

# Draft Summative Alternate English Language Proficiency Assessments for California Threshold Score Validation Study Report

**Contract #CN22002**

**Prepared for the California Department of Education by ETS**

**Presented March 3, 2023**



Table of Contents

[Overview and Context 3](#_Toc128986766)

[Description of the Alternate ELPAC 3](#_Toc128986767)

[Standard Setting Process 4](#_Toc128986768)

[Alternate ELPAC Threshold Score Validation Study 4](#_Toc128986769)

[Study Sample 5](#_Toc128986770)

[Contrasting Groups Method 6](#_Toc128986771)

[Analytic Methods 7](#_Toc128986772)

[Results and Interpretation 11](#_Toc128986773)

[Summary 13](#_Toc128986774)

[Conclusion 13](#_Toc128986775)

[References 14](#_Toc128986776)

List of Tables

[Table 1. Contrasting Groups Study Sample 6](#_Toc128986784)

[Table 2. Educator Ratings by Alternate ELPAC Performance Levels (PLs) for Kindergarten 7](#_Toc128986785)

[Table 3. Educator Ratings by Alternate ELPAC Performance Levels for Grade One 7](#_Toc128986786)

[Table 4. Educator Ratings by Alternate ELPAC Performance Levels for Grade Two 8](#_Toc128986787)

[Table 5. Educator Ratings by Alternate ELPAC Performance Levels for Grades Three Through Five 8](#_Toc128986788)

[Table 6. Educator Ratings by Alternate ELPAC Performance Levels for Grades Six Through Eight 8](#_Toc128986789)

[Table 7. Educator Ratings by Alternate ELPAC Performance Levels for Grades Nine and Ten 8](#_Toc128986790)

[Table 8. Educator Ratings by Alternate ELPAC Performance Levels for Grades Eleven and Twelve 9](#_Toc128986791)

[Table 9. Percent Agreement on Classification as EL (Levels 1 and 2) and Fluent English Proficient (Level 3) 9](#_Toc128986792)

## Overview and Context

The purpose of this report is to provide the results of the validation study conducted to evaluate the thresholds for the Summative Alternate English Language Proficiency Assessments for California (ELPAC). Data for the study was collected using an in-test survey that test examiners completed at the beginning of the operational field test administration prior to testing the student. For context, a description of the Alternate ELPAC and a description of the process used to develop threshold scores are included. The main sections of the report include a description of the validation study, study sample, method, data collection tool, and results, as well as a summary of the study.

### Description of the Alternate ELPAC

The Alternate ELPAC is designed to assess the English Language Proficiency (ELP) of English learners (ELs) and potential ELs with the most significant cognitive disabilities, as determined by their individualized education program team. The Alternate ELPAC is aligned to the [2012 California English Language Development Standards](https://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf), which were adopted by the California State Board of Education (SBE), via the English Language Development (ELD) Connectors.

The Alternate ELPAC consists of two assessments: the Initial Alternate ELPAC and the Summative Alternate ELPAC. Their purposes are as follows:

1. The Initial Alternate ELPAC provides information to determine a student’s initial classification as an EL, or as initial fluent English proficient, for students with the most significant cognitive disabilities.
2. The Summative Alternate ELPAC provides information on annual student progress toward ELP and supports decisions on student reclassification as fluent English proficient for students with the most significant cognitive disabilities.

The operational field test administration of the Summative Alternate ELPAC took place from November 1, 2021, through May 31, 2022.

As with other California alternate assessments, the Alternate ELPAC is administered online and is administered one-on-one with an educator serving as test examiner. It is important that this educator is familiar with the student and specific to the Alternate ELPAC, familiar with the student’s preferred communication modes. The test items are categorized as receptive—Listening and Reading—or expressive—Speaking and Writing—allowing for an integrated approach to the four language domains. This important feature of the Alternate ELPAC allows students to use any communication mode they typically use during classroom instruction, including, but not limited to, sign language, pointing, eye gaze, and an Augmentative and Alternative Communication device. Furthermore, the test administration can take place over more than one day.

### Standard Setting Process

In May 2019, the SBE approved the Alternate ELPAC general performance level descriptors (PLDs) as listed in the [*Proposed High-Level Test Design for the Alternate ELPAC*](https://www.cde.ca.gov/ta/tg/ep/documents/proposedhltdaltelpac.pdf). These general, or policy, PLDs convey the degree of a student’s ELP using three levels: Novice English Learner (Level 1), Intermediate English Learner (Level 2), and Fluent English Proficient (Level 3). In fall 2020, range PLDs were reviewed by California educators and approved by the California Department of Education (CDE). The PLDs were used in standard setting to distinguish between the meaning of each performance level at a grade level or grade span.

In February 2020, to develop threshold-score recommendations, ETS conducted a standard setting workshop for the Alternate ELPAC using virtual meeting applications. All items in the Alternate ELPAC item pool were considered in the process of standard setting. The Modified Angoff and Extended Angoff standard setting methods were applied, as appropriate. The standard setting included a diverse group, representative of California educators familiar with instructing this student population, who participated as panelists in the standard setting sessions. These educators were familiar with the ELD Connectors, were engaged in the daily instruction of ELs with the most significant cognitive disabilities and understood and represented the diverse group of students eligible to take the Alternate ELPAC. For each grade level or grade span, the standard setting panel recommended threshold scores to indicate the score that must be earned for a student to reach the beginning (i.e., threshold) of two of the three performance levels (Level 2 and Level 3) for the Alternate ELPAC total score. California educators used the ELD Connectors, the Alternate ELPAC General PLDs (CDE, 2019), and the range PLDs (CDE, 2022).

The SBE adopted the CDE’s recommended threshold scores in May 2022. Following a similar process that was used for the Summative ELPAC Threshold Score Validation Study (CDE, 2018), CDE and ETS collected evidence to evaluate the placement of the thresholds using data from the first operational field test administration of the Alternate ELPAC.

## Alternate English Language Proficiency Assessments for California Threshold Score Validation Study

At the beginning of the test administration, the educator was asked to complete an in-test survey. The survey was delivered online, and the educator completed the survey prior to students taking the test. The focus of this report is on the first question in the survey, the purpose of which was to collect evidence from the educator on what they perceive as the student’s current level of overall ELP. The data was collected for all grade levels and grade spans. All educators received the same question, described in the next section. The overarching goal of the study method was to compare the performance levels of the students as indicated by the assessment to the evaluations made by test examiners in the in-test survey.

The Alternate ELPAC in-test survey was developed by research staff at ETS in consultation with national experts, technical advisors, and the CDE. The various groups provided guidance in terms of the length of the survey and the questions to consider. The goal of the survey was to gather validity evidence to support continuous program improvement. This report focuses on one use of the Alternate ELPAC in-test survey, to gather an external measure of student ELP and provide a concurrent check on the validity of the threshold scores used for ELP classification. Three options were provided to the educator, corresponding to the three general, or policy, PLDs.

The question and options are as follows:

* Based on your interactions with this student **during classroom instruction,** which of the following best characterizes this student’s current level of overall English language proficiency?
1. **High or fluent English proficient—**Students at this level have **sufficient** English language proficiency. They may need **occasional** linguistic support to enable them to access adapted grade-level content in English.
2. **Medium or intermediate English learner—**Students at this level have **moderate** English language proficiency. They may need **frequent** linguistic support to enable them to access adapted grade-level content in English.
3. **Low or novice English learner—**Students at this level have **minimal** English language proficiency. They need **substantial** linguistic support to enable them to access adapted grade-level content in English.

### Study Sample

The results from the survey are based on all tested grade levels and grade spans, from kindergarten through grade span eleven and twelve. All available Alternate ELPAC student scores and performance levels were included in the student data. All educators were required to respond to the in-test survey questions via the test delivery system (TDS). Some questions were to be answered prior to the administration with the student, including the first question, which is the focus of this report. Educators and local educational agencies (LEAs) were provided with access to the survey questions in the web-based *Alternate ELPAC Operational Field Test Manual* and other test administration resources prior to entering the TDS to limit the time an educator sat with a student for test administration.

The study sample includes all students for whom both pieces of evidence were available: the Alternate ELPAC performance level based on the operational field test and the performance level provided by the educator in response to the first in-test survey question. These data allow analyses to compare student classifications using the contrasting groups method, as described below. Table 1 provides the number of students that were evaluated by test examiners in the in-test survey.

Table 1. Contrasting Groups Study Sample

|  |  |
| --- | --- |
| **Grade Level or Grade Span** | **Number of Students** |
| Kindergarten | 1,273 |
| Grade 1 | 1,247 |
| Grade 2 | 1,198 |
| Grade Span 3–5 | 4,161 |
| Grade Span 6–8 | 3,571 |
| Grade Span 9–10 | 1,720 |
| Grade Span 11–12 | 3,228 |

### Contrasting Groups Method

The survey data provided input from educators familiar with the ELD Connectors and with the students in their classrooms. The analysis conducted provided a comparison of the student performance levels based on the threshold scores to performance levels based on the educator judgments. This method is known as the contrasting groups method of standard setting (Zieky, Perie, & Livingston, 2008).

In this method, educators familiar with the ELD Connectors and with the students in their classroom who took the Alternate ELPAC were asked to make judgments about the students’ performance levels based on the approved Alternate ELPAC PLDs. Using data from the operational field test of the Alternate ELPAC, student performance levels based on the threshold scores were compared to performance levels based on the educator judgments. Results from the validation study will allow the CDE to consider information across standard setting methods, as described herein.

Approximately seven months into the school year, educators familiar with students in their classroom were asked to classify students according to the approved Alternate ELPAC general PLDs. The contrasting groups method requires a large number of educator ratings of students from a representative sample of LEAs. It is also desirable to have ratings of students with a wide range of performance and to require a reasonable number of student ratings from each educator.

The judgment of the educators was based on their knowledge and understanding of their own students’ levels of proficiency, relative to the California-approved final PLDs. Note that California-approved PLDs were the starting point for this contrasting groups study, thereby maintaining the meaning of the performance levels from the standard setting studies for consistency and standardization. A statistical analysis was conducted comparing students’ Alternate ELPAC scores to educators’ ratings.

The results of the statistical analyses and educator ratings can be used in concert with other information, including the results from the panel-based standard setting and post–standard setting considerations, such as the impact of threshold scores on the Alternate ELPAC score distributions.

#### Analytic Methods

##### Rater Agreement Analysis

Table 2 through table 8 show the cross-tabulation of the number of students classified in each of the three Alternate ELPAC performance levels based on two methods. Each table presents, for a grade-level or grade-span test, the number of students classified based on the educator ratings and based on the Summative Alternate ELPAC threshold score (Alternate ELPAC performance level). In each table, the number of students classified as the same level by both methods can be found on the diagonal. For example, for kindergarten, 435 students (80%) were classified as Level 1 both by educator rating and by Alternate ELPAC performance level (PL1), 192 students (36%) were classified as Level 2 by both methods, and 52 students (27%) were classified as Level 3 by both methods. Table 9 provides a summary of the agreement by grade level and grade span, when comparing classification as EL (Level 1 and Level 2) to classification as Fluent English Proficient (Level 3). The table shows the percent agreement between educator ratings and student performance levels based on Alternate ELPAC performance level. Figure 1 shows the trend across grade levels and grade spans for Level 3 based on both the educator ratings and Alternate ELPAC thresholds.

Table 2. Educator Ratings by Alternate ELPAC Performance Levels (PLs) for Kindergarten

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 435 | 80 | 262 | 49 | 55 | 28 |
| 2 | 79 | 15 | 192 | 36 | 88 | 45 |
| 3 | 27 | 5 | 83 | 15 | 52 | 27 |
| **Total** | **541** | **100** | **537** | **100** | **195** | **100** |

Table 3. Educator Ratings by Alternate ELPAC Performance Levels for Grade One

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 363 | 76 | 283 | 51 | 47 | 22 |
| 2 | 81 | 17 | 199 | 36 | 85 | 41 |
| 3 | 35 | 7 | 77 | 14 | 77 | 37 |
| **Total** | **479** | **100** | **559** | **100** | **209** | **100** |

Table 4. Educator Ratings by Alternate ELPAC Performance Levels for Grade Two

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 360 | 68 | 193 | 41 | 29 | 14 |
| 2 | 126 | 24 | 191 | 41 | 79 | 39 |
| 3 | 40 | 8 | 86 | 18 | 94 | 47 |
| **Total** | **526** | **100** | **470** | **100** | **202** | **100** |

Table 5. Educator Ratings by Alternate ELPAC Performance Levels for Grades Three Through Five

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 837 | 72 | 695 | 37 | 144 | 13 |
| 2 | 246 | 21 | 796 | 42 | 486 | 44 |
| 3 | 84 | 7 | 398 | 21 | 475 | 43 |
| **Total** | **1167** | **100** | **1889** | **100** | **1105** | **100** |

Table 6. Educator Ratings by Alternate ELPAC Performance Levels for Grades Six Through Eight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 573 | 68 | 464 | 36 | 156 | 11 |
| 2 | 186 | 22 | 532 | 41 | 592 | 41 |
| 3 | 84 | 10 | 301 | 23 | 683 | 48 |
| **Total** | **843** | **100** | **1297** | **100** | **1431** | **100** |

Table 7. Educator Ratings by Alternate ELPAC Performance Levels for Grades Nine and Ten

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 307 | 64 | 192 | 26 | 30 | 6 |
| 2 | 110 | 23 | 281 | 39 | 176 | 34 |
| 3 | 59 | 12 | 253 | 35 | 312 | 60 |
| **Total** | **476** | **100** | **726** | **100** | **518** | **100** |

Table 8. Educator Ratings by Alternate ELPAC Performance Levels for Grades Eleven and Twelve

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Educator Rating Level** | **Alternate ELPAC PL 1** | **Percent Students PL 1** | **Alternate ELPAC PL 2** | **Percent Students PL 2** | **Alternate ELPAC PL 3** | **Percent Students PL 3** |
| 1 | 590 | 65 | 351 | 28 | 54 | 5 |
| 2 | 189 | 21 | 488 | 39 | 284 | 27 |
| 3 | 126 | 14 | 428 | 34 | 718 | 68 |
| **Total** | **905** | **100** | **1267** | **100** | **1056** | **100** |

Table 9 provides a summary of the agreement by grade level and grade span, when comparing classification as EL (Level 1 and Level 2) to classification as Fluent English Proficient (Level 3). The table shows the percent agreement between educator ratings and student performance levels based on Alternate ELPAC performance level.

Table 9. Percent Agreement on Classification as EL (Levels 1 and 2) and Fluent English Proficient (Level 3)

|  |  |
| --- | --- |
| **Grade Level or Grade Span** | **Percent Agreement** |
| Kindergarten | 80.1% |
| 1 | 80.4% |
| 2 | 80.5% |
| 3–5 | 73.3% |
| 6–8 | 68.3% |
| 9–10 | 69.9% |
| 11–12 | 72.4% |
| **Overall** | **73.3%** |

Figure 1 shows the trend across grade levels and grade spans for Level 3 based on both the educator ratings and Alternate ELPAC thresholds. The y-axis displays the percentage of students from zero to 60 and the x-axis shows results from kindergarten through grades eleven and twelve.

Figure 1 Percent of students at Level 3 across grade levels and grade spans

### Results and Interpretation

Results from the rater-agreement analysis of the contrasting groups study indicated some trends by grade level and by performance level. Educator ratings of expected performance levels (referred to as Levels) were compared to the students’ performance levels based on the 2021–22 threshold scores (referred to as PL1, PL2, PL3). The comparison indicated moderate agreement rates, with some differences by grade level and performance level. Agreement was strongest for Alternate ELPAC performance level one (PL1), where between 64–80 percent of educators’ ratings agreed with Alternate ELPAC classification, in kindergarten through grade twelve. This means, for example, where the Alternate ELPAC indicated a PL1 for kindergarten students, 80 percent of educator ratings agreed that Level 1 best characterizes the students’ current level of overall English language proficiency. Compared to the agreement for PL1, the results for PL2 and PL3 are more moderate.

In kindergarten and grade one, educators rated about half of the students as Level 1 (49% and 51% respectively). This indicates that the educators believed that the students are performing at a lower level than the student performance on the Alternate ELPAC would suggest. The pattern continues in PL3. The percent agreement is 27 percent for kindergarten and 37 percent for grade one. Most kindergarten and grade one educators rated students’ current level of overall ELP below the Alternate ELPAC PL3 level.

In grade two, the percent agreement for PL2 is 41 percent; however, educators rated 41 percent of students as Level 1 and only rated 18 percent of students as Level 3. This pattern of agreement is also observed for grade two, grades three through five, and grades six through eight for PL2. Educator ratings trend toward a lower level when there is disagreement. By combining educator ratings of levels 1 and 2 for students classified as PL2 based on Alternate ELPAC performance, it can be observed that most educator ratings across grades two through eight for students classified as PL2 are considered by educators as either Level 1 or Level 2 (77% to 82%).

The pattern of percent agreement for PL3 for grades two through eight suggests slightly less than half of the educator ratings agreed with the Alternate ELPAC PL3 classification. In grade two, 47 percent of educator ratings concurred with the PL3 Alternate ELPAC classification. Percent agreement was similar for grades three through five (43%) and six through eight (48%). The educator ratings that were not in agreement indicated a similar pattern to that observed in kindergarten and grade one. Educators rated students at a lower level than the Alternate ELPAC classification of PL3. Specifically, for students classified as PL3, 39 percent of grade two students, 44 percent of grade three through five students, and 41 percent of grade six through eight students were rated as Level 2.

In the high school grade levels (table 7 and table 8), the pattern of results is slightly different for PL2 and PL3. Thirty-nine percent of students classified on the Alternate ELPAC as PL2 were rated by educators as Level 2. However, in comparison to the lower grade levels, a larger percentage of students were rated at a higher level by educators. Of the students at PL2 in grades nine and ten, 35 percent were rated as Level 3; in grades ten and eleven, 34 percent were rated as level 3. The educator ratings for students who would be classified as PL2 on the Alternate ELPAC suggest the expectation that these students would perform at a higher level than the assessment results indicate. However, educator ratings for almost one third of students classified as PL2 suggest the opposite. In grades nine and ten, 26 percent of students were rated as Level 1; in grades eleven and twelve, 28 percent of students were rated by educators as Level 1. These results may indicate that educators’ judgments for the upper grade level students may be somewhat more lenient when considering performance Level 3 and may deserve further exploration. The percent agreement for students who were classified as PL3 on the Alternate ELPAC was strongest in the high school grade levels. In grades nine and ten, there was 60 percent agreement between the educator ratings and assessment results; in grades eleven and twelve, there was 68 percent agreement. Across all grade levels, the range of agreement for PL3 increased from kindergarten (27%) through grades eleven and twelve (68%), with the exception of grades three through five.

Because of the importance of PL3 in decisions about classification of the proficiency levels for EL students, table 9 presents the percent agreement for students classified as ELs (PL1 and PL2) and students who would be classified as Fluent English Proficient (PL3). This comparison provides some evidence that the educator ratings are slightly more aligned with the alternate ELPAC classifications in kindergarten through grade two, relative to the upper grade levels. The general trend in percent agreement between the two sources for classification, comparing EL versus Fluent English Proficient students, is a decrease in agreement from kindergarten (80.1%) to grades eleven and twelve (72.4%), with some small fluctuations across the seven grade levels and grade spans. Overall, educators and the assessment agreed 73.3 percent of the time on the classification of a student as an EL or Fluent English Proficient.

Figure 1 shows the trend across grade levels and grade spans for student classification as PL3 based on the Alternate ELPAC thresholds and as Level 3 based on educator ratings. There is an increase in the percent of students in PL3 in grade span six through eight, however generally both educator ratings and Alternate ELPAC performance levels increase from kindergarten to high school. This provides further support for the SBE approved Alternate ELPAC thresholds, particularly for students classified as Fluent English Proficient.

## Summary

The standard setting workshop conducted in February 2020 resulted in recommendations for threshold scores, which resulted in performance levels for students who took the 2021–22 Summative Alternate ELPAC operational field test. Subsequently, a threshold score validation study was conducted, using a contrasting groups standard setting method. Educators considered the performance of EL and potential EL students with the most significant cognitive disabilities in the classroom and provided expected performance levels for their students. A comparison of these expected performance levels with the levels for those students based on the threshold scores was conducted. The results indicate some patterns by performance level and some patterns related to grade level. The strongest agreements across grade levels and grade spans were found in PL1, defined as Low or novice English learner—Students at this level have minimal English language proficiency. They need substantial linguistic support to enable them to access adapted grade-level content in English. There was also a pattern observed for PL3. As the grade level of students increased, educator ratings were more aligned with the Alternate ELPAC PL3 classification.

## Conclusion

The purpose of conducting the threshold score validation study was to evaluate the degree to which the threshold scores and performance levels of the Summative Alternate ELPAC distinguish between levels of students’ ELP, based on educator ratings from a multistep process (standard setting in February 2020 and the validation study). Implementing a multistep process offers increased confidence in decisions using threshold scores based on Alternate ELPAC results. Because the classification of the proficiency levels for EL students with the most significant cognitive disabilities entails relatively high-stakes decisions for individual students’ academic paths, school program funding, and resource plans, it is crucial to validate the threshold scores for each proficiency level to the extent possible. Across all grade levels and grade spans, the percent agreement in classification of students as EL or FEP ranged from 72.4 percent to 80.5 percent and the overall percent agreement was 73.3 percent. As well, a trend was observed across methods, which was that the percent of students classified as FEP increased as student grade level increased (kindergarten through high school).

## References

California Department of Education. (2018). *ELPAC threshold score validation study report.* [Unpublished report] Sacramento, CA*:* California Department of Education*.*

California Department of Education. (2019). *NGSS for California public schools, K-12.* Sacramento, CA*:* California Department of Education*.*

California Department of Education. (2022). *California standard setting technical report for the Alternate English Language Proficiency Assessments for California*. [Unpublished report] Sacramento, CA: California Department of Education.

Zieky, M.J, Perie, M., & Livingston, S.A. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.