.

##### Proposed Actions:

1. Revise the Speaking-administration model such that test examiners are able to administer the Speaking test to the student utilizing just the student’s testing device and a Speaking *DFA*. In this model, test examiners would enter the student’s Speaking scores in the DEI immediately after the administration of Speaking, alleviating the need to coordinate the student’s screen and the DEI simultaneously. For example, provide mock score sheets within the Speaking *DFA* that would allow test examiners to still score in the moment and keep track of the scores that would be entered in the DEI immediately after the administration.
2. Update the *DFAs* with specific directions for the test examiner to sit face-to-face with the student or at a 90-degree angle with the student while viewing the student testing interface between them.
3. Update the *DFAs* with directions for the test examiner to begin the audio recording of Speaking responses before the test examiner asks the questions. Thus, the student would be able to provide a response immediately without waiting for the test examiner to begin the recording.

#### Recommendation 8

Improve the organization of the *DFAs* so that the test examiners’ materials for administration are easier to use.

##### Context:

Findings from the Usability Pilot indicated that test examiners had difficulties using and navigating the *DFAs* (see Table 16). Test examiners expressed a desire for increased clarity and improved organization to support their test administration. For more information, see Section 6.1 and Tables 14 and 15 in Section 6.3.

##### Proposed Actions:

1. Put task-type administration directions within the domain where they are used instead of in tables at the beginning of the *DFAs*.
2. For Speaking, ensure that there is complete consistency between what is on the screen and the language in the *DFAs*. Add scripting in *DFAs* for transitions between task types and audio recording practice. Make the layout of the Speaking *DFA* very close to the previous version of the *Examiner’s Manual*, where scripts and instructions on prompting, pointing, and when to push audio record are all on the same page and familiar to experienced test examiners.

#### Recommendation 9

Enhance training for test examiners.

##### Context:

The Usability Pilot was the first effort to train test examiners to administer the computer-based ELPAC. Some of the participating test examiners did not have any previous experience administering a computer-based assessment (e.g., CAASPP). Because all training materials were still in development (e.g., User Acceptance Testing environment), it was not possible to conduct a comprehensive training. Accordingly, test examiners were not fully trained to on all of the administrative considerations to support the range of students taking the computer-based ELPAC. For more information, see Section 6.1, Section 6.2, and Tables 14 through 16 in Section 6.3.

##### Proposed Actions:

1. Ensure that the fully developed test examiner training as planned is developed and delivered to test examiners in advance of the field test and the operational launch. Based on the experience of the Usability Pilot, make any adjustments to the plans for execution of this training that may improve its effectiveness while remaining within the scope of the current contract.
2. Consider having the current in-person trainings by the state focus on new computer-based technology and, in particular, on the one-on-one administrations of kindergarten to grade two and Speaking.
3. Utilize the training tests for kindergarten to grade two and Speaking (all grade levels and grade-spans) to model administration and allow participants to practice computer-based administration during the in-person trainings.
4. During in-person trainings, demonstrate exemplary administration models and allow time for participants to collaborate with others to discuss new training models (i.e., utilizing a coaching approach with guided practice before test examiners administer to students).

## References

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

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. (2018). *ELPAC practice tests*. Sacramento, CA: California Department of Education. Retrieved from <https://www.elpac.org/resources/practicetests/>

California Department of Education. (2018). *Matrix One: Universal tools, designated supports, and accommodations for the California Assessment of Student Performance and Progress system.* Sacramento, CA: California Department of Education. Retrieved from ~~https://www.cde.ca.gov/ta/tg/ai/documents/caasppmatrixone0918.docx~~ (This link is no longer active)

California Department of Education. (2019). *English Language Proficiency Assessments for California accessibility resources for operational testing.* Sacramento, CA: California Department of Education.

Educational Testing Service (2019, August). *Addendum: English language proficiency assessments for California usability pilot report*. (Unpublished manuscript). Sacramento, CA: California Department of Education.

Educational Testing Service (2019, March). *Proposed high-level test design for the transition to computer-based ELPAC.* (Unpublished manuscript). Sacramento, CA: California Department of Education.

Educational Testing Service (2019, March). *Specifications for conversion of ELPAC task types for computer-based delivery.* (Unpublished manuscript). Sacramento, CA: California Department of Education.

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.

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.

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~~ (This link is no longer active)

U.S. Department of Education (2017). Resource guide: Accountability for English learners under the ESEA: Non-regulatory guidance. Retrieved from ~~https://www2.ed.gov/programs/sfgp/eseatitleiiiresourceaccountelsguide.pdf~~ (This link is no longer active)

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.

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.

Wolf, M. K., Guzman-Orth, D., & Wain, J. (2015). *Investigating the usability of technology-enhanced assessment items during the ELPA21 development process: ELPA21 cognitive lab study report. Final Deliverable to the English Language Proficiency Assessment for the 21st Century consortium.* Retrieved from ~~https://www.elpa21.org/resources/recdoSRr39KO7OvFC/~~ (This link is no longer active)

Wolf, M. K., Lopez, A., Oh, S, & Tsutagawa, F. (2017). *Comparing the performance of young English language learners and native English speakers on speaking assessment tasks*. In M. K. Wolf & Y.G. Butler (Eds.), English language proficiency assessments for young learners (pp. 171–90). New York, NY: Routledge.

## Accessibility Information

### Alternate Text for Figure 1

Figure 1. A screenshot of a sample Listening item from the paper-based ELPAC. There is an image of a girl sitting next to an adult and is talking. Under the image, there is a question displayed "What does Nadia love to do?" along with three response options: (a) play basketball, (b) read books, and (c) draw pictures.

### Alternate Text for Figure 2

Figure 2. A screenshot of a sample Listening item from the computer-based ELPAC . This is an example of an item within the listening domain. On the left side is an image of a child who is sitting on the bed next to an adult and appears to be talking. Underneath the image, there is a play/pause toggle button that is shown. On the right side, there is a question displayed "What does Nadia love to do?" with three response options: (a) play basketball, (b) read books, and (c) draw pictures. There is a play button above the question to play the audio track in which a student can listen to the question and the response options.

### Alternate Text for Figure 3

Figure 3. A screenshot of a sample Speaking item from the paper-based ELPAC. It includes an image of four students and a teacher engaging in various art activities.

### Alternate Text for Figure 4

Figure 4. A screenshot of a sample Speaking item from the computer-based ELPAC. It includes an image on the left side of four students and a teacher engaging in various art activities. On the right side, there is microphone and play button to record the student's response.

### Alternate Text for Figure 5

Figure 5. A screenshot of a sample Writing item from the paper-based ELPAC. There is a short description of an image at the top of the page. Under the description, there is an image of a student throwing a basketball into a basket, with two other individuals standing behind the student. Underneath the image, there is a caption: "The students are taking turns. The girl just careful throwed the ball. The boy is looking at the ball. The ball is in the air." Underneath this caption, there is direction for the student: "Look at this sentence. The students are taking turns. Rewrite this sentence with more details." Under this description, there are two lines for the student to rewrite this sentence with more details.

### Alternate Text for Figure 6

Figure 6. A screenshot of a sample Writing item from the computer-based ELPAC. There is a short description of an image on the left side of the page. Under the description, there is an image of a student throwing a basketball into a basket, with two other individuals standing behind the student. Underneath the image, there is a caption: "The students are taking turns. The girl just careful throwed the ball. The boy is looking at the ball. The ball is in the air." On the right side, there is direction for the student: "Look at the sentence below. In the box, rewrite the sentence so that it contains more details about the students. The students are taking turns." Under this description, there is a box for the student to type their response. On top of this box, there is a tool bar that include formatting features.