

**Research Evidence to Support Alternate English Language Proficiency Assessments for
California (ELPAC) in Reclassification Decisions**

Kelly B. Bacher¹, Patricia Baron², Molly Faulkner-Bond³, Shuqin Guo¹, Alesha Moreno Ramirez¹,

Mao Vang¹

¹California Department of Education, ²ETS, ³WestEd

Abstract

State and federal laws require that all students whose primary language is a language other than English be assessed for English language proficiency (ELP). The legal basis for requiring ELP testing is that all students have the right to an equal and appropriate education, and any English language limitations left unidentified and/or unaddressed could preclude a student from accessing that right. Because standardized assessment of English learner (EL) students is a required component for evaluating student progress toward English proficiency, it is critical to establish the validity of such assessments. Evidence of external validity can be collected by comparing its results to other measures, such as teacher evaluations and results of other assessments. This report summarizes the results from two studies conducted to evaluate a large-scale ELP assessment designed for students with the most significant cognitive disabilities. Specifically, we present a series of analyses to evaluate and validate the state's current recommended thresholds for reclassifying EL students with the most significant cognitive disabilities. These dually-identified students represent an understudied group in the assessment literature.

Keywords: assessment, English learners, students with disabilities, reclassification

Study 1 was conducted by ETS with collaboration and oversight by CDE under contract CN220002.

Study 2 was conducted by CDE with training, coaching, and technical support from the Regional Education Laboratory West (REL West) at WestEd. CDE wishes to express their appreciation for the cross-organizational partnerships that supported this important work.

Research Evidence to Support Alternate English Language Proficiency Assessments for California (ELPAC) in Reclassification Decisions

English learners with the most significant cognitive disabilities in California

In the state of California, over 2.3 million K–12 students speak a language other than English at home and over one million of these students are classified as English learner (EL) students (DataQuest, 2023). As in the general population, some ELs (17.5%) have learning, physical, or cognitive disabilities that require an individualized education program (IEP) or a Section 504 plan. A smaller percentage (no more than 1%, under the regulations of the Every Student Succeeds Act (ESSA)), have significant disabilities that preclude them from participating in standardized assessments, even with accommodations. In the 2021–22 school year, California had approximately 20,000 such students.

State and federal laws require that all students whose primary language is a language other than English be assessed for English language proficiency (ELP). The legal basis for requiring ELP testing is that all students have the right to an equal and appropriate education, and any English language limitations left unidentified and/or unaddressed could preclude a student from accessing that right. Local educational agencies (LEAs) must ensure that English learners are making progress toward English proficiency, and remedy any academic deficits incurred while learning English, within a reasonable amount of time. In California, this legal basis comes from *California Code of Regulations*, Title 5 (5 CCR), Section 11518; *California Education Code (EC)* sections 313 and 60810; and Titles I and III of the federal Every Student Succeeds Act, which reauthorized the Elementary and Secondary Education Act.

This report summarizes efforts made by the California Department of Education (CDE) to develop an ELP assessment and associated policies for the approximately 20,000 EL students in the state with the most significant cognitive disabilities. As noted above, these students must participate in annual assessments of ELP. The results from this assessment may be used for multiple purposes, the

most high-stakes of which is reclassification, or the decision to exit an EL student with significant cognitive disabilities from EL status and its associated linguistic services and supports. Although students with disabilities are not expected to outgrow their disabilities over time, it is a goal for all EL students to develop proficiency in English and exit out of this status. In California, these decisions are made on the basis of four criteria, one of which is a statewide standardized assessment of ELP (*EC* Section 313(f)). In the next section, we describe the CDE's efforts to develop such an assessment for EL students with the most significant cognitive disabilities, and to set a statewide threshold for reclassification on said assessment. Specifically, we present a series of analyses to evaluate and validate the state's current recommended thresholds for reclassifying EL students with the most significant cognitive disabilities. These dually-identified students represent an understudied group in the assessment literature.

Development of the Alternate ELPAC

Between 2019 and 2021, the CDE and its testing contractor, ETS, developed the Summative Alternate English Language Proficiency Assessments for California (Summative Alternate ELPAC) to be used as a statewide, standardized assessment of ELP for EL students with the most significant cognitive disabilities. The Alternate ELPAC is aligned to the 2012 California English Language Development Standards, which were adopted by the California State Board of Education (SBE), via the English Language Development (ELD) Connectors.

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

In May 2019, the SBE approved the Alternate ELPAC general performance level descriptors (PLDs). 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 of 2020, range PLDs were reviewed by California educators and approved by the CDE. The PLDs were used in standard setting to distinguish between the meaning of each performance level at a given grade level or grade span.

In February 2020, ETS conducted a standard setting workshop for the Alternate ELPAC using virtual meeting applications to develop threshold-score recommendations. 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.

The CDE provided interim guidance to LEAs recommending that EL students with the most significant cognitive disabilities be eligible for reclassification when they earn an Overall Performance Level (PL) of 3 (out of three possible levels) on the Summative Alternate ELPAC. Following a similar process that was used for the Summative ELPAC Threshold Score Validation Study (CDE, 2018), we collected evidence to evaluate the placement of the thresholds as it pertains to the interim guidance for meeting Criterion 1 for reclassification using data from the first operational field test administration of the Alternate ELPAC.

About This Report

The purpose of this report is to provide the results of two validation studies conducted to evaluate the thresholds for the Summative Alternate ELPAC. The purpose of the first study was to evaluate the degree to which threshold scores and performance levels of the Summative Alternate ELPAC agree with levels of students' English proficiency based on the test examiner ratings. The second study examined the relationship between student performance on the Summative Alternate ELPAC relative to student performance on the California Alternate Assessment for English language arts/literacy (CAA for ELA). Altogether, the goal of these studies was to evaluate whether to maintain the current threshold score of the Summative Alternate ELPAC Overall PL 3 (Fluent English Proficient) for Criterion 1 (assessment of English language proficiency) of the four required criteria in reclassification decisions for EL students with the most significant cognitive disabilities. The details of each study are presented in turn, followed by a concluding discussion of the research.

Study 1: Threshold Score Validation

Data Sources and Sample

Data for the study included two types of evidence: student performance data on the Alternate ELPAC, and data from educators collected prior to testing students. 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 educators about students' ELP, collected in the in-test survey.

Evidence based on student performance on Alternate ELPAC included all available student scores and performance levels in the student data. Data collected from educators utilized an in-test survey that test examiners completed at the beginning of the operational field test administration. The data from the survey was based on responses to the first question; the purpose was to collect evidence from the educator on what they perceive as the student's current level of overall ELP. The results are based on all tested grade levels and grade spans, from kindergarten through grade span eleven and twelve. All educators were required to respond to the in-test survey questions, and all educators received the same question, described in the next section.

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 full survey was to gather validity evidence to support continuous program improvement. The first question in the survey provided an external measure of student ELP and 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?
 - A. **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.
 - B. **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.

- C. **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.

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

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 et al., 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 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

those based on the educator judgments. Results from the validation study allowed 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 judgments of the educators were based on their knowledge and understanding of their own students' levels of proficiency, relative to the California-approved final PLDs. Note that the 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.

Results

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 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

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
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 11–12.

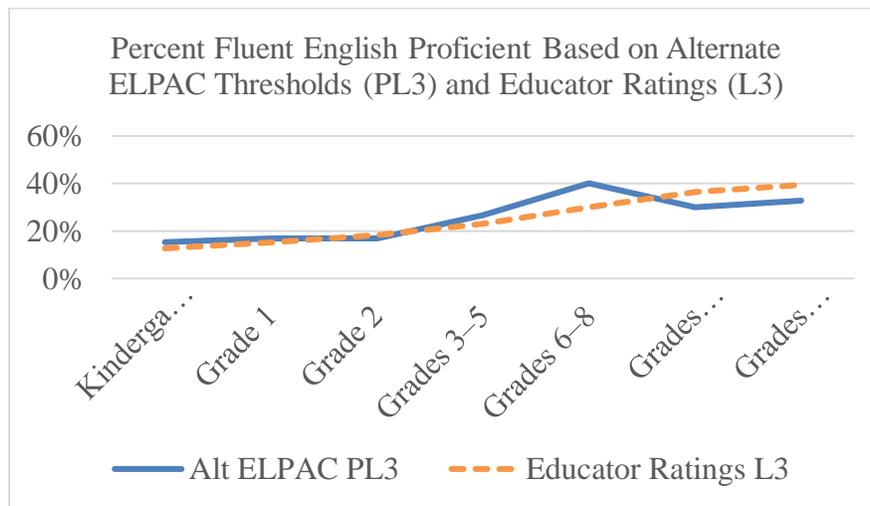


Figure 1. Percent of students at Level 3 across grade levels and grade spans

Caption: The x-axis displays the grade spans on the Alternate ELPAC – Kindergarten, Grade 1, Grade 2, Grades 3–5, Grades 6–8, Grades 9–10, and Grades 11–12. The y-axis displays the percent of students classified at Level 3 by Alternate ELPAC (indicated by a solid blue line) or by educator ratings (indicated by the dashed green line). The educator ratings gradually increase from left to right, with a low of approximately 12% in kindergarten to a high of approximately 40% in Grades 11–12.

The Alternate ELPAC line also generally increases from left to right, starting at approximately 16% in kindergarten, but peaks in Grades 6–8 at approximately 40%.

Discussion

The results of the contrasting groups study indicated some trends by grade level and by performance level. The study compared educator ratings of expected performance levels (Levels) to the students' performance levels (PL1, PL2, PL3) based on the 2021–22 threshold scores. Moderate agreement rates were observed, with variations across grade and performance levels. The strongest agreement occurred for Alternate ELPAC performance level one (PL1), with 64–80 percent alignment between educator ratings and PL1 classification for kindergarten through grade twelve. For instance, when 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.

Contrasting PL1, PL2 and PL3 showed more moderate agreement. Kindergarten and grade one educators frequently rated students at Level 1 (49% and 51% respectively), implying lower performance than Alternate ELPAC suggested. This pattern persisted in PL3, with agreement at 27% for kindergarten and 37% for grade one. A trend for grades 2–8 suggests that educators leaned toward lower levels in disagreements. Combining levels 1 and 2 for PL2 students indicated most educator ratings across grades 2–8 were considered Level 1 or 2 (77% to 82%).

For PL3 in grades 2–8, educator agreement with Alternate ELPAC was below 50%. In grade two, 47% agreed with PL3 classification, similar to grades 3–5 (43%) and 6–8 (48%). The trend was that educators rated PL3 students lower than Alternate ELPAC's PL3. In high school, PL2 results differed: 39% aligned with educator-rated Level 2, and a larger portion rated higher. PL3 agreement was strongest in high school, with 60% in grades 9–10 and 68% in 11–12.

A comparison of EL (PL1–2) and Fluent English Proficient (PL3) classifications showed better alignment in lower grades, decreasing from kindergarten (80.1%) to grades 11–12 (72.4%), though fluctuating. Overall agreement on EL vs. Fluent English Proficient classification was 73.3%.

Figure 1 depicted trends in PL3 student classification based on Alternate ELPAC thresholds and educator ratings across grade levels. Generally, both educator ratings and Alternate ELPAC performance levels increased from kindergarten to high school, supporting the approved thresholds, especially for Fluent English Proficient students.

In 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.

Study 2: Supplemental Empirical Analyses

Data Sources and Sample

Data for this study are based on the results of the Alternate ELPAC operational field test, which took place during the 2021–22 school year. The CDE gathered data from the Alternate ELPAC section

of the 2021–22 ELPAC data file and from the CAA for ELA section of the 2022 California Assessment of Student Performance and Progress (CAASPP) data file. The two data files were merged based on each student’s statewide student identification number (SSID) and records were kept only if the observation had valid test scores from either source file and had matching enrollment and test grades for grades three through eight and eleven. This resulted in a sample comprised of 32,389 total student observations with valid CAA for ELA test scores. Among these 32,389 students, 8,077 were EL students who also had valid tests scores from the Summative Alternate ELPAC (meaning EL students with valid CAA for ELA scores but lacking valid Alternate ELPAC scores are also excluded from the analyses).

The CDE conducted analyses to compare the study sample against the population of all students who participate in alternate assessments (see Table 10), and the full population EL students with the most severe cognitive difficulties (see Table 11). The first comparison provides information about any group-level differences between EL and non-EL students that could affect study findings or interpretations. The second comparison provides information about the extent to which the study sample, which is limited only to grades that also participate in the CAA for ELA, resembles the full K–12 population of EL students with the most significant cognitive disabilities, who all must participate in the Alternate ELPAC.

Relative to non-EL students who participate in alternate assessments (see Table 10), the CDE found that the EL students in the analytic sample differ from the non-EL students in terms of their distribution across grades and their prevalence of autism and intellectual disabilities. This information is particularly relevant for the results from the descriptive box plot method, which includes CAA for ELA results for EO, IFEP, and RFEP students as a point of comparison for the Alternate ELPAC students. It is also relevant for the exploratory logistic regression analyses using the mean CAA for ELA scores of EO students as a reference point (not shown in this report).

Table 10: Number and Percentage of Students in the Study Sample by English Language Proficiency Status, Primary Disability, and Grade Level

Categories: Primary Disability and Grade Level	Number of Students with Valid CAA ELA Scores	Percent Students with Valid CAA ELA Scores	Number of EO, IFEP, and RFEP Students with Valid CAA ELA Scores	Percent of EO, IFEP, and RFEP Students with Valid CAA ELA Scores	Number of EL Students with Valid Scores on Both Alternate ELPAC and CAA for ELA	Percent EL Students with Valid Scores on Both Alternate ELPAC and CAA for ELA
Total	32,389	–	24,312	–	8,077	–
Autism	14,441	44.6	11,489	47.3	2,952	36.6
Intellectual Disability	11,483	35.5	7,904	32.5	3,579	44.3
Other Disability	6,452	19.9	4,912	20.2	1,540	19.1
Missing Primary Disability	13	0.0	7	0.0	6	0.1
Grade Three	4,730	14.6	3,365	13.8	1,365	16.9
Grade Four	4,771	14.7	3,483	14.3	1,288	16.0
Grade Five	4,644	14.3	3,383	13.9	1,261	15.6
Grade Six	4,724	14.6	3,567	14.7	1,157	14.3
Grade Seven	4,651	14.4	3,471	14.3	1,180	14.6
Grade Eight	4,646	14.3	3,612	14.9	1,034	12.8
Grade Eleven	4,223	13.0	3,431	14.1	792	9.8

NOTE: The study sample excludes 816 non-EL students who were missing a flag for EO, IFEP, or RFEP status. It also excludes EL students who had valid CAA for ELA scores but lacked valid Alternate ELPAC scores.

The CDE’s comparisons of the analytic sample to the full population of students who participated in the Alternate ELPAC suggests that the analytic sample of EL students used in this study is not meaningfully different from the population of EL students with the most significant cognitive disabilities in terms of their home languages, disabilities, or distribution across grades. That said, there remain several limitations to this study’s sample, which are discussed in the next section.

Table 11: Comparison of the Full Alternate ELPAC Population and the Study Sample by Home Language and Grade Level

Categories: Home Language and Grade Level	Number Eligible for Alternate ELPAC	Percent All Eligible Alternate ELPAC	Number Tested with Valid Alternate ELPAC Scores	Percent All Eligible Alternate ELPAC	Number Matched Valid Alternate ELPAC and CAA for ELA Scores	Percent All Matched Valid Alternate ELPAC and CAA for ELA Scores
Total	19,697	–	16,669	84.6	8,077	–
Spanish	15,172	77.0	12,884	84.9	6,325	78.3
Chinese (Cantonese)	293	1.5	217	74.1	112	1.4
Chinese (Mandarin)	298	1.5	257	86.2	129	1.6
Other Home language	3,934	20.0	3,311	84.2	1,511	18.7
Grade Three	1,664	8.5	1,475	88.6	1,365	16.9
Grade Four	1,589	8.1	1,388	87.4	1,288	16.0
Grade Five	1,546	7.9	1,368	88.5	1,261	15.6
Grade Six	1,464	7.4	1,252	85.5	1,157	14.3
Grade Seven	1,497	7.6	1,258	84.0	1,180	14.6
Grade Eight	1,354	6.9	1,114	82.3	1,034	12.8
Grade Eleven	1,165	5.9	896	76.9	792	9.8

Sample Limitations and Considerations. Because the methods utilized scores from both the Alternate ELPAC and the CAA for ELA, they are necessarily limited to the tested grades for the CAA for ELA (grades three through eight, and grade eleven). Since the Alternate ELPAC is administered to all grades, K–12, and because the EL population is generally larger in the lower grades, this limitation means that these analyses are based on less than half of the student population who participated in the Alternate ELPAC operational field test (8,077 students out of 16,669 tested). Study 1, the contrasting groups study, used teacher judgment about students’ language acquisition status and allows for the inclusion of all tested grades.

Additionally, it is significant that these analyses are based on a single year of data, which was also the first year the Alternate ELPAC was administered statewide. Because of the policy shift from

locally determined instruments to a statewide one, it is likely that the performance distribution of students in the first year may differ from future years in important ways. For example:

- The operational field test census sample may include a “backlog” of students who would have been reclassified sooner had the Summative Alternate ELPAC been in place instead of locally-determined instruments.
- The operational field test marked the first time that the vast majority of teachers and students participated in the administration of the Alternate ELPAC, and both groups will likely become more familiar and confident with the assessment as time goes on.

To the extent that this is the case, future score distributions on the assessment may not look similar to the distributions observed in the first year. For this reason, the CDE intends to conduct follow-up analyses in future years as more data and information become available.

Method

For this supplemental study, REL West coached the CDE to apply three analytic methods that were developed by the US Department of Education as part of a National Evaluation of Title III (see Cook et al., 2012). These methods were designed specifically to support states to empirically identify appropriate thresholds for ELP, and they have since been used by many states—including California—to identify and evaluate thresholds on both general and alternate summative ELP assessments used for accountability (Linguanti et al., 2018; Cook, 2014).

All three methods are based on the facts that (1) English learner status is intended to address language barriers that may affect students’ academic learning and performance, and (2) monolingual English students demonstrate a range of performance on assessments of academic content like ELA and mathematics—they are not, in other words, uniformly proficient in these areas.

Accordingly, the methods’ developers argue, EL students should not be expected to achieve academic proficiency as a condition of exiting EL status. Rather, researchers and decision-makers

should look for evidence that students' ELP performance no longer strongly predicts or correlates with their academic content performance. This lack of a strong relationship serves as evidence that students' language proficiency is no longer affecting or driving their academic performance; a sign that the removal of the language services and supports of EL status are less likely to be harmful for students.

In the words of the authors: "...there is a point at which EL students have sufficient English language skills to adequately function in English on content assessments; accordingly, there should be observable decreases in the relationship between the two assessments. At or beyond this point is where the ELP performance standard might be considered..." (Cook et al., 2012, p. 8). The methods operationalize this concept by generally seeking the point at which students' ELP performance predicts their academic content achievement (proficient, or not proficient) at the level of chance. In other words, if students with the same level of ELP have a 50-50 chance of being proficient on an academic content assessment, this suggests that their ELP is not making a difference in their academic achievement. A stronger level of prediction in either direction would suggest that their ELP is driving their academic content achievement (so removal of services could be harmful).

Accordingly, all three methods use scores from both an ELP assessment and academic content assessments. For the current study, the CDE used scores from the CAA for ELA as the academic content assessment in the analysis, alongside the Alternate ELPAC. Each of the three methods seeks to identify the point of separation between performance in language and academic content in a different way:

1. **Descriptive box plot analysis** examines the distribution of overall scale scores on the CAA for ELA for students at each performance level on the Summative Alternate ELPAC, for each applicable grade level. This analysis also includes the distribution of overall scale scores on CAA for ELA for RFEP, IFEP, and EO students for comparison. The purpose of the analysis is to identify an Alternate ELPAC performance level where a) EL students have an equal

likelihood of scoring above or below the CAA for ELA Level 3 threshold (Understanding); or
b) EL student's score distribution on the CAA for ELA is very similar to that of EO students statewide.

2. **Logistic regression analysis** estimates the probability of reaching Level 3 (Understanding) on CAA for ELA for each Alternate ELPAC overall scale score. This approach helps to identify the Alternate ELPAC overall scale score range in which EL students have a probability equal to or greater than 50 percent of attaining that standard on the CAA for ELA.
3. **Decision consistency analysis** analyzes Alternate ELPAC and CAA for ELA proficient-level categorizations and optimizes consistent categorization of EL students at or above the current CAA for ELA threshold score for Level 3 (Understanding). For this analysis, consistency is defined as being proficient either on both the CAA for ELA and Alternate ELPAC, or on neither. Because states are expected to set their reclassification thresholds at a place where students have developed the linguistic knowledge and skills they need to function independently in classrooms where English is the language of instruction, a cut score that results in large numbers of "inconsistent" classification scenarios (i.e., students who are considered proficient in ELP but who overwhelmingly fail to reach proficiency in ELA, or who are proficient in ELA but still classified as English learners) would suggest that the ELP threshold is not set in the right place. Accordingly, this analysis determines the Alternate ELPAC overall scale score range that maximizes the amount of agreement between achieving Alternate ELPAC proficiency and CAA for ELA proficiency.

The CDE applied all three methods separately to each grade-level of the Alternate ELPAC for students in grades three through eight, and eleven and separately ran models for the two largest disability groups in the student sample – students with autism, and students with intellectual disabilities (see Table 1). The CDE also ran the logistic regression and decision consistency models

using three different performance thresholds on the CAA for ELA as a performance target. In all, these different approaches resulted in nearly 100 different results to consider collectively.

Results

This section provides selected data visualizations and summary information about the overall results for each method, including the analyses conducted for specific student groups or using different CAA for ELA scores as outcomes. In other words, the grades and groups that are not shown have results that point to the same achievement level as those shown in the report.

EL Students' Performance on California Alternate Assessments for English Language Arts and Alternate English Language Proficiency Assessments for California. Table 12 shows a cross tabulation of the number of students at each performance level on the Alternate ELPAC (under "Row Total") and achievement level on the CAA for ELA (under "Column Total"). The Table shows that students generally scored at a higher proficiency level on the Alternate ELPAC than they did on the CAA for ELA. The number of students at each of the three Alternate ELPAC performance levels are more evenly distributed than the number of students at each of the CAA for ELA proficiency levels, with more students scoring at Alternate ELPAC Levels 2 and 3 (3,345 and 2,711, respectively) compared to Level 1.

On the CAA for ELA, by contrast, half of all students (4,032) scored at the lowest achievement level, and approximately one in seven students scored at the highest level. This imbalance is further evident in the interior cells of the table, which show the number of students who earned each combination of performance levels across the two assessments (e.g., Alternate ELPAC Level 1 and CAA for ELA Level 2). While most students who scored at Alternate ELPAC Level 1 also scored at CAA for ELA Level 1 (1,821 out of 2,021 students), the Table shows that students who scored at Alternate ELPAC Level 2 were most likely to score at CAA for ELA Level 1 (1,887 out of 3,345). Similarly, students who scored at Alternate ELPAC Level 3 were most likely to score at CAA for ELA

Level 2. In other words, students generally performed at one level lower on the CAA for ELA than they did on the Alternate ELPAC.

Table 12: Alternate ELPAC Performance Level by CAA for ELA Achievement Level

Alternate ELPAC Performance	CAA for ELA Level 1	CAA for ELA Level 2	CAA for ELA Level 3	Row Total
Level 1	1,821	192	8	2,021
Level 2	1,887	1,262	196	3,345
Level 3	324	1,448	939	2,711
Column Total	4,032	2,902	1,143	8,077

Results from Descriptive Box Plot Analyses. Figure 2 displays box plots showing the CAA for ELA score distributions for all EL, EO, IFEP, and RFEP students with valid scores in grades four, seven, and eleven, respectively. Score distributions for current EL students are further broken out by overall performance level on the Alternate ELPAC (1, 2, or 3). Results for the other tested grades, as well as for students with autism and students with intellectual disabilities, align with the results shown here. The figures show three important results:

1. As expected, the mean and distribution of CAA for ELA scores are progressively higher for students at each Alternate ELPAC performance level. In other words, students who do better on the Alternate ELPAC, also generally do better on the CAA for ELA.
2. With the exception of Grade 3, none of the distributions reflect a situation where half the students score above, and half below, the proficiency cut score on the CAA for ELA, the students at Alternate ELPAC level 3 are consistently closest to this goal, particularly relative to the other groups. Across all grades, between a half and quarter of the students at Alternate ELPAC Level 3 score at or above proficient on the CAA for ELA, and the proportion for this group is consistently the largest relative to all other student groups.
3. The figures show that students at Alternate ELPAC Level 3 generally outperform students who are EO, RFEP, and IFEP on the CAA for ELA. In addition, across all the grades (three through

eight, and eleven) EL students at Alternate ELPAC Overall Level 2 had about same mean and median scores as EO and RFEP students on the CAA for ELA.

Taken together, these results generally confirm Alternate ELPAC Level 3 as an appropriate threshold for reclassification, given the observed relationship between students’ Alternate ELPAC scores and CAA for ELA at this performance level. The achievement gap between Alternate ELPAC Level 3 students and EO and RFEP students on the CAA for ELA is notable and warrants additional consideration, as noted in the ‘equity check’ analyses at the end of this study.

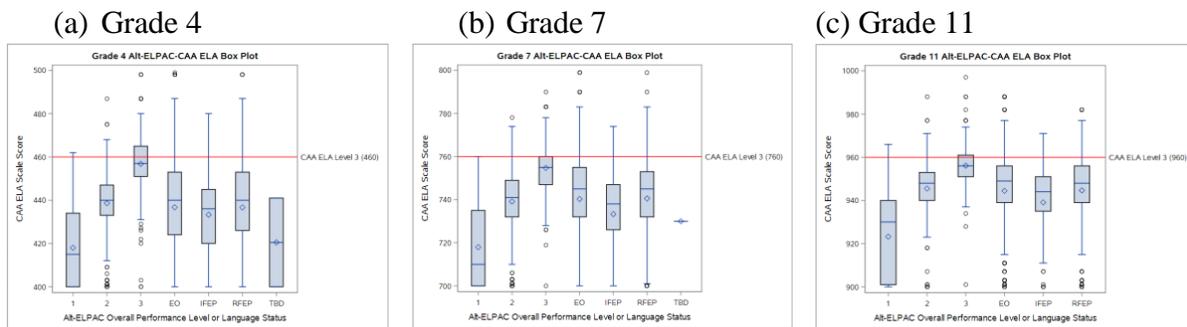


Figure 2. Descriptive Box Plots Alt-ELPAC – CAA ELA

The charts in Figure 2 (above) show the descriptive box plots for Grades 4, 7, and 11, respectively. In the charts show that approximately 40 percent of EL students in Grade 4, and 25 percent of students in Grades 7 and 11 perform at Level 3 on the Alternate ELPAC, achieve proficiency on the CAA for ELA and outperform other student groups. For all grade levels shown, EL students at Level 2 approximate the non-EL students on the CAA for ELA.234

Results from Logistic Regression Analysis. Selected results for the logistic regression analysis are displayed in Figure 3, which shows the probability of reaching Level 3 (Understanding) on the CAA for ELA for each Alternate ELPAC overall scale score value. As noted in the “Analytic Methods” section, the goal for this method is to identify the Alternate ELPAC scale score at which students have a 50 percent probability of scoring at Level 3 (Understanding) on the CAA for ELA. The figures show three notable results:

1. For all grades and groups, the scale score at which the probability is 50 percent falls in performance level 3 on the Alternate ELPAC.
2. The score at which the probability reaches 50 percent is higher in higher grades. In the grade eleven sample, there is no score at which the probability reaches 50 percent (see Figure 6).
3. The probability of achieving a Level 3 on the CAA for ELA at lower Alternate ELPAC performance levels is very low for all grades and does not vary by more than ten percentage points across the range of scores within the performance level. For example, the probability that a student at Alternate ELPAC Level 2 would score at Level 3 on the CAA for ELA ranges from approximately 5 to 15 percent for all grade levels shown here. This suggests there is a strong relationship between the ELP and ELA scores at these lower levels (i.e., students who attain any scale score within Alternate ELPAC Level 2 are very likely to not be proficient on the CAA for ELA).

Taken together, these results also affirm Alternate ELPAC Level 3 as the level at which the relationship between ELP and ELA decreases. Given the observed distributions of CAA for ELA performance among EO and RFEP students in the descriptive box plots methods (see figures 1 to 3 above), the CDE also ran logistic regression models using the mean EO achievement on CAA for ELA (not shown in this report), as well as the midpoint of Level 2 on the CAA for ELA, as alternate targets for the Alternate ELPAC to predict. Some of these results are shared below in the “equity check” analyses section.

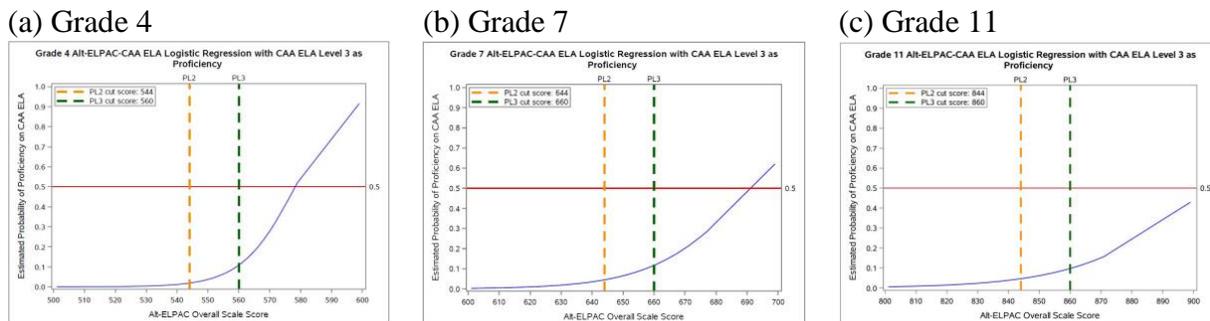


Figure 3. Logistic Regression, Proficiency Cuts

Caption: The charts in Figure 3 (above) show the logistic regressions for Grades 4, 7, and 11, respectively. The curved line crosses the horizontal reference line to the right of the vertical reference line for Level 3 on the Alternate ELPAC for Grades 4 and 7, meaning EL students must be beyond the Level 3 threshold to have at least a fifty percent chance of reaching proficiency on the CAA for ELA. The curved line does not cross the horizontal reference line at all for Grade 11, which means EL students at Level 3 on the Alternate ELPAC are not able to have a fifty percent chance of reaching proficiency on the CAA for ELA.

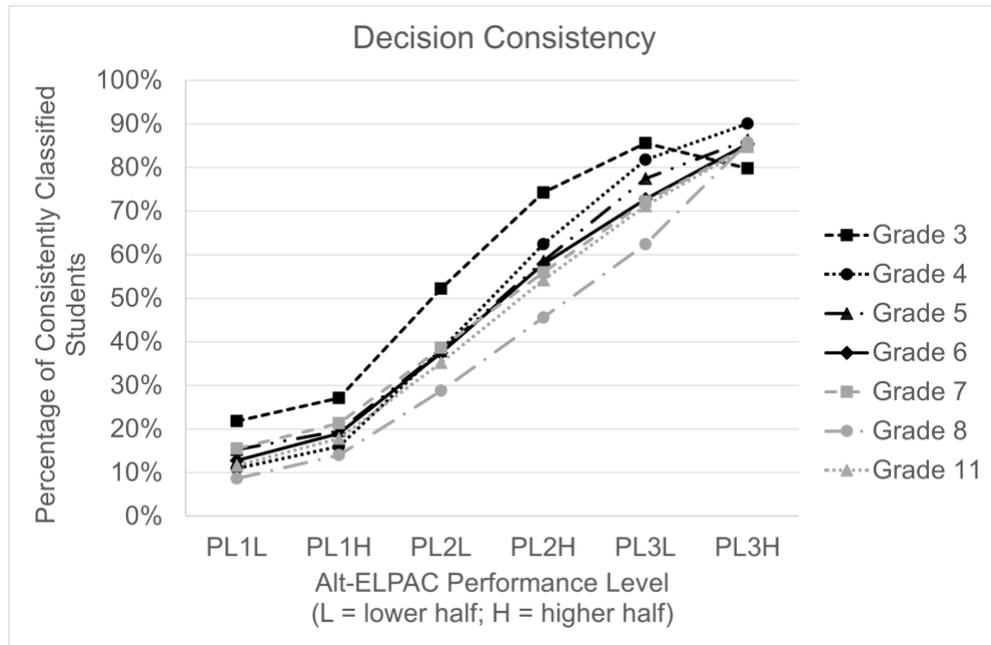
Results from Decision Consistency Analysis. Figure 4 shows the results of the decision consistency analysis for all tested grades in the study sample. As described above under “Analytic Methods,” this method optimizes consistent categorization of EL students as either proficient in both ELP and ELA, or in neither. For this analysis, students are further sorted into smaller groups based on whether their scale scores place them in the lower half (L) or higher half (H) of the score range for each Alternate ELPAC achievement level.

As Figure 4 illustrates, the results of these analyses generally corroborate earlier findings from the box plot and logistic regression analyses. The results are clearest for grade three, which shows a peak in consistent classifications at the lower end (threshold to below the midpoint) of the Alternate ELPAC Overall Level 3 scale score range, after which the percentage of consistent decisions decreases again. The other grades do not reach a similar peak (i.e., the curves at the median value of both the lower and higher half of Alternate ELPAC Overall Level 3 continue to have increasing slopes through the scale score range), but they do affirm that, across the performance continuum, consistent classifications are highest at Alternate ELPAC Level 3.

It is worth noting that this analysis is the most impacted, of the three, by the uneven score distributions on the Alternate ELPAC and CAA for ELA that are discussed above and summarized in Table 3. Because it was common for students to score one performance level higher on the Alternate

ELPAC versus the CAA for ELA, there are very few instances of inconsistent classifications in the other direction (students who are proficient in ELA but not in ELP), which likely explains why most grades do not show a point of inflection and subsequent decrease that generally occurs in this analysis.

Figure 4: Using CAA ELA Level 3 as Proficiency Cutoff



Caption: The chart shows only students in grade 3 reach the peak at performance level 3 low. All other grades are increasing across the figure such that higher consistency occurs with higher performance levels. The highest consistency for grades 4–8 and 11 occurs at performance level 3 high and is between 80–90 percent consistent.

5Results from an ‘Equity Check’ Exploratory Analysis Setting California Alternate Assessment for English Language Arts Target at a Level Other than “Proficient”. One function of EL status is to prepare EL students – including those with the most significant cognitive disabilities – to meet the same rigorous academic achievement standards set for all students. At the same time, it is also important—particularly given the language of the CDE’s reclassification Criterion 4—to try to avoid a situation wherein EL students appear to be held to a higher standard than their EO peers. These two goals appear to be somewhat in tension in the results above, specifically given that

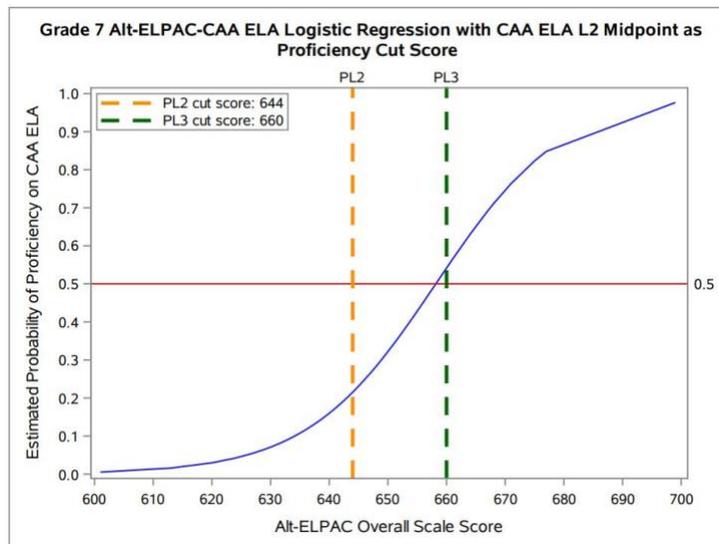
1. As shown in the descriptive box plot results (see figures 1 to 3), students at Alternate ELPAC Level 3 generally outperformed EO and RFEP students, and
2. The results from all three methods, as well as the descriptive results for the sample generally, suggest that Level 3 on the CAA for ELA is a high standard to meet.

For this reason, as noted above, the CDE also conducted several exploratory analyses using different CAA for ELA scores as targets for proficiency as a sort of ‘equity check.’ Specifically, the CDE ran logistic regression and decision consistency analyses using two alternative scores as a definition for acceptable performance on the CAA for ELA. These were:

1. The average score of EO students on the CAA for ELA in each grade level, and
2. The mid-point of Level 2 on the CAA for ELA

Results for the logistic regression models using the mid-point of Level 2 on the CAA for ELA as a target show that, with this score as the performance target, the point at which EL students’ probability of attaining proficiency on the CAA for ELA reaches the 50 percent target (suggesting equal odds of being proficient or non-proficient) is at or above the midpoint of Alternate ELPAC Level 2. Results are similar in other grades, and when using the mean CAA for ELA score of EO students as a target. An example for grade seven is shown in Figure 5.

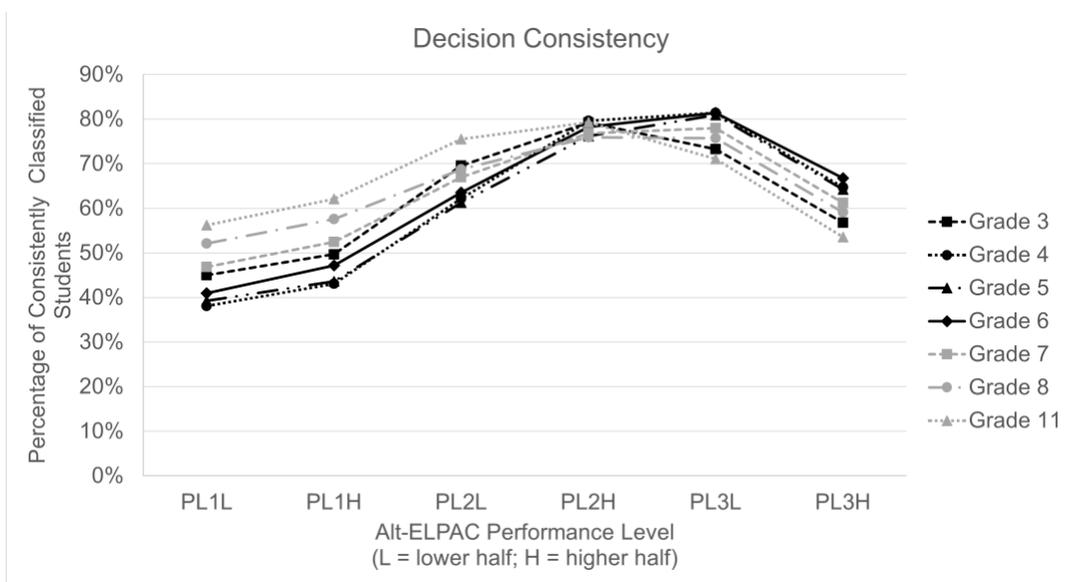
Figure 5: Logistic Regression Grade Seven, CAA ELA L2 Mid-Point Cut



Caption: The curved line crosses the horizontal reference line at approximately 658 (in between two vertical reference lines), which means students must be beyond the Level 2 threshold but lower than the Level 3 threshold on Alternate ELPAC to have at least a fifty percent chance of reaching the midpoint of Level 2 on CAA for ELA.

Results for the exploratory decision consistency analyses using the mid-point of CAA for ELA Level 2 are shown in Figure 6 for all grades. With this target, all grades now show a clear point of inflection, beyond which the percentage of consistent classifications decreases (meaning more students are proficient in ELA but not ELP). The Figure shows that, in all grades except grade three, the peak in consistent classifications occurs in the lower half of Alternate ELPAC Level 3. (For grade three, consistency peaks in the higher half of Alternate ELPAC Level 2). These exploratory results affirm Alternate ELPAC Level 3 as an appropriate reclassification threshold.

Figure 6: Using Midpoint of CAA ELA Level 2 as Proficiency Cutoff



Caption: The chart shows all grades peaking at performance level 2 high or performance level 3 low, at approximately 75–80 percent consistency.

Discussion

Taken together, the results from these exploratory analyses suggest Alternate ELPAC Level 3 as an appropriate reclassification threshold. Choosing between Levels 2 and 3 requires a careful consideration of the trade-offs between potentially removing EL services earlier in a student's language development trajectory (which is more likely to occur if the threshold is set at Level 2), versus potentially maintaining EL services past the point of their being useful (which is more likely to occur if the threshold is set at Level 3). An informed decision on this front requires understanding the nature and quality of English language development instruction that EL students with the most significant cognitive disabilities currently receive, and whether this instruction has any negative consequences in terms of access to content instruction or other learning opportunities. Recommending a Level 2 threshold would also constitute a policy change from the CDE's current recommended threshold, which could have implementation implications for the field.

As a result of these analyses, the CDE recommended and the SBE approved the current threshold of Overall Performance Level 3 on the Summative Alternate ELPAC for use in meeting Criterion 1 for reclassification decisions for EL students with the most significant cognitive disabilities. This was based on the following evidence and considerations:

1. The results for all methods, all grade-levels, and all student groups (including the two largest disability groups within the overall sample) identify Alternate ELPAC Level 3 as the performance level at which the relationship between ELP and ELA decreases relative to lower Alternate ELPAC achievement levels. This suggests that language proficiency is becoming less likely to drive academic achievement at this performance level compared to other performance levels.
2. Exploratory decision consistency analyses using lower achievement targets on the CAA for ELA also affirm Level 3 as an appropriate target for reclassification.

3. The results of the analyses conducted by the CDE consistently affirm this standard as the performance level at which language and academic achievement become less strongly related to each other, which is an appropriate range for reclassifying students out of EL status and ceasing their EL services.

Although the recommendations in this report are strongly backed by data, there are certain unavoidable limitations to these analyses. Specifically, the current analyses are based on only a single cohort of students – who, importantly, are the very first cohort of students ever to be recommended for reclassification using the Alternate ELPAC, rather than a local measure. Also, interest holders in the field also expressed a desire for more information about programming and services for EL students with the most significant cognitive disabilities – a desire shared by the CDE. In light of these considerations, future work could include replicating and expanding these analyses when additional years of data become available and collecting more information about the services provided to EL students with the most significant cognitive disabilities.

Conclusion

Two studies were conducted to evaluate the placement of the thresholds on the Summative Alternate ELPAC, a test of English proficiency for students with the most significant cognitive disabilities. 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.

The purpose of Study 1 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. 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).

The purpose of Study 2 was to conduct supplemental analyses of test-based performance data to evaluate the placement of the thresholds as it pertains to the interim guidance for meeting Criterion 1 for reclassification. Specifically, results on the Alternate ELPAC were compared with results on the CAA for ELA using a series of industry-standard analyses to identify and evaluate reclassification thresholds using scores on both ELP and academic content assessments.

Taken together, these studies support the continued use of PL 3 to meet Criterion 1 for reclassification. This conclusion is supported by the varied analyses presented here as well as political and logistical considerations for state and local educators and leaders who will be tasked with implementing new standardized reclassification procedures for EL students with the most significant cognitive disabilities now and into the future.

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 ELPAC. [Unpublished report] Sacramento, CA: California Department of Education.
- California Department of Education DataQuest Website (2023). <https://dq.cde.ca.gov/dataquest/>
- California *Education Code* Section 313(f). https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=313&lawCode=EDC
- Christensen, L.L., Mitchell, J.D., Shyyan, V.V., & Ryan, S. (2018). Characteristics of English learners with significant cognitive disabilities: Findings from the Individual Characteristics Questionnaire. ALTELLA. <https://altella.wceruw.org/pubs/ICQ-Report.pdf>
- Cook, H. G. (2014). Examining relationships between alternate ACCESS and state alternate assessments: Exploring notions of English language proficiency. WIDA Research Report. University of Wisconsin, Madison, Wisconsin Center for Education Research. <http://wida.wisc.edu/sites/default/files/resource/Report-Examining-RelationshipsBetweenAlternateAccessandStateAlternateAssessmentsAL.pdf>
- Cook, H.G., Linquanti, R., Chinen, M., & Jung, H. (2012). National Evaluation of Title III Implementation Supplement Report – Exploring Approaches to Setting English Language Proficiency Performance Criteria and Monitoring English Learner Progress. Washington, DC: American Institutes for Research. <https://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.pdf>
- Linquanti, R., Huang, M., & Crane, E. (2018) *Report on Supplemental Empirical Analyses of the English Language Proficiency Assessments for California*. Sacramento, CA: California Comprehensive Center at WestEd.
- 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.