

Supporting the California Department of Education in Examining Data to Inform the Setting of Thresholds on the California Alternate English Language Proficiency Assessments for California

Acronyms in this Memo:

- **ACCESS for ELLs** – Assessing Comprehension and Communication in English State-to-State for English Language Learners. The name of the ELP assessments used by member states of the WIDA consortium.
 - **CAA for ELA** – California alternate assessment for English language arts. The content assessment used in California to measure the ELA achievement of students with the most significant cognitive disabilities.
 - **CDE** – California Department of Education
 - **EL** – English learner
 - **ELP** – English language proficiency
 - **ELPAC** – English language proficiency assessments for California. Refers to a suite of four assessments used in the state of California to classify students and students with disabilities as English learners and annually measure their ELP.
 - **ETS** – Formerly stood for Educational Testing Service, an assessment organization. The company now officially goes by “ETS” only.
 - **SBE** – State Board of Education
 - **SCD** – Significant cognitive disabilities
 - **WIDA** – Formerly stood for World-class Instructional Design and Assessment, an assessment consortium. The organization now officially goes by “WIDA” only.
-

High-Leverage Problem

California’s K–12 education system serves the largest population of students classified as English learners (ELs) of any state in the country (National Center for Education Statistics, 2019). Of the 1,111,421 EL students in California in 2021/22, 20,450 (1.84%) were identified as EL students with the most significant cognitive disabilities (SCD) (California Department of Education [CDE], 2022). As for all students who are EL-classified, CDE’s goal is for EL students with the most significant cognitive disabilities to develop English proficiency and exit EL programs and services. According to state legislation,¹ in California, these exit decisions are based on the following four criteria:

¹ See California Education Code section 313(f):

https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=313.&lawCode=EDC

1. Assessment of language proficiency using an objective assessment instrument, including, but not limited to, the English language development test that is developed or acquired pursuant to Section 60810².
2. Teacher evaluation, including, but not limited to, a review of the pupil’s curriculum mastery.
3. Parental opinion and consultation.
4. Comparison of the performance of the pupil in basic skills against an empirically established range of performance in basic skills based upon the performance of English-proficient pupils of the same age, that demonstrates whether the pupil is sufficiently proficient in English to participate effectively in a curriculum designed for pupils of the same age whose native language is English.

Starting in the 2022/23 school year, local education agencies in California will be required to use scores from the Alternate English Language Proficiency Assessments for California (ELPAC) for reclassification criterion 1 above. (Prior to 2022/23, local education agencies had discretion to develop or select an assessment of English language proficiency for this criterion). The Alternate ELPAC is California’s first statewide alternate English language proficiency (ELP) assessment. The Alternate ELPAC was developed by the CDE’s assessment vendor, Educational Testing Service (ETS) in 2018. In the 2021/22 school year, ETS completed an operational field test of the Alternate ELPAC with California’s population of EL students with the most significant cognitive disabilities.³

Based on various information that ETS collected systematically from California teachers who work with EL students with SCD, ETS and the CDE have previously recommended that the highest achievement level on Alternate ELPAC (Level 3) represent fluency in English, and that students who achieve this level on the assessment be eligible to exit from EL services. The California State Board of Education (SBE) requested that the CDE conduct additional analyses to confirm that Alternate ELPAC Level 3 is an appropriate threshold for exit from services (also referred to as reclassification).

In response to this request from the SBE, the CDE requested technical support from REL West so the CDE could conduct the requested analyses. REL West, well known in the field for their expertise in this area, provided training to build the capacity of the CDE to conduct the analyses. The results of these analyses will be presented to the California SBE in May, as one piece of evidence (along with results from other studies and analyses in which REL West did not provide coaching) to recommend that the state maintain its current exit threshold of Alternate ELPAC Level 3, while continuing to evaluate its appropriateness based on additional data and analyses that the state plans to conduct in the future as more data become available.

Methods Summary

The methods the CDE used in this project were developed and disseminated with support from the U.S. Department of Education’s Office of Planning, Evaluation, and Policy Development (OPEPD) to support

² Section 60810 of the California Ed Code falls in Chapter 7, “Assessment of Language Development” and describes the state’s roles, responsibilities, and approach for designing or selecting, administering, and using assessments of English language proficiency for statewide purposes and to ensure compliance with various federal laws and rulings.

³ For more information on the Alternate ELPAC operational field test and standard setting, see: <https://www.elpac.org/test-administration/alternate/>

state education agency (SEA) personnel and technical assistance providers in using empirical data to inform decisions about thresholds and policy for EL reclassification (Cook et al., 2012). The methods were developed using real student data from multiple states in the WIDA consortium, which produces the most widely used assessment and alternate assessment of ELP in the country (the ACCESS for ELLs and Alternate ACCESS, which are currently used in 36 of the 50 US states, as well as in DC and several other territories). Since their development, these methods have been used in member states of the WIDA consortium (e.g., Tennessee Department of Education, 2019) and in California to inform the setting of thresholds on the general ELPAC assessment. They have also been used to inform the thresholds set on the WIDA Alternate ACCESS assessment (Cook, 2014).

The methods are designed to reflect the language and logic of the two most recent reauthorizations of the Elementary and Secondary Education Act by supporting states to identify the point at which language proficiency and academic achievement become less tightly coupled for students who are assessed in both areas. Specifically:

Empirically, researchers can define “English language proficient” as the point at which EL students’ academic content achievement assessed using English becomes *less* related to their ELP. That is, 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, and empirical procedures can help to identify this level of performance. (Cook et al., 2012, p. 8)

An important point by Cook and colleagues is that not all English-only (EO) students are proficient on a state’s academic content assessments. Relatedly, the purpose of the supports and services of EL status is to remove any linguistic barriers that may affect or inhibit students’ ability to access instruction or demonstrate knowledge or skills; it is not the goal or the expectation of EL status to ensure every student is proficient in academic content areas before exiting. Through this lens, attention to achievement in academic content areas is used more to gauge whether students’ ELP scores appear to be overdetermining their academic achievement in either direction (i.e., most or all students at a given ELP level are non-proficient in academic content areas, which could suggest that language is posing a barrier, or most or all students at a given ELP level are proficient in academic content areas, which could suggest that EL students are being maintained in language support services in order to meet a higher academic achievement standard than many English proficient students can meet).

There are three methods in total: descriptive box plots, logistic regression, and decision consistency. Each of the methods is designed to identify a score, or range of scores, at which there is evidence that students’ ELP is not shaping or inhibiting their ability to demonstrate their skills and knowledge on the state standardized content area assessments.⁴ Across the three methods:

- All are descriptive, in the sense that no two groups are being statistically compared for differential outcomes or impacts.

⁴ In the CDE’s case, the English language arts assessment was used, but the original methods can be used with any academic content assessment, and the authors recommend using both ELA and mathematics scores when possible.

- Two of the three methods (Methods 1 and 2) are entirely descriptive, meaning they only involve arraying extant assessment data in new groupings for consideration.
- Two of the three methods (Methods 2 and 3) use only data from EL students with disabilities. Method 1 also considers outcomes from students with disabilities who are not ELs as an additional group.

Each is described in more detail below, using the assessment names and details of the CDE’s particular analysis (e.g., the academic achievement assessment is referred to as the CAA for ELA, and the number of ELP performance levels is based on the number of performance levels on the Alternate ELPAC). In all cases, the methods have not been modified from their generic articulation in the original methods paper, other than to substitute in the names and details of California’s specific case.

The use of three related methods (rather than selecting only one of the three) reflects best practice recommendations from the assessment community (see, e.g., *Standards for Educational and Psychological Testing*; Hambleton & Pitoniak in *Educational Measurement, 4th edition*), which advises using multiple methods for standard setting as a means of triangulating and validating results to ensure that high-stakes decisions about achievement thresholds are not made based on a single judgment if possible. It is also informed by the potentially competing challenges of setting an ELP threshold at an appropriate place given the language demands that EL students will face in their academic instruction, while also not holding EL students to higher academic achievement standards than their English-only peers currently meet. The use of three methods allows for decisionmakers to use multiple sources of evidence to identify a threshold value, or in instances where findings diverge, can provide important context for the recommendation being made. In the case of this specific study, it also may be that the model-based logistic regression method may not have sufficient power with the student population, and in that case the descriptive analyses will provide important sources of information to triangulate the model-based results.

Method 1: Descriptive Box Plot

For the descriptive box plot analysis, the purpose is to produce box and whisker plots showing the scale score distribution of academic achievement scores (here, the CAA for ELA) for up to six distinct groups of students with the most significant cognitive disabilities in each tested grade. The groups are:

- Current ELs who scored at Alternate ELPAC Level 1
- Current ELs who scored at Alternate ELPAC Level 2
- Current ELs who scored at Alternate ELPAC Level 3
- Students who were screened to become English learners but scored above the state’s threshold for English proficiency (called initially fluent English proficient, or IFEP students)
- Students who were previously ELs but who have since reached the state’s threshold for English proficiency (called reclassified fluent English proficient, or RFEP students)
- Students who have never been screened for or classified as English learners (called “English only” or EO students)

All six groups are mutually exclusive (i.e., a student can only fall into one of the six categories), and, for this study, all are limited only to students with the most significant cognitive disabilities who participate

in alternate assessments. The goal of this method is to look for the Alternate ELPAC level or levels where:

1. Roughly half of all current EL students are above, and half below, the proficiency cut score on the CAA for ELA assessment. This is interpreted to be the Alternate ELPAC achievement level where students have a roughly equal chance of being proficient or non-proficient on the CAA for ELA—suggesting that their academic achievement is not being determined by their language proficiency.
2. The distribution of current EL students most closely resembles the distribution of one or more of the non-EL student groups (EO, IFEP, or RFEP). This is interpreted to be the Alternate ELPAC achievement level where the academic achievement distribution for current EL students is most similar to the academic achievement distribution for non-EL students with the most significant cognitive disabilities.

Ideally, both of the indicators above will identify the same Alternate ELPAC achievement level. If they do not, both levels should be considered in concert with the outcomes from the other two methods, which rely only on the achievement of current EL students.

Method 2: Decision Consistency

The second method consists of decision consistency analyses. For this method, students' outcomes on both the CAA for ELA and Alternate ELPAC were analyzed to identify the ELP performance range where the most EL students are either proficient on both assessments or are proficient on neither assessment, given the proficiency thresholds on the CAA for ELA assessment. This method identifies the Alternate ELPAC achievement level that maximizes the number of students with consistent achievement classifications (proficient or not proficient) across the two assessments.

This method is designed to explicitly reflect requirements from the current reauthorization of the Elementary and Secondary Education Act, which require that ELP standards and assessments be “aligned with” or “correspond with” academic content and performance standards. In other words, well-designed ELP standards and assessments reflect the linguistic knowledge and skills that students need to meet their states' academic content and achievement standards in ELA, mathematics, and science. If a peak is observed, where the proportion of students who are proficient in ELP and on a given content assessment peaks and then declines, this would suggest that, after the peak, the divergent outcomes reflect not language proficiency but content proficiency in the academic domain being tested on the content assessment. The expectation is not that the peak is 100%, but rather that there will be an increase in the decision consistency rate up until the peak, at which point there will be a decline. (This pattern has been observed consistently in data from the WIDA ACCESS for ELLs assessment, and was also observed in California's data when these methods were applied to the general ELPAC assessment.) The methods' authors recommend conducting this analysis for multiple academic content areas, as the peaks may occur in different places for the different content areas, depending on their differing language demands.

Method 3: Logistic Regression

The third method uses logistic regression analyses to estimate the probability that a student scores at the “proficient” performance level on the CAA for ELA given their Alternate ELPAC scale score. The logic behind this method is that a probability of 0.5 indicates a relationship between the ELPAC and the CAA

that is essentially a toss-up, with equal odds of being proficient or not proficient on the latter. States are encouraged to consider a reclassification threshold that is at or above the Alternate ELPAC scale score that predicts this probability.

Rationale for CDE’s Use of the Methods and REL’s Technical Support Role

To maintain consistency with the general ELPAC analyses completed by the state in 2018/19, the CDE requested support from REL West to use Cook and colleagues’ (2012) methods to explore the relationship between EL students’ scores on the Alternate ELPAC and scores on the California Alternate Assessment (CAA) for ELA. The REL team coached the CDE team to conduct all three approaches—descriptive box plot analyses, decision consistency analyses, and logistic regression analyses (Cook et al., 2012).

- All analyses were conducted using the single year of data available from the Alternate ELPAC operational field test, which took place in 2021/22. Because the Alternate ELPAC is vertically scaled with different achievement level cut scores in each grade, all methods were conducted separately for students in each grade overall, and then by specific disability where there were sufficient students in the pool ($N > 10$) to conduct the analysis. Additionally, the CDE conducted separate analyses for only students with the two most prevalent disability categorizations—autism, and intellectual disabilities—to explore variations by disability (none were found).
- Additionally, in response to initial findings from all three methods that students at Alternate ELPAC Level 3 outperformed all three non-EL groups with significant cognitive disabilities (EO, IFEP, and RFEP) on the CAA for ELA, REL West and the CDE collaboratively decided that the CDE should also rerun the logistic regression and decision consistency methods using two alternative targets on the CAA for ELA: the mean scale score earned by English-only students, and the mid-point of the second achievement level (out of three) on the CAA for ELA.

Once all methods and analyses were complete, REL West coached the CDE to examine the results from all analyses in relation to one another, as well as with the proposed thresholds from the CDE and ETS’s prior standard-setting efforts to identify a recommended threshold. The REL West team coached the CDE to examine the different thresholds identified through the various methods and make a conclusion about the degree to which the evidence and proposed thresholds align. The REL West team also trained the CDE to identify and discuss instances where the different models converged or diverged in terms of findings, and facilitated discussions about how to resolve or reconcile divergences to arrive at a defensible recommendation.

Part of this work included supporting the CDE to identify important interest holder groups with whom to share their initial findings and recommendations so that the CDE could receive feedback from the field on their priorities and preferences for a recommended threshold. The CDE ultimately shared their results with four major interest holder groups that collectively represented the assessment, special education, and English learner communities. Information about the composition and purpose of these groups is provided in the Appendix below. All groups were shown the same presentation and invited to ask questions and share feedback and reactions. In general, none of the groups expressed major concerns about maintaining the current Level 3 threshold, but almost all of the groups sought to contextualize the recommendation by asking questions about services and instruction for students with SCDs, including EL students in this group. Broadly speaking, special education and English learner interest holder groups were more concerned about students losing access to services, while the state

assessment technical advisory group members were more concerned about students being retained in EL status. The questions and concerns that these groups raised informed the CDE's final recommendation.

Summary of the CDE’s Findings and Recommendations

Detailed results from the CDE’s application of the three methods above are available in a report by the CDE which will be made publicly available ahead of the May SBE meeting at which CDE will present the slides that accompany this memo. A summary version of the CDE’s conclusions are as follows:

The CDE recommends that the SBE maintain 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 SCDs. This recommendation is based on the following evidence and considerations:

1. The results for all methods, all grade levels, and all student groups in this analysis (including the two largest disability groups within the overall sample, students with autism and students with intellectual disabilities) 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.

Recommendations for Future Analysis

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 locally developed or determined measure of English language proficiency. Additionally, the operational field test sample on which these analyses are based:

- 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.
- Took place in the early years of the COVID-19 pandemic, which may have had a particularly negative effect on access and learning opportunities for students classified as ELs and students with SCDs—both separately, and as overlapping groups.
- Is necessarily limited to students who are also tested on the CAA for ELA (i.e., grades 3 through 8 and high school), despite the fact that the Alternate ELPAC is administered to students in all grades, K–12.

Finally, 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, the CDE recommends the following activities and analyses for future work:

1. **Replicate the current analyses with additional years of data as they become available.** As more students take the Summative Alternate ELPAC, and as students and educators become more familiar and confident with the Summative Alternate ELPAC’s content and administration, there

is potential for the relationships and distributions to shift over time. By continuing to replicate these analyses as future years of data become available, the CDE will be able to monitor the findings of these analyses.

2. **Conduct additional analyses on current and future data.** In addition to replicating the current analyses with future data, the CDE could also expand on these analyses by, for example, replicating them with CAA for math scores, exploring the decisions and outcomes from reclassification decisions that have already been made for students who participated in the first two years of the Alternate ELPAC, or more closely evaluating the cut scores between the Alternate ELPAC achievement levels to see if these would benefit from minor adjustments (as was done for the general Summative ELPAC).
3. **Collect more detailed information about language instruction services for EL students with significant cognitive disabilities.** As noted above, the current recommendation errs on the side of maintaining services for classified students, which reflects the priorities of most advocacy and interest holder groups for this population. With more detailed information, the CDE could develop a more nuanced understanding of the benefits and trade-offs of continuing to receive services (such as whether students receiving ELD instruction miss out on other learning opportunities). This information can provide both the CDE and the field with a more detailed sense of what reclassification really means for students and ensure that the Summative Alternate ELPAC threshold is serving students' best interest.
4. **Identify opportunities for data collection, professional learning, and technical assistance related to academic instruction for all students with significant cognitive disabilities.** As the descriptive box plot results showed, EO and RFEP students generally are not performing at high levels on the CAA for ELA. It may be the case that special educators and leaders need more supports or resources to implement instruction that prepares their students to meet the performance standards of the assessment. There may be opportunities for the CDE to bolster supports for special educators that may improve the achievement of students with the most significant cognitive disabilities.

With support from the SBE, the CDE will pursue these analyses to the extent practicable.

Appendix. Interest Holder Group Descriptions

The following groups were engaged by CDE throughout this process:

- **Regional County Office of Education English Learner Specialists (RELS):**
 - Leads from each of the 11 California County Superintendents' regions are engaged for the purpose of guiding local educational agencies (LEAs) in the development, implementation, and evaluation of programs and services for English learners. RELS convenings are facilitated by the California Department of Education. RELS are part of the Statewide System of Support and a subset of the Bilingual Coordinators' Network.
- **Regional Assessment Network (RAN):**
 - The County Office of Education (COE) RAN members represent a sub-committee of the Curriculum and Instruction Steering Committee (CISC) of the County Superintendents Association. The purpose of RAN is to develop a knowledge and understanding about California's assessment and accountability systems by facilitating communication between the California Department of Education (CDE) and COEs, which represent the schools, districts, and county offices in their region.
- **Assessment Interest Holder Meeting:**
 - Monthly meeting hosted by the Assessment Development and Administration Division at CDE to provide assessment updates to interested parties. Includes representatives from the California Teachers Association, Californian's Together, California Association for Bilingual Educators, Charter School Schools Association, Association of California School Administrators, California School Boards Association, California County Superintendents, California Association of Science Educators, Capitol Advisors, and public advocates.
- **California Assessment System Technical Advisory Group (TAG):**
 - A group of national technical experts and district representatives who provide feedback on various assessment-related topics for California's assessment programs. California has separate but overlapping TAGs for content and English language assessments, with approximately 16 individuals represented across the two groups. TAG members' expertise include psychometrics, assessment design and development, assessment validity, education policy, language development, and special education.
- **Special Education Division at the CDE:**
 - Members of the CDE's Special Education Division were also convened to discuss the findings and potential implications for the field. Helpful suggestions were provided to the Multilingual Support Division with regard to communications out to the field.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.
- ETS. (2023, March). *Draft Summative Alternate English Language Proficiency Assessments for California Threshold Score Validation Study Report*. Sacramento, CA: California Department of Education.
- California Department of Education. (2022). *2021–22 enrollment by English language acquisition status and grade: Statewide report*. Retrieved on June 24, 2022, from <https://dq.cde.ca.gov/dataquest/>
- Cook, G., Linqanti, R., Chinen, M., & Jung, H. (2012). *National evaluation of Title III Implementation Supplemental Report: Exploring approaches to setting English language proficiency performance criteria and monitoring English learner progress*. Office of Planning, Evaluation and Policy Development, U.S. Department of Education.
- Cook, G. (2014). *Examining relationships between Alternate ACCESS and state alternate assessments: Exploring notions of English language proficiency*. Wisconsin Center for Education Research, University of Wisconsin-Madison. <https://wida.wisc.edu/sites/default/files/resource/Report-ExaminingRelationshipsBetweenAlternateAccessandStateAlternateAssessmentsAL.pdf>
- Hambleton, R., & Pitoniak, M. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Praeger Publishers.
- McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., & Barmer, A. (2019). *The condition of education 2019* (NCES 2019-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019144>
- Tennessee Department of Education. (2019). *ESL Program Policy 3.207*. Retrieved from <https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/2019-sbe-meetings/july-25%2C-2019-sbe-workshop/7-25-19%20ESL%20Program%20Policy.pdf>.