California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-imb-adad-aug20item02

# **MEMORANDUM**

**DATE:**  August 10, 2020

**TO:** MEMBERS, State Board of Education

**FROM:**  TONY THURMOND, State Superintendent of Public Instruction

**SUBJECT:**  Update on the Results of the Mode Comparability Study and a Panel Review for the Transition to Computer-based English Language Proficiency Assessments for California.

## Summary of Key Issues

This California State Board of Education (SBE) Information Memorandum provides a summary of two studies related to the transition of the English Language Proficiency Assessments for California (ELPAC) from a paper–pencil test to a computer-based assessment. Refer to Attachment 1 for *A Study of Mode Comparability for the Transition to a Computer-based English Language Proficiency Assessments for California: Executive Summary* and to Attachment 2 for *A Panel Review of the Transition to Computer-based English Language Proficiency Assessments for California.*

## Background

The ELPAC is the state-administered test developed to address the requirement by the state of California for: 1) determining whether a student is an English learner (EL) or 2) annually determining students’ progress in English language proficiency (ELP)**.** In November 2018, the SBE approved Amendment 6 of the California Assessment of Student Performance and Progress (CAASPP) contract, which combined the CAASPP with the ELPAC into the same contract and allowed for the transitioning of the paper-based ELPAC to a computer-based assessment. In October 2019, the computer-based Summative and Initial ELPAC were field-tested. In an effort to carefully examine the potential impact of this change in format, two studies were conducted: 1) an empirical mode comparability study, which evaluated the field test data and, 2) a qualitative study conducted in late January 2020, in which a group of educators convened to review and analyze ELPAC items in each mode.

The purpose of the mode comparability study was to examine the mode differences between paper–pencil testing and computer-based assessment performance using the data from the October 2019 field test. This quantitative analysis examined students’ Oral (Listening and Speaking domains) and Written (Reading and Writing domains) composite scores to identify significant differences in students’ performance on the basis of test delivery mode. The design of the first study randomly assigned two groups of students from each grade or grade span to a paper–pencil test and a computer-based administration of the same assessment. At the item level, the researchers examined how the computer-based test preserved the substantive meaning of the reported score scales developed from the paper–pencil test. Refer to Attachment 1 for the executive summary of this study. At the end of August, a full report of the quantitative analysis, including a summary of the sampling design, methodology, and results of the psychometric analyses, will be posted on the CDE ELPAC web page at <https://www.cde.ca.gov/ta/tg/ep/>, in the Technical Documents section.

## Additional qualitative analyses were conducted during the transition review panel meeting in order to collect educator feedback about the comparability of items across test administration modes. The focus of educator participation was on experience in working with English learner students, and previous participation in the ELPAC standard setting workshops. Refer to Attachment 2 for further details.

## Findings

Results of the mode comparability study show that, for most grades or grade spans, the students performed similarly on the paper–pencil test and the computer-based assessment. The researchers used three methods of comparison—single group, equivalent group, and hierarchical linear modeling (HLM) to identify mode differences on each of the two composite scores, Oral and Written, and then examined at both the item and test-score level differences. There were only a few cases that showed significant mode differences at the test-score level (see table 1), and these cases showed no consistent pattern favoring one mode over the other. For example, in the equivalent group approach, kindergarten students showed a higher written composite score on the computer-based assessment than on the paper–pencil test, while students in grade spans three through five, six through eight, and nine and ten had differences favoring the paper–pencil test.

Table 1. Summary of Mode Comparability Results

| **Grade or Grade Span** | **Single Group of Oral Score** | **Single Group of Written Score** | **Equivalent Group of Oral Score** | **Equivalent Group of Written Score** |  **HLM of Oral Score** | **HLM of Written Score** |
| --- | --- | --- | --- | --- | --- | --- |
| K | N/A | N/A | Not significant | Significant | Not significant | Not significant |
| 1 | N/A | N/A | Significant | Not significant | Significant | Not significant |
| 2 | N/A | N/A | Not significant | Not significant | Not significant | Not significant |
| 3–5 | N/A | N/A | Significant | Significant | Significant | Not significant |
| 6–8 | N/A | Significant | Not significant | Significant | Not significant | Not significant |
| 9–10 | N/A | Significant | Not significant | Significant | Not significant | Significant |
| 11–12 | Not significant | Not significant | Not significant | Not significant | Not significant | Not significant |

At the item level, only a small number of items (5 out of 266) were identified as performing differently across administration modes based on differential item functioning analyses. Results show strong evidence that the links created between the two modes preserve the substantive meaning of the established reporting score scale. Thus, there is enough evidence to support reliable and valid score interpretations of computer-based assessments for future operational administrations. The reporting score scale that was developed for the paper–pencil test is consistent with the score scale for the computer-based assessment; thus, the change in the mode does not impact the interpretation of results in future ELPAC administrations.

Qualitative results from the transition review panel meeting show that the educators agreed that when evaluating pairs of items presented on the paper–pencil and computer-based assessments, the items measured the same language skills for all grade and grade span tests. The panel of educators had experience reviewing ELPAC items, as they all had participated in the standard setting for ELPAC. They discussed what the items measure and found that across modes, the same skills were measured in the Reading, Listening, Speaking, and Writing domains. They noted that although the paper–pencil test has some advantages with respect to students’ interactions with the items (e.g., annotating text or writing notes on the Test Booklet), the computer-based assessment is better at presenting the stimulus, items, and instructions, especially in showing color images and in standardizing the listening stimulus. The educators concluded that the language skills that are measured or assessed by the items are the same on the paper–pencil test and the computer-based assessment.

In summary, the quantitative results show no change in the meaning of student scores on the computer-based assessment compared to the paper–pencil test. The transition review panel meeting showed some of the nuances in the qualitative data that give further support to the quantitative results of the mode comparability study.

## Attachment(s)

* Attachment 1: *A Study of Mode Comparability for the Transition to a Computer-based English Language Proficiency Assessments for California: Executive Summary* (5 Pages).
* Attachment 2: *A Panel Review of the Transition to Computer-based English Language Proficiency Assessments for California* (14 Pages).



**California Department of Education Assessment Development & Administration Division**



**A Study of Mode Comparability for the Transition to the Computer‑based English Language Proficiency Assessments for California: Executive Summary**

**Submitted June 29, 2020**

**Educational Testing Service**



**Contract No. CN140284**

**Table of Contents**

[**Executive Summary 3**](#_Toc43791153)

[1.1. Study Goals 3](#_Toc43791154)

[1.2. Methods 3](#_Toc43791155)

[1.3. Analyses 4](#_Toc43791156)

[1.3.1. Mode Differences for Items 4](#_Toc43791157)

[1.3.2. Mode Differences for Test Scores 4](#_Toc43791158)

[1.3.3. Linking Computer-based Scores 5](#_Toc43791159)

[1.4. Conclusion 5](#_Toc43791160)

[1.5. References 5](#_Toc43791161)

**Executive Summary**

The English Language Proficiency Assessments for California (ELPAC) is the state-administered test developed to address the requirement by the state of California for: 1) determining whether a student is an English learner (EL), or 2) annually determining students’ progress in English language proficiency (ELP). It is given to students whose primary language is a language other than English based on results of a home language survey.

In November 2018, the California State Board of Education (SBE) approved Amendment 6 of the California Assessment of Student Performance and Progress contract, which, among other things, included the transition of the current paper–pencil ELPAC to a computer-based assessment. As part of the transition to the operational computer-based ELPAC administration, Educational Testing Service (ETS) field-tested the ELPAC items in an online environment between October 1 and November 8, 2019.

**1.1. Study Goals**

ETS conducted analyses to support the following goals of the Summative ELPAC mode comparability study:

* Examine differences between paper–pencil testing and computer-based assessment performance at both item and test-score levels
* Establish links that preserve the substantive meaning of the reported score scale and allow for the valid comparison of ELPAC paper–pencil scores and computer-based assessment scores

**1.2. Methods**

The study included kindergarten, grade one, grade two, and the following grade spans: three through five, six through eight, nine and ten, and eleven and twelve.

The ELPAC includes two composites, one for oral language skills and one for written language skills. The oral composite includes Listening and Speaking. The written composite includes Reading and Writing. To examine performance differences between the paper–pencil test and the computer-based assessment, two groups of students from each grade or grade span participated with sample sizes ranging from 93 to 1,159 (see table 1). Each group responded to one paper–pencil and one computer-based version of the same composite using a counter-balanced design to control for the possible effect on order of administration.

In addition, a computer-based form was used to establish links between paper–pencil scores and computer-based assessment scores for each grade or grade span. A third group of students responded to each of these forms, which contained both oral composite skills and written composite skills.

The number of students in each of these three groups are provided, by grade and grade span, in table 1.

**Table 1. Number of Students by Study Group**

|  |  |  |  |
| --- | --- | --- | --- |
| Grade or Grade Span | Oral Comparability | Written Comparability | Linking |
| Kindergarten | 176 | 700 | 993 |
| 1 | 361 | 644 | 751 |
| 2 | 348 | 581 | 909 |
| 3–5 | 612 | 1,159 | 1,642 |
| 6–8 | 522 | 1,128 | 1,303 |
| 9–10 | 272 | 439 | 705 |
| 11–12 | 93 | 100 | 738 |

**1.3. Analyses**

The next subsections provide an overview of the types of analyses conducted.

**1.3.1. Mode Differences for Items**

Researchers identified 5 items out of 266—1.9 percent—as performing differently across administration modes. Content experts reviewed these items and identified no clear patterns as to why these items might have performed differently. These five items are viewed as isolated cases.

**1.3.2. Mode Differences for Test Scores**

Evidence of mode differences at the test-score level were investigated using *single-group* and *equivalent-groups* designs. The single-group design involved comparing paper–pencil and computer-based results from a group of students taking both versions of the assessment. The equivalent-groups approach involved comparing results for two groups: one administered the paper-based test and one administered the computer-based test.

With seven grades and grade spans and two composites (oral skills or written skills), up to 14 comparisons were performed for each design.

For the single-group design, the analyses make assumptions about the data that are needed for the results to be meaningful; these were not met for all comparisons. Because of this, single-group analyses could be performed for 4 of the 14 comparisons. Of those four, ETS observed differences for two that were of statistical significance but were not large enough to be of practical significance.

ETS found that 6 of the 14 comparisons were statistically significant in analyses based on the equivalent-groups approach. However, the mean differences were not all in the same direction. Evidence of higher performance for the computer-based group was observed for one of these comparisons (written composite skills in kindergarten), while the remaining five comparisons favored students in the paper–pencil test groups (oral composite skills in grade one and grade span three through five; and written composite skills in grade spans three through five, six through eight, and nine and ten). These differences would all be described as “small” based on Cohen’s (1988) often-used classifications of effect size. These results suggest modest differences that can be bridged by standard scale linking methods.

As a follow-up to the equivalent-groups analyses, hierarchical linear modeling was used (Bryk & Raudenbush, 2002). This approach controlled for differences among schools. ETS found that 3 of the 14 comparisons were statistically significant: oral composite skills in grade one and grade span three through five; and written composite skills in grade span nine and ten. Each favored students in the paper–pencil test group.

**1.3.3. Linking Computer-based Scores**

The final stage of the mode comparability study linked scores from the computer-based Summative ELPAC field test to the reporting scale for the operational paper–pencil Summative ELPAC.

This effort succeeded because careful consideration was given to the evaluation and treatment of items appearing in both computer-based and paper–pencil tests. During the scaling linking process, performances were compared for items appearing in both computer-based and paper–pencil forms. Fifteen of those items performed differently between modes and so were treated as unique items within the computer-based and paper–pencil forms. Some were more difficult and others were less difficult in computer-based forms. When a single item proved to be more difficult online, the raw-score-to-scale-score conversion table reported a higher scale score for a given raw score. When a single item proved to be less difficult online, the raw-score-to-scale-score conversion table reported a lower scale score for a given raw score.

Finally, after completing the process to link the scores from the computer-based test to the paper–pencil scale, ETS evaluated whether proficiency classifications were the same for both spring 2019 and fall 2019. Some improvement in student performance was observed in the fall 2019 computer-based administration relative to the spring 2019 paper–pencil administration, which was to be expected for students testing after additional instruction. After a psychometric procedure was applied to adjust for the additional instruction, proficiency classifications were deemed reasonable across modes of administration.

**1.4. Conclusion**

The results of this study provide sufficient evidence that the links created between the Summative ELPAC paper–pencil test and the new computer-based Summative ELPAC preserve the meaning of the established reporting score scale and will support reliable and valid score interpretations of computer-based tests for future operational administrations.

**1.5. References**

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York, NY: Routledge Academic.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.



**A Panel Review of the Transition to Computer-based English Language Proficiency Assessments for California**

**Contract #CN140284**

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

**Presented May 29, 2020**



**Educational Testing Service**

**660 Rosedale Road**

**Princeton, NJ 08541**

**Table of Contents**

[**Background 3**](#_Toc42504457)

[**Time and Location 4**](#_Toc42504458)

[**Panelists 4**](#_Toc42504459)

[**Process 4**](#_Toc42504460)

[Premeeting Activities 4](#_Toc42504461)

[Educator Panel Activities 5](#_Toc42504462)

[Discussion and Data Collection 6](#_Toc42504463)

[Staffing, Logistics, and Security 7](#_Toc42504464)

[**Results 7**](#_Toc42504465)

[Group Feedback, Kindergarten Through Grade Two 7](#_Toc42504466)

[Group Feedback, Grade Three Through Grade Twelve 10](#_Toc42504467)

[Additional Panel Commentary 14](#_Toc42504468)

[**Summary 15**](#_Toc42504469)

**Background**

The ELPAC is the state-administered test developed to address the requirement by the state of California for: 1) determining whether a student is an English learner (EL), or 2) annually determining students’ progress in English language proficiency (ELP). The ELPAC is aligned with the [*2012* *California English Language Development Standards: Kindergarten Through Grade Twelve*](https://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf) (2012 ELD Standards) and is composed of two separate assessments: the Initial ELPAC and the Summative ELPAC.

In November 2018, the California State Board of Education (SBE) approved Amendment 6 of the California Assessment of Student Performance and Progress contract, which included the transition of the current ELPAC paper-pencil test (PPT) to a computer-based assessment. As part of the transition to the operational computer-based ELPAC administration, Educational Testing Service (ETS) field-tested the ELPAC items in an online environment between October 1 and November 8, 2019.

The plan for the transition was for the Summative ELPAC to move online for the 2019‒‍2020 administration, beginning February 1, 2020; and the Initial ELPAC to transition for the 2020‒2021 administration, beginning July 1, 2020. The planned administration dates have been impacted by the novel coronavirus 2019. The operational computer-based Summative ELPAC administration may be extended through fall 2020, and the first operational computer-based Initial ELPAC administration will occur August 20, 2020, through June 30, 2021.

During the transition year, it is particularly important to collect validity evidence for the new computer-based ELPAC. This is a best practice in educational assessment. The validity evidence includes statistical analyses, described in *A Study of Mode Comparability for the Transition to a Computer-Based ELPAC: Executive Summary* [Unpublished Report]; and an evaluation of the content and skills measured by the assessment, described in this report.

With this in mind, and at the request of the CDE, ETS conducted an educator review panel meeting. The purpose of the review was to collect input from educators on items in both the PPT and computer-based modes; they were asked to consider whether the items measured the same or different skills across modes. Educators also provided input on the computer-based mode that may help inform future assessment development and item writer workshop instruction for the computer-based ELPAC. This report presents the results and discussions from the educator review panel meeting.

**Time and Location**

A Transition Review Panel meeting took place over four days: January 27–28, 2020, for kindergarten and grades one and two; and from January 30–31, 2020, for grade spans three through five, six through eight, and nine through twelve. The meeting was held at the Sacramento County Office of Education (SCOE) offices, in Mather, California.

**Panelists**

The panels were composed of educators who participated in standard setting for either the Summative or Initial ELPAC. Panelists all had experience teaching students who took or will take the ELPAC, and were representative of the ethnic, cultural and geographic diversity in California. The standard setting panelists were recruited with similar qualifications in mind, as well as for those panelists’ knowledge of the 2012 ELD Standards.

Thirty-six panelists were selected after CDE approval and were assigned to six groups based on the panelists’ experience working with students from kindergarten, grade one, or grade two; or grade spans three through five, six through eight, or nine through twelve. Recruitment of the panelists specifically focused on educators who had participated in the standard setting workshops for the ELPAC. During those workshops, these educators spent three days reviewing the assessment, considering how students would interact with the assessment tasks, and discussing what the ELPAC items measure. Because of that experience, the panelists were particularly well suited to do this additional work of comparing items in both the PPT and computer-based ELPAC.

**Process**

**Premeeting Activities**

Prior to the Transition Review Panel meeting, confirmed panelists were mailed a letter describing the purpose of the study and were asked to complete a premeeting assignment. The letter included the purpose of the assignment and a hyperlink to the computer-based ELPAC training tests. The panelists took the training test online to become familiar with the task types for the grade or grade span they would review at the meeting.

The computer-based ELPAC training test gives students, parents/guardians, families, teachers, administrators, and others an opportunity to become familiar with both the types of test questions on the ELPAC and the new computer-based platform. The premeeting request to take the test and become familiar with the tasks provided the preliminary work necessary for the meeting while limiting the burden on panelists.

**Educator Panel Activities**

At the meeting, facilitators provided the agenda and overall schedule, the goals of the activities, and instructions on each step in the process. The facilitators modeled the process of item review, notetaking, and discussion for the panel prior to the start of independent reviews.

The educators were seated at three tables each day in groups of six. The six groups were assembled, and worked together, based on the educators’ teaching experience, as outlined in table 1.

**Table** **1. Panel Configuration**

| **Date** | **Group 1** | **Group 2**  | **Group 3**  |
| --- | --- | --- | --- |
| January 27–28 | Kindergarten educators | Grade one educators | Grade two educators |
| January 30–31 | Grade span three through five educators | Grade span six through eight educators | Grade span nine through twelve educators |

Each group of six educators was composed of three pairs. Each pair reviewed small sets of test items from the assigned grade or grade span. The number of items in each set differed by domain: Reading sets ranged from 11 to 16 items, Listening sets ranged from 7 to 12 items, Speaking sets ranged from 8 to 11 items, and all Writing sets included 6 items.

The test items were representative of the task types presented in that grade or grade span. They included items with characteristics that appear different across the delivery modes, as well as items that appear similar across the modes.

In the first phase of the process, educator pairs reviewed items in both delivery modes. Each pair was provided one laptop computer and two copies of the paper test. Each educator reviewed the items first in one mode and then in the other, resulting in the pair reviewing both the paper-pencil format and computer-based format. The educators independently made judgments and kept notes during the review. Each pair discussed each set of items in a domain prior to moving to the next domain.

The goal of the review was for each educator to respond to two questions:

* Question 1 (Q1)—The first time you encounter the item: What language skill or skills is the item measuring? That is, what is it that the student needs to know and be able to do to get the item correct?
* Question 2 (Q2)—The second time you encounter the item: What language skill or skills is the item measuring? If different from your answer in Q1, note why in the column “Notes for Discussion.”

A list of skills for all domains was provided to each panelist. A data-collection form was used to collect the panelists’ feedback; a separate form was used for each domain.

The order of review differed by domain for logistical reasons, such as the ease of participants reviewing Reading items independently and the need for them to review Speaking items on the laptop in pairs. For the Reading domain, for all grades, one educator in each pair encountered a set of items in the paper-based format first while the other educator encountered that set of items in the computer-based format first. For the Listening and Speaking domains, for all grades, the pair worked independently on the paper-based items first and then reviewed the computer-based items on the laptop. For the Writing domain, grade spans three through five, six through eight, and nine through twelve were implemented in the same way as the Reading domain items. However, for the lower grades, no items were reviewed because Writing items continue to be administered in a paper-based mode in kindergarten and grades one and two. Prior to the start of the review for each domain, the facilitator provided examples of differences across modes, such as that in the Listening domain, students hear the items “paced” in paper-pencil format while the students have control of the pace in the computer-based mode.

The instructions to the educators were as follows:

1. Respond to all questions in this domain and answer Q1 after taking the item in the first format you encounter (paper-pencil or computer-based).
2. Switch with your partner. Take the items in the other format and answer Q2 for each item in this domain. Make notes to contribute to further discussion.
3. Discuss with your partner your responses to Q1 and Q2 for all items in this domain. Make notes to contribute to further discussion.

The partners were not required to reach consensus; however, the two colleagues were encouraged to actively listen and to share their rationales for the responses provided to the questions.

**Discussion and Data Collection**

After all three pairs of panelists at each table completed steps 1 through 3, the table group discussed the similarities and differences that were noted for the group’s assigned grade or grade span. A notetaker at each table took notes to contribute to later discussion.

Three facilitators observed the group discussions. After all sets in a domain were reviewed and discussed by the three pairs at a table, the facilitator collected the forms, entered the data, and summarized the data for discussion later.

On January 28, a full-group discussion allowed educators from the lower grades (kindergarten and grades one and two) to discuss the task types reviewed by the group and discuss the similarities and differences across the grades. Two ETS notetakers documented the discussion, and those comments are included as part of the results in the [Results](#_Results) section of this report.

A similar discussion took place on January 31, allowing the educators from grade levels three through twelve to discuss the similarities and differences across the upper grades. These results are also in the [Results](#_Results) section of this report. The ELPAC administration differs between the lower and upper grades, and the focus in these two meetings incorporated that difference.

**Staffing, Logistics, and Security**

Dr. Patricia Baron, technical liaison for the California Assessment of Student Performance and Progress and ELPAC at ETS, led the introductory training session and provided general oversight of the review process. In addition, ETS provided three assessment development specialists to help monitor the process and respond to questions about items, assessment processes, and content validity for the duration of the meeting. They also helped collect and summarize data. Representatives of ETS’ Program Management staff and the CDE were present to observe and respond to panelists’ questions about ELPAC administration or policy.

To protect assessment security, groups were provided with numbered materials on the first day at the time of registration and as needed during the three-day process. At the end of the process, ETS staff collected and securely destroyed all confidential materials.

**Results**

The primary focus of the Transition Review Panel meeting was to collect educator responses about the comparability of items across test administration modes. The judgment asked for was, for pairs of items being measured across two modes, what language skill or skills is the item measuring? Panelists’ individual notes were used for a pair-wise discussion and then for a table discussion. A notetaker for each table summarized the discussion in writing and shared those comments during a full panel discussion.

For all grades and grade spans, the results indicate that the panelists’ agreed that the items presented on the PPT and the computer-based test measured the same skills. In addition, during the discussion, educator comments were captured about what differences were noticed and how those differences might impact the cognitive load placed on students. The discussion also included some questions about test administration, providing educators the opportunity to offer some suggestions to the CDE for continuous improvement.

**Group Feedback: Kindergarten Through Grade Two**

The results from the kindergarten, grade one, and grade two panels are presented in table 2 for three domains in the order in which the educators reviewed them: Reading, Listening, and Speaking. The table details some of the panelists’ specific comments in addition to the participants’ judgments about the skills the two testing modes measured.

Note that the Writing domain was excluded from this chart as Writing items for these grade levels continue to be administered in a PPT format.

**Table 2. Kindergarten Through Grade Two Panel Feedback**

| **Domain** | **Kindergarten** | **Grade 1** | **Grade 2** |
| --- | --- | --- | --- |
| Reading | The PPT and computer-based test assess the same Reading skills. In one item, the picture appeared different in the two modes, and educators thought it should be edited to make the prompt in the two modes more similar. | The PPT and computer-based test assess the same Reading skills. | The PPT and computer-based test assess the same Reading skills. |
| Listening | The PPT and computer-based test assess the same Listening skills. Educators found the computer-based test more engaging. The group agreed it was good to hear the audio recording of two students instead of having the test examiner try to read the lines for both speakers while pointing to pictures of two students. | The PPT and computer-based test assess the same Listening skills. Educators like that color images are provided for students.  | The PPT and computer-based test assess the same Listening skills. Educators thought it was appropriate that students will be able to hear the stimulus only once in the operational test; it is a good feature because students are taught to read questions ahead of responding when taking the Smarter Balanced test in grade three. Educators asked that the *Directions for Administration* provide explicit guidance about whether a student can pause the stimulus, answer a question, and then continue playing the stimulus. |

Table 2 *(continuation)*

| **Domain** | **Kindergarten** | **Grade 1** | **Grade 2** |
| --- | --- | --- | --- |
| Speaking | The PPT and computer-based test assess the same Speaking skills. Educators thought the use of color images in the computer-based items may elicit more elaboration. | The PPT and computer-based test assess the same Speaking skills. Educators said using a paper version of the *Examiner’s Manual* to administer the test seemed to elicit more natural conversation than working with the computer during the test.  | The PPT and computer-based test assess the same Speaking skills. Educators commented on the size of images and the need for scrolling; some educators suggested making the image for Talk about a Scene fit the screen to avoid scrolling.  |

The educators working in kindergarten and grades one and two provided feedback indicating that the PPT and computer-based test assess the same skills in the Reading, Listening, and Speaking domains. The educators noted that the computer-based test is better at presenting the items compared to the PPT, especially in showing color images and in standardizing the Listening stimulus.

**Group Feedback: Grade Three Through Grade Twelve**

Table 3 presents the feedback from the panels for grade spans three through five, six through eight, and nine through twelve for the four domains in the order in which the educators reviewed them: Reading, Listening, Speaking, and Writing. The table details some of the panelists’ specific comments in addition to the participants’ judgments about the skills the two testing modes measured.

**Table 3. Grade Spans Three Through Five, Six Through Eight, and Nine Through Twelve Panel Feedback**

| **Domain** | **Grade Span 3–5** | **Grade Span 6–8** | **Grade Span 9–12** |
| --- | --- | --- | --- |
| Reading | The PPT and computer-based test assess the same Reading skills. Educators noted that more than one skill was measured by some items. Some partners disagreed on which was the primary skill. Table discussion resulted in agreement that there was no mode difference in the skill measured. | The PPT and computer-based test assess the same Reading skills. Educators observed that students are taught to read closely and annotate text on paper. Some educators were concerned that these skills did not transfer well to computer-based text. Other educators thought that students could exercise the same skills by writing annotations about the text on blank paper. | The PPT and computer-based test assess the same Reading skills. One educator noted that hovering over the [**Next**] and [**Back**] buttons on the computer-based test changes the test question; others thought that this was similar to “flipping pages” in the PPT. |

Table 3 *(continuation one)*

| **Domain** | **Grade Span 3–5** | **Grade Span 6–8** | **Grade Span 9–12** |
| --- | --- | --- | --- |
| Listening | The PPT and computer-based test assess the same Listening skills. A number of positive comments were made about the standardization in the Listening domain. Educators valued that students could answer questions at the students’ own pace and play the questions and options as much as the students wanted. This was much better than playing the audio aloud for a group administration because some students had trouble keeping pace with the pauses provided between questions. When those students went back to the book to find the information necessary to answer questions, it essentially became a reading test. Test examiners appreciated no longer needing to read the stimuli aloud, because it was tiring to reread the stimuli to students during numerous administrations. | The PPT and computer-based test assess the same Listening skills. One educator noted that when testing Listening, reading ability should not affect Listening domain scores and thought the computer-based test would support that goal.  | The PPT and computer-based test assess the same Listening skills. The group thought that the computer-based assessment is a more effective means of assessing Listening. |

Table 3 *(continuation two)*

| **Domain** | **Grade Span 3–5** | **Grade Span 6–8** | **Grade Span 9–12** |
| --- | --- | --- | --- |
| Speaking | The PPT and computer-based test assess the same Speaking skills.The group noticed that some of the Speaking skills are subsumed by others. The group discussed this and recognized that this reflects that some skills are drawn from the Part I standards and other skills are drawn from the Part II standards. | The PPT and computer-based test assess the same Speaking skills. Educators said that the computer-based test is more visual and more engaging and the pictures will support understanding. Some educators thought that test examiner questions could be on audio, which would provide a consistent voice across the state. Educators thought that the division between graphs and charts was better in the computer-based version than in the PPT. One educator said that it might be difficult for some students to see some images, so it is important to show students how to use the zoom function. | The PPT and computer-based test assess the same Speaking skills. One educator was concerned that student responses might be different because the PPT version lets students take notes directly on graphics in the Test Book, but the computer-based version provides blank paper for notes.  |

Table 3 *(continuation three)*

| **Domain** | **Grade Span 3–5** | **Grade Span 6–8** | **Grade Span 9–12** |
| --- | --- | --- | --- |
| Writing | The PPT and computer-based test assess the same Writing skills. Educators noticed differences in some test examiner wording across modes.In the large group, educators talked about students’ ability to enter Writing responses via keyboard instead of hand writing. Participants had some concerns about students who have not taken Smarter Balanced yet. Additionally, educators suggested it is important to teach keyboarding skills so that students have practice providing written responses on the computer. | The PPT and computer-based test assess the same Writing skills.Educators commented that the tasks are much easier to read on the computer-based test, where there were bullet points and the instructions are more precise.Educators discussed questions that arose about student stamina and writing to the prompt on the computer versus on paper. | The PPT and computer-based test assess the same Writing skills. |

Similar to the lower grades, the educators in the upper grades identified some advantages to the computer-based mode, compared to the PPT, specifically in the presentation of the stimulus, items, and instructions. There were also comments suggesting that the way students interact with the item, such as annotating text or writing notes on the PPT Test Book, is more similar to the way Writing is currently being taught. For all four domains, the educators agreed that the PPT and computer-based test assess the same skills.

**Additional Panel Commentary**

Recommendations from the kindergarten, grade one, and grade two group discussion included comments about the requirement to provide grade two students with one-on-one administrations instead of small-group administrations. Educators were concerned about the resources required to administer the grade two test (i.e., having sufficient time and personnel for a one-on-one test administration). The participants reported that because many students can enter the responses on the computer themselves, this should be reconsidered. However, some educators said that many of the grade two students would not be able to complete the computer-based test, and would prefer the flexibility to decide whether the student should complete the test one-on-one with the test examiner or participate in a small-group administration.

Educators recommended being provided with audio recordings for two Speaking domain task types (Retell a Narrative and Summarize an Academic Presentation) at grades one and two to improve standardization and efficiency. For kindergarten, there was a discussion about whether it would be better to use an audio recording or to continue to have the test examiner use the script. Some felt that those students need the personal interaction with a familiar adult to elicit responses because that is how much of the information is presented in class. Others felt that standardizing the audio would address issues of accents or of examiners who are not engaging. There was also support for allowing the teacher or test examiner to decide whether to use prerecorded audio.

Educators in the panel for grade spans three through five, six through eight, and nine through twelve shared a few ideas about close reading as an instructional strategy. The educators discussed how to teach students to read closely and annotate text when reading on a computer rather than on paper. One teacher pointed out that it is possible to use a blank piece of paper to annotate what is read online. Another commented that the questions on the ELPAC do not require close reading; rather, students are reading to locate information, and students may need to adjust the reading style used when reading for this purpose. One educator said that students tend to take less time on an online test because accessing digital information quickly and briefly is a familiar process, and another suggested that students practice reading and then skimming the text for information.

Some educators shared that it may be necessary to change the way reading is taught. Students now read a large amount of text on cell phones, so this mode is not new to them. For writing, it is important to teach keyboarding skills so that students have practice providing written responses on the computer. An additional recommendation was for students to have a checklist to ensure the requirements of the task are met. Educators commented that students would benefit from the computer-based test by being able to play the directions as needed for each Writing task type, which would make it easier for students who have better listening comprehension skills than reading comprehension skills to understand the directions.

Educators noted that some concerns about students’ lack of familiarity with the technology are similar to those voiced during the transition to the online Smarter Balanced assessments. The panel thought that the skills assessed on the two modes are the same and thought that at grade four and above, students have experience taking an assessment that requires keyboarding skills. The educators agreed that in grade three, students should use the practice tests to gain experience with the interface. In addition, panelists discussed the need for teachers to make scratch paper available prior to the test administration for all students to use during the Writing section.

The panel appreciated the inclusion of science, social studies, and graphs in the ELPAC, stating that this encourages teachers of English learners to collaborate with subject area teachers to ensure that these students are getting the appropriate instruction.

Educators noted that students benefit from the computer-based version by being able to play the directions as needed for each task type. The audio of the prompt “levels the playing field” because students can hear the prompt while reading it. However, there is no overview of all the task types that will be completed. The panelists stated it might be beneficial to have a brief description at the beginning of the task types that the students will perform.

Panelists asked how the ELPAC works on different brands and models of computers, adding that educators should be encouraged to update the computers and use the ELPAC practice and training tests before accessing the operational test. Teachers in the upper grades panel added that it was very helpful to have a workshop that included educators from grade spans three through five, six through eight, and nine through twelve, saying that it was helpful to hear the expectations for students at each of the grade spans.

**Summary**

Across all grade and grade-span panels, educators’ judgments indicated that the tasks on the ELPAC are measuring the same skills in the paper-pencil mode and the computer-based mode of the assessment. Panelists commented that, overall, the computer-based test is aligned to the knowledge and skills being assessed across the range of students who are taking this test. The educators liked the computer-based presentation, stating that the color images are vivid and engaging. In addition, it was noted that the audio files for directions and prompts help level the playing field for students who have better listening comprehension skills than reading comprehension skills.