California Department of Education
Assessment Development & Administration Division

Summative English Language Proficiency Assessments for California Technical Report

2017–18 Administration

Final Submitted October 11, 2019

Educational Testing Service

Contract No. CN140284

Table of Contents

List of Appendices

List of Tables

List of Figures

Acronyms and Initialisms Used in the *English Language Proficiency Assessments for California Technical Report*

Term	Definition
2PL	two-parameter logistic
AERA	American Educational Research Association
APA	American Psychological Association
CAASPP	California Assessment of Student Performance and Progress
CCR	*California Code of Regulations*
CDE	California Department of Education
CDS	county/district/school
CELDT	California English Language Development Test
COE	county office of education
CR	constructed response
CSEM	conditional standard error of measurement
DIF	differential item functioning
EC	*Education Code*
ECD	evidence-centered design
ELA	English language arts/literacy
EL	English learner
ELD Standards	English Language Development Standards
ELP	English language proficiency
ELPAC	English Language Proficiency Assessments for California
ESKM	Enterprise Score Key Management
ESSA	Every Student Succeeds Act
ETS	Educational Testing Service
GPC	generalized partial credit
HOSS	highest obtainable scale score
IBIS	Item Banking Information System
IEP	individualized education program
IFEP	initial fluent English proficient
IRT	item response theory
LEA	local educational agency
LOSS	lowest obtainable scale score
MC	multiple choice
MH DIF	Mantel-Haenszel differential item functioning
MSICL	multiple-selection inline choice list
NCME	National Council on Measurement in Education
NS	No Score
OIB	ordered item booklet
ONE	Online Network for Evaluation
OTI	Office of Testing Integrity

Table of Acronyms and Initialisms *(continued)*

Term	Definition
PIN	problem item notification
PLD	performance level descriptor
Pre-ID	pre-identification
QC	quality control
RFEP	reclassified fluent English proficient
SBE	State Board of Education
SCOE	Sacramento County Office of Education
SD	standard deviation
SEM	standard error of measurement
SMD	standardized mean difference
SSR	Student Score Report
TCC	test characteristic curve
TIPS	Technology and Information Processing Services
TOMS	Test Operations Management System
USC	United States Code

Chapter 1: Introduction

1.1. Background

The English Language Proficiency Assessments for California (ELPAC) "is the required state test for English language proficiency (ELP) that must be given to students whose primary language is a language other than English. State and federal law require that local educational agencies (LEAs) administer a state test of ELP to eligible students in kindergarten through grade twelve" (California Department of Education [CDE], 2019). California *Education Code (EC)* Section 313(a) requires that the assessment of ELP be done upon initial enrollment and annually thereafter until the LEA reclassifies the student as English proficient.

In November 2012, the California State Board of Education (SBE) adopted the 2012 *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards). At that time, the California English Language Development Test (CELDT) was used as both the initial assessment and the summative ELP assessment. To provide an assessment aligned with the 2012 ELD Standards, the CDE contracted with Educational Testing Service (ETS) to develop the ELPAC. The CDE began transitioning from the paper-pencil CELDT to the paper-pencil ELPAC in 2017–18. In that year, the CELDT continued as the ELP assessment for initial identification, and the Summative ELPAC was used as the annual ELP assessment in spring 2018.

This technical report describes the development, administration, and results of the 2017–18 administration of the Summative ELPAC.

1.2. Test Purpose

The ELPAC consists of two assessments: the Initial ELPAC and the Summative ELPAC. The Initial ELPAC identifies whether a student is an English learner (EL) and would therefore benefit from additional instructional supports. Students identified as ELs on the Initial ELPAC go on to take the Summative ELPAC. The Summative ELPAC is administered annually to students in kindergarten through grade twelve who have been identified as ELs.

1.3. Test Content

Under the Every Student Succeeds Act (ESSA) and California state regulations, students who are identified as ELs are required to take the ELPAC in the domains of Listening, Speaking, Reading, and Writing.

The content of the Summative ELPAC is aligned with the 2012 ELD Standards. The test content corresponds to the *California Common Core State Standards: English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.* Items on the Summative ELPAC also correspond to the *California Common Core State Standards for Mathematics* as well as the *Next Generation Science Standards for California Public Schools, Kindergarten through Grade Twelve*.

1.4. Testing Window

California Code of Regulations (CCR), Article 1, Section 11518(d) establishes the Summative ELPAC testing window from February 1 through May 31 annually. During this time period, any student identified as an EL must be administered the Summative ELPAC.

The testing window for the 2017–18 administration of the Summative ELPAC was from February 1, 2018, through May 31, 2018.

CCR Section 11518(m) establishes the Initial ELPAC testing window from July 1 through June 30 of each school year.

1.5. Intended Population

All students who previously took the CELDT, who were identified as ELs, and who were enrolled between February 1, 2018, and May 31, 2018, were required to take the Summative ELPAC. All students classified as ELs must be tested annually during the Summative ELPAC window until they are reclassified as fluent English proficient (RFEP) based on the CDE's established guidelines for reclassification established by the SBE (*EC* 313[f]).

1.6. Intended Use and Purpose of Test Scores

The SBE approved the reporting hierarchy of the Summative ELPAC in September 2017. Individual student scores for the Summative ELPAC for all grades (i.e., kindergarten through grade twelve) included

- an overall score based on a continuous scale;
- an oral language subscore which reflects performance on the Listening and Speaking domains based on a continuous scale;
- a written language subscore which reflects performance on the Reading and Writing domains based on a continuous scale; and
- the student's proficiency within each domain (i.e., Listening, Speaking, Reading, and Writing) based on three levels.

Each student who took the 2017–18 paper-pencil Summative ELPAC received an overall score, an oral language subscore, and a written language subscore, which placed the student within one of the four ELPAC proficiency levels:

1. Beginning stage
2. Somewhat developed
3. Moderately developed
4. Well developed

The three scale scores—overall score, oral language subscore, and written language subscore—were all linked to the four ELPAC proficiency levels.

Similar scale scores across adjacent grade levels or grade spans and adjacent editions indicated a comparable degree of ELP. For example, similar scale scores on the grade two and grade span three through five assessments, or the grade span three through five and grade span six through eight assessments, indicated similar degrees of ELP. Further, similar scale scores from the 2017–18 and 2018–19 administrations will indicate similar degrees of ELP.

The purpose of the scale scores is to track student progress in ELP from year to year. For example, the scale scores will be used to track student progress in ELP from 2017–18 to 2018–19, once students who continue to be designated as ELs take the 2018–19 Summative ELPAC.

1.7. Limitations of the Assessment

Students who are identified as ELs must be tested annually during the annual assessment window—February 1 through May 31—until they are reclassified as RFEP. Because the Summative ELPAC is the ELP assessment developed pursuant to *EC* Section 60810, scores from the Summative ELPAC are one set of criteria used to determine whether individual students qualify for RFEP. Results from the Summative ELPAC may also be used to plan for instruction.

1.8. Organizations Involved with the ELPAC Program

1.8.1. State Board of Education

The SBE is the state agency that establishes educational policy for kindergarten through grade twelve in the areas of standards, instructional materials, assessment, and accountability. The SBE adopts textbooks for kindergarten through grade eight, adopts regulations to implement legislation, and has the authority to grant waivers of the *EC*.

In addition to adopting the rules and regulations for itself, its appointees, and California's public schools, the SBE also is the state educational agency responsible for overseeing California's compliance of the ESSA and the state's Public School Accountability Act, which measures the academic performance and progress of schools on a variety of academic metrics (CDE, 2017).

1.8.2. California Department of Education

The CDE oversees California's public school system, which is responsible for the education of more than 6,200,000 children and young adults in more than 10,450[1] schools. California aims to provide a world-class education for all students, from early childhood to adulthood. The CDE serves the state by innovating and collaborating, as a team, with educators, school staff, parents/guardians, and community partners to prepare students to live, work, and thrive in a highly connected world.

Within the CDE, the Performance, Planning & Technology Branch oversees programs promoting innovation and improving student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data (CDE, 2018c).

1.8.3. California Educators

A variety of California educators, including teachers experienced in teaching ELs and school administrators, were selected based on their qualifications, experiences, demographics, and geographic locations and invited to participate in the ELPAC development process. In this process, California educators participated in tasks that included defining the purpose and scope of the assessment, assessment design, item development, standard setting, score reporting, and scoring constructed-response items.

1.8.4. Contractors

1.8.4.1 Primary Contractor—Educational Testing Service

The CDE and the SBE contract with ETS to develop and administer the ELPAC. As the prime contractor, ETS has the overall responsibility for working with the CDE to implement and maintain an effective assessment system and to coordinate the work of ETS with its

[1] Retrieved from the CDE Fingertip Facts on Education in California – *CalEdFacts* web page at https://www.cde.ca.gov/ds/sd/cb/ceffingertipfacts.asp

subcontractors. Activities directly conducted by ETS include, but are not limited to, the following:

- Providing management of the program activities
- Providing tiered help desk support to LEAs
- Developing all ELPAC items
- Constructing, producing, and controlling the quality of ELPAC test forms and related testing materials, including grade- and content-specific *Examiner's Manuals*
- Hosting and maintaining a website with resources for the ELPAC
- Developing, hosting, and providing support for the Test Operations Management System (TOMS)
- Producing and distributing score reports
- Developing a score reporting website
- Completing all psychometric procedures

1.8.4.2 Subcontractor—Sacramento County Office of Education

ETS contracted with the Sacramento County Office of Education to manage all activities associated with training and outreach, including the following:

- Supporting and training county offices of education, LEAs, and charter schools
- Developing informational materials
- Recruiting and logistics for educator trainings
- Producing training videos

1.9. Overview of the Technical Report

This technical report addresses the characteristics of the ELPAC administered in spring of the 2017–18 school year and contains 11 additional chapters, as follows:

- Chapter 2 presents an overview of the processes involved in a testing cycle for the ELPAC. This includes item development, test construction, test administration, test participation, and accessibility.
- Chapter 3 describes the procedures followed during item development, various reviews (e.g., item content and bias and sensitivity reviews), and the process of item review.
- Chapter 4 describes the process of test assembly, including the content being measured, as well as the content and psychometric criteria. Also discussed is materials development.
- Chapter 5 details the processes involved in the actual 2017–18 administration. It also describes the procedures followed to maintain test security throughout the test administration process.
- Chapter 6 summarizes the standard setting process that established the base year (i.e., 2017–18) ELPAC scores. Details include the performance level descriptors, an overview of the standard setting methodology, and the process to establish the threshold scores that define the score ranges for each ELPAC level. These standard

setting processes were based on student testing results from the spring 2017 stand-alone field test, which occurred between March and April 2017.

- Chapter 7 provides information on the scoring processes and summarizes the types of scores and score reports.
- Chapter 8 summarizes the statistical procedures and results for 2017–18, including
  - classical item analysis;
  - differential item functioning analysis;
  - item response theory calibration, linking, and scaling;
  - reliability analyses; and
  - analyses of the consistency and accuracy of the performance level classifications.
- Chapter 9 discusses the procedures designed to ensure the validity of score uses and interpretations.
- Chapter 10 highlights the quality-control processes used at various stages of the 2017–18 Summative ELPAC administration, including item development, test assignment, test administration, scoring procedures, psychometric analysis processes, and score reporting.
- Chapter 11 describes the ELPAC field test and how it was scored and analyzed. Included are the results of special studies that were conducted.
- Chapter 12 details the ongoing means of program improvement.

References

California Department of Education. (2014). *California 2012 English Language Development Standards: Kindergarten Through Grade 12*. Adopted by the State Board of Education November 2012. Sacramento, CA: Author. Retrieved from http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

California Department of Education. (2019). English Language Proficiency Assessments for California (ELPAC). Retrieved from https://www.cde.ca.gov/ta/tg/ep/

California Department of Education. (2018c, July). *Organization.* Retrieved from http://www.cde.ca.gov/re/di/or/

California Department of Education. (2017, October). *State Board of Education responsibilities.* Retrieved from http://www.cde.ca.gov/be/ms/po/sberesponsibilities.asp

Chapter 2: Overview of ELPAC Development Processes

This section describes the processes used to develop a high-quality pool of items, assemble tests, administer tests, and provide accommodations as required by *Education Code* Section 60810 for the 2017–18 Summative English Language Proficiency Assessments for California (ELPAC).

2.1. Item Development

To construct test forms for the 2017–18 Summative ELPAC, an appropriate pool of items needed to be developed. The item development process started with the creation of item development specifications, which described the quantity of items to be created and the process to be followed. After the item development specifications were reviewed and approved by the California Department of Education (CDE), Educational Testing Service (ETS) assessment specialists worked with two groups of people to draft test items: California educators and ETS contractors. In February 2017, California educators attended an item-writer training workshop, where they received training, generated lists of topics for items, and drafted items. The educators focused on the development of Speaking and Writing items as well as shorter Listening and Reading items.

At the same time, the ETS contractors developed topics for longer Listening sets and Reading sets. ETS then compiled the topics from both groups and submitted them to the CDE for review. Once approved, the topics for the longer sets were sent to five ETS contractors with prior experience in developing Listening and Reading sets. The contractors then submitted their draft items to ETS for review.

All items drafted by California educators and ETS contractors went through internal ETS reviews, including two content reviews, a fairness review, and an editorial review. The items were then submitted to the CDE for review and approval.

Each item was then reviewed during two educator meetings: a Content Review Panel meeting and a Bias and Sensitivity Review Panel meeting.

During the Content Review Panel meeting, California educators considered whether each item would appropriately measure the aligned standard(s), whether each item was appropriate for the designated grade level or grade span, and whether each item was presented clearly and effectively. Multiple-choice (MC) items were also reviewed to ensure that each one had a single best key and distractors that were all plausible yet wrong. In addition, constructed-response (CR) items were reviewed to make sure that each prompt would elicit a response that allows students to demonstrate their language abilities, as described by the 2012 *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards).

During the Bias and Sensitivity Review Panel meeting, educators considered whether each item was free of content that was potentially biased against or offensive to any identified group, such as students from other countries or students who are deaf or hard of hearing. If an item contained potentially biased or offensive content, the educators considered whether the item could be revised to remove the potentially biased or offensive content.

Educators at both the Content Review Panel meeting and the Bias and Sensitivity Review Panel meeting had the option of making one of three decisions regarding each item:

approve the item as is, approve the item with revisions, or reject the item. Whenever an item was approved with revisions, educators specified the revisions needed to text or images and the reasons for the proposed revisions.

After the educator meetings, CDE staff reviewed the proposed revisions and made final decisions as to whether each educator's proposed revisions should be implemented. ETS assessment specialists then applied the CDE-approved revisions.

After the items were revised, CDE staff confirmed that revisions were entered correctly. After ETS implemented any necessary final revisions, the CDE approved the items for use as field test items.

All items that were used in the 2017–18 Summative ELPAC were administered in the 2017 stand-alone field test and approved for operational use as described in subsection *2.2.4 Psychometric Review*.

2.1.1. Item Format

The 2017–18 paper-pencil ELPAC contained three item formats: (1) single-selection multiple choice (MC) items, (2) multiple-selection inline choice list (MSICL) items, and (3) constructed response (CR).

1. **MC items** contained a question that was followed by three or four options as answer choices, one of which was the correct option.
2. **MSICL items,** which were found in the kindergarten Reading test, contained a series of questions. After the test examiner assessed each of the student's responses to the series of MSICL questions as correct, incorrect, or no response, scoring rules were used to assign the student with full, partial, or no credit. This item format was treated the same as the CR item format for statistical analysis.
3. **CR items** consisted of a prompt that elicited either a spoken response or a written response. A rubric was used to assess the quality of the response on a scale of 0–1, 0–2, 0–3, or 0–4. The rubrics described typical characteristics of a response at each score point based on criteria that were derived from the 2012 ELD Standards.

2.1.2. Item Writing Guidelines

Item writing guidelines were developed to define the task types and content of the items to provide guidance to item writers and drive consistency and efficiency in item development. The guidelines were used to facilitate the development of comparable items that measure appropriate skills and content aligned with the 2012 ELD Standards.

2.1.3. Item Banking

After items were drafted, they were entered in the ETS Item Banking Information System (IBIS). IBIS contains fields for entering item content and information about items for MC and CR items. IBIS was used to store item text, graphics, scripts for audio recordings, scoring information, statistical information, and metadata. After ETS staff drafted and reviewed items in IBIS, CDE staff used IBIS to review items in preparation for educator reviews and to ensure that ETS had revised items accurately after the educator reviews.

2.2. Test Assembly

ETS assessment specialists assembled the Summative ELPAC tests, which were reviewed and approved by the CDE. This process began with the creation of test development specifications, which described the content characteristics, psychometric characteristics,

and quantity of items to be used in the 2017–18 Summative ELPAC. ETS created the test development specifications that the CDE reviewed and approved.

After the test development specifications were approved, ETS assessment specialists assembled the tests in IBIS according to the specifications. IBIS then generated form planners, which are spreadsheets containing essential item information such as the number of items, the alignment of items according to the 2012 ELD Standards, and the keys to MC items. ETS assessment specialists and psychometricians reviewed the form planners before they were delivered to the CDE for review. The CDE reviewed and approved the form planners after ETS revised the form planners as needed.

2.2.1. Test Design

The Summative ELPAC is administered to the following grade levels and grade spans: kindergarten (K), grade one, grade two, grades three through five, grades six through eight, grades nine and ten, and grades eleven and twelve.

Four domains of English language proficiency (ELP) were assessed in the 2017–18 Summative ELPAC: Listening, Speaking, Reading, and Writing. Students in K and grade one were tested one-on-one in all four domains. Students in grade two were tested one-on-one in the Speaking domain. In the Listening, Reading, and Writing domains, grade two students were tested in small groups of up to 10 students. Students in grade spans three through five, six through eight, nine and ten, and eleven and twelve were tested one-on-one in the Speaking domain and in a group administration in the Listening, Reading, and Writing domains. A proctor assisted the test examiner during test administrations to groups comprised of more than 20 students.

2.2.2. Test Blueprints

Test blueprints were developed to describe the content of the Summative ELPAC. The test blueprints contain four tables with information about the task types in each of the four language domains of Listening, Speaking, Reading, and Writing. Task types are individual items or sets of items that require a student to perform an activity to elicit information about the student's ELP.

The test blueprints provide information about the number of items and points that are administered per task type within each grade level and domain. The Summative ELPAC test blueprints also provide two types of alignment between task types and the standards: "primary" and "secondary." Primary alignment indicates there is a close or strong match in terms of the language knowledge, skills, and abilities covered by both the task type and the standard. Secondary alignment indicates that there is a moderate or partial match between the standard and the item in terms of language knowledge, skills, and abilities.

2.2.3. Test Length

Because the blueprints identify the numbers of items to be tested within each domain, they govern test length. When the Summative ELPAC test blueprints were developed, the goal was to include sufficient numbers of items to provide valid and reliable assessments of ELP, while keeping the administration time at a reasonable level. The number of items increases from K through grade span three through five to make the length of the test appropriate for students as they gain the ability to focus for longer periods of time.

The Summative ELPAC is an untimed test. Estimated administration times were provided in the *Examiner's Manuals,* but only as a basis for planning, because students were allowed as much time as they need to complete their responses in each domain. Additionally, the

testing schedule might have been altered to give students sufficient breaks to avoid fatigue, and testing may be administered over the course of several days.

Test examiners were trained to administer an entire domain in a single sitting except for the Reading and Writing domains at grades three through twelve, which could have been administered in either one or two sittings.

2.2.4. Psychometric Review

All operational items in the 2017–18 Summative ELPAC were field tested in the 2016–17 stand-alone field test. After the administration of the field test, all items underwent statistical analysis. The ETS statistical analysis team used student responses to compile classical item statistics and flag any items that fell outside of acceptable parameters. Assessment specialists reviewed each flagged item and made one of three recommendations:

1. Keep the flagged item as is and classify it as operationally ready
2. Revise the flagged item and classify it as field-test ready for a future form
3. Reject the flagged item and discontinue using it

Items that were classified as operationally ready were used to develop the 2017–18 Summative ELPAC.

After assessment specialists assembled the tests, ETS psychometricians reviewed the statistical characteristics of the tests to ensure that the full range of ELP would be assessed. Tests were revised if needed based on the feedback of ETS psychometricians, and then the tests were submitted to CDE psychometricians for review. ETS assessment specialists made any further revisions needed to the tests to obtain approval from the CDE psychometricians.

2.2.5. CDE Review

The CDE used a two-stage process to review all test materials: (1) request for review (RFR) and (2) request for CDE approval (RFCA). Test materials for review and approval by the CDE included form planners, *Examiner's Manuals*, Test Books, Answer Books, braille versions of *Examiner's Manuals* and forms, and large-print versions of forms. All test materials were approved at RFCA before they were submitted to vendors for reproduction.

For the first stage, ETS initiated the review by submitting an RFR to the CDE. CDE consultants performed the initial RFR review and returned comments and requests for revisions to ETS. ETS staff then revised the documents as requested and returned them to the CDE consultants, who then reviewed the updated materials. If the test materials needed additional revisions, they were returned to ETS for further revisions.

For RFCA, if the CDE consultants approved the test materials during the RFR stage, then the CDE submitted the test materials to the CDE administrator with a request for CDE approval (RFCA). Test materials that were approved with revisions were revised by ETS and resubmitted for approval. Test materials that were not approved needed significant revisions and had to be submitted to the consultants for RFR again before they could be resubmitted for RFCA. Test materials that were approved without edits moved on to the composition phase.

2.3. Test Administration

Standardization and security of the ELPAC is of utmost importance in order to maintain the integrity and validity of the assessment. ELPAC test administration manuals provided information to LEAs and testing personnel on how to efficiently receive, organize, administer, and return test materials for scoring.

2.3.1. Test Security and Confidentiality

All testing materials for the 2017–18 Summative ELPAC—Test Books, Answer Books, and *Examiner's Manuals*—were considered secure documents. Every person having access to test materials was required to maintain the security and confidentiality of the test materials. ETS' Code of Ethics requires that all test information, including tangible materials (e.g., test booklets, test questions, test results), confidential files, processes, and activities are kept secure.

To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI).

In the pursuit of enforcing secure practices, ETS and the OTI strive to safeguard the various processes involved in a test development and administration cycle. For the 2017–18 Summative ELPAC, those processes included the following:

- Test development
- Item and data review
- Item banking
- Transfer of forms and items to the CDE
- Security of electronic files using a firewall
- Printing and publishing
- Test administration
- Test delivery
- Processing and scoring
- Data management
- Transfer of scores via secure data exchange
- Statistical analysis
- Reporting and posting results
- Student confidentiality
- Student test results

2.3.2. Procedures to Maintain Standardization

ELPAC processes were designed so the tests are administered and scored in a standardized manner. ETS took all necessary measures to ensure the standardization of the ELPAC, as described in this section.

2.3.2.1 Test Administration

Roles and responsibilities for each person involved in the ELPAC administration were defined in the *Summative ELPAC Test Administration Manual* (CDE, 2017a). Providing clear definitions and delineation for each role ensured test security and standardized administration. These processes are discussed in more detail in subsection *5.1 Procedures to Maintain Standardization*.

2.3.2.2 Test Directions

A series of instructions compiled in detailed manuals is provided to testing personnel. For the 2017–18 Summative ELPAC, such documents included, but are not limited to, the following:

Examiner's Manuals—These were grade-level manuals used by test examiners to administer the ELPAC to students and were to be followed exactly so that all students have an equal opportunity to demonstrate their level of English proficiency. (Refer to *5.1.4.1 Examiner's Manual* in chapter 5 for more information.)

Summative ELPAC Test Administration Manual—This manual contained test administration procedures for local educational agency (LEA) ELPAC coordinators and site ELPAC coordinators (CDE, 2017a). (Refer to *5.1.4.2 Summative ELPAC Test Administration Manual* in chapter 5 for more information.)

Test Operations Management System (TOMS) Guide for the ELPAC—This manual provided instructions for LEA ELPAC and site ELPAC coordinators to perform tasks in TOMS in support of the program, including providing organization information, adding and managing users, searching and viewing student information, ordering pre-identification labels that contain student demographic data, ordering test materials, viewing reports, and accessing audio files (CDE, 2017b). (Refer to *5.1.4.3 TOMS Guide for the ELPAC* in chapter 5 for more information.)

2.4 Participation

California *Education Code* Section 313 requires LEAs to administer the Initial ELPAC to all eligible students in K through twelve whose primary language is a language other than English. LEAs are required to administer the Summative ELPAC annually to students identified as English learners until they are reclassified as fluent English proficient (RFEP).

Table 2.A.1 through table 2.A.4, in appendix 2.A, provide the number of participants and the percent of participation of all students and select demographic groups for each test during the 2017–18 administration. Note that the data in the *Number Enrolled* column includes students who were enrolled within a grade and eligible for Summative ELPAC during the 2017–18 administration. The *Number Tested* columns include students who tested at the current grade level and exclude off-grade testers and students registered who did not test.

2.4. Accessibility

To ensure a fair and valid testing experience for all students who took the 2017–18 Summative ELPAC, ETS provided accessible versions of the test materials for each ELPAC grade level and grade span. Braille and large-print test materials were available for students who have a Section 504 plan or whose individualized education program (IEP) indicated a need for an accommodated version of the ELPAC.

2.4.1. Resources for Selection of Accessibility Resources

The CDE developed Matrix Four and made the document available on the CDE's website to assist LEAs in understanding the accessible resources that are available for an ELPAC administration. Matrix Four follows a three-tiered accessible approach that includes universal tools, designated supports, and accommodations (CDE, 2018).

The following types of accessibility resources were available to students taking the 2017–18 Summative ELPAC:

- **Universal tools** were available to *all* students on the basis of student preference and selection.
- **Designated supports** were available to *all* students when determined for use by an educator or team of educators—with parent/guardian and student input, as appropriate—or specified in the student's IEP or Section 504 plan.
- **Accommodations** were to be permitted on ELPAC tests to all eligible students if specified in the student's IEP or Section 504 plan.

Accessibility resources allowed all students to show what they know and can do. These resources were not intended to give a testing advantage, but, rather, to allow students the opportunity for a fair and valid testing experience.

2.4.2. Delivery of Accessibility Resources

ELPAC test materials were available in braille and large-print for each ELPAC grade level and grade span. Additionally, as noted previously, Matrix Four outlined the accessibility resources that are permitted during the ELPAC administration.

The percentages of accessibility resources used during the 2017–18 Summative ELPAC field test by grade level and for each domain are presented in appendix 2.B, in table 2.B.1.

2.4.3. Unlisted Resources

Unlisted resources are not universal tools, designated supports, or accommodations. Unlisted resources are made available if specified in an eligible student's IEP or Section 504 plan and only on approval by the CDE (CDE, 2018).

To request the use of an unlisted resource for the 2017–18 Summative ELPAC, the LEA ELPAC coordinator submitted a request to the CDE a minimum of 10 business days before the student's first day of testing. The CDE replied to the request within four business days.

Approval of an unlisted resource that was not previously identified may have been granted by the CDE on the basis of the IEP team's or Section 504 plan's designation and if the unlisted resource did not compromise test security. Prior to administration, the CDE determined if the unlisted resource changed the construct being measured. If so, the LEA ELPAC coordinator was instructed to mark *Alternate Assessment* on the Answer Book for all affected domains. The student received the lowest obtainable scale score for any domains in which *Alternate Assessment* was marked.

References

California Department of Education. (2017a). *ELPAC test administration manual.* Sacramento, CA: California Department of Education. Retrieved from https://www.elpac.org/s/pdf/ELPAC_test-admin-manual-summative.2018.pdf

California Department of Education. (2018). *Matrix Four: Universal tools, designated supports, and accommodations for the English Language Proficiency Assessments for California*. Retrieved from https://www.cde.ca.gov/ta/tg/ep/documents/elpacmatrix4.docx

California Department of Education. (2017b). *Test Operations Management System guide for the ELPAC*. Sacramento, CA: California Department of Education. Retrieved from https://www.elpac.org/s/pdf/ELPAC.TOMS-users-guide.2017-18.pdf

Chapter 3: Item Development and Review

3.1. Item Development

This section describes the work performed to develop a high-quality pool of items for the English Language Proficiency Assessments for California (ELPAC).

3.1.1. Overview

The ELPAC technical proposal from Educational Testing Service (ETS) stated that ETS would undertake several item development efforts during the first ELPAC contract. This section describes the first pilot of ELPAC items, task types that were developed, and the development of the items.

3.1.2. Pilot of ELPAC Items

The 2017–18 Summative ELPAC included task types that were not administered in the California English Language Development Test (CELDT). ETS and the Sacramento County Office of Education (SCOE) conducted a pilot of new task types in the four domains of Listening, Speaking, Reading, and Writing from November 30, 2016, through December 8, 2016.

The purpose of the pilot was to evaluate the quality of the new task types by examining whether they gathered appropriate evidence about the student English-language abilities they were intended to measure as described by the 2012 *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards). The item pilot laid the foundation for development of a high-quality item pool by gathering actionable information for revising early drafts of the test blueprints and the item writing guidelines.

The research questions, pilot procedures, and results were compiled in the *Report on the First Pilot of Items for the ELPAC* (CDE, 2016b). A total of 184 items that represented the range of 26 new task types were developed and administered to students in the pilot. Student responses for certain items were video recorded, while responses to other items were collected and reviewed. Feedback was also collected from test examiners, school administrators, teachers, reading specialists, and students about the viability of the new task types. Based on the pilot results and feedback, ETS assessment specialists implemented revisions to item content and task type directions as well as the rubrics for Speaking and Writing items. The results were also used as a basis for refining the Summative ELPAC test blueprints.

3.1.3. Test Blueprints

The State Board of Education adopted the *Proposed Test Blueprints for the ELPAC* on November 4, 2015, which was prior to the first piloting of the ELPAC items (CDE, 2015). The pilot results provided crucial input for the refinement and streamlining of the Summative ELPAC test blueprints. As ETS and SCOE recommended in the *Report on the First Pilot of Items for the ELPAC* (CDE, 2016b), the ELPAC test blueprints were revised to include those task types that best elicited the types of responses needed to assess students' English language abilities as described by the 2012 ELD Standards.

Before the first pilot, the test blueprints contained a total of 32 proposed task types across the domains of Listening, Speaking, Reading, and Writing. During the evaluation of pilot data, six task types were removed and one was added, making a total of 27 task types in the *Proposed Test Blueprints for the ELPAC* (CDE, 2015).

3.1.4. Task Types

The 2017–18 Summative ELPAC contained 27 task types. Each task type required a student to perform an activity to elicit information about the student's English language proficiency (ELP). Each task type had one or more items that aligned with the 2012 ELD Standards. While the 2012 ELD Standards are organized according to three modes of communication (collaborative, interpretive, and productive communication), federal Title I requirements of the Every Student Succeeds Act of 2015 call for scores to be reported according to the four language domains of Listening, Speaking, Reading, and Writing.

The Listening domain of the Summative ELPAC had five task types, the Speaking domain had six task types, the Reading domain had nine task types, and the Writing domain had seven task types. When a task type required the use of integrated language skills, such as listening and speaking, the task type was classified according to the language skill used to provide the response. For instance, the task type *Summarize an Academic Presentation* required a student to listen to a presentation and then summarize the presentation by speaking to the test examiner. Because the student provided the summary as a spoken response, the task type was classified as a Speaking task type.

The next subsections describe the task types used to assess ELP within each domain of the Summative ELPAC.

3.1.4.1 Listening Task Types

Listening task types assessed the ability of an English learner (EL) to comprehend spoken English (conversations, discussions, and oral presentations) in a range of social and academic contexts. Students listened to a stimulus and then demonstrated their ability to actively listen by answering multiple-choice (MC) questions. For the 2017–18 Summative ELPAC, test examiners used scripts in the *Examiner's Manuals* to read Listening stimuli aloud to K–2 students. Students at grades three through twelve heard audio recordings of the Listening stimuli. The following are descriptions of the stimuli provided for the five Listening task types:

- **Listen to a Short Exchange, kindergarten (K) through grade twelve:** Students hear a two-turn exchange between two speakers and then answer a question about the exchange.
- **Listen to a Classroom Conversation, grades three through twelve:** Students hear a multiple-turn conversation between two speakers and then answer three questions about the conversation.
- **Listen to a Story, K through grade five:** Students hear a multiple-turn conversation between two speakers and then answer three questions about the conversation.
- **Listen to an Oral Presentation, K through grade twelve:** Students hear an oral presentation on an academic topic and then answer three to four questions about it.
- **Listen to a Speaker Support an Opinion, grades six through twelve:** Students hear an extended conversation between two classmates. In the conversation, one classmate makes an argument in support of an opinion or academic topic. After listening to the conversation, students answer four questions.

3.1.4.2 Speaking Task Types

Speaking task types assess the ability of an EL to express information and ideas and to participate in grade-level conversations and class discussions. All task types included one or more constructed-response items. Test examiners scored student responses in the moment using scoring rubrics. The following are descriptions of the six Speaking task types:

- **Talk About a Scene, K through grade twelve:** The student is presented with an illustration of a familiar scene. The test examiner first asks three who-, what-, and when-type questions about the scene. The test examiner then administers three items intended to generate longer responses.
- **Speech Functions, grades two through twelve:** Students state what they would say in a situation described by the test examiner.
- **Support an Opinion, K through grade twelve:** The student listens to a presentation about two activities, events, materials, or objects, and is asked to give an opinion about why one is better than the other. At K and grades one and two and grade span three through five, students view a picture of the choices for context and support.
- **Retell a Narrative, K through grade five:** The student listens to a story that follows a series of pictures, and then the student uses the pictures to retell the story.
- **Present and Discuss Information, grades six through twelve:** The student views a graph, chart, or image that provides information. The student is prompted to read the information and then asked to respond to two prompts. The first prompt asks for a summary of the information in the graph, chart, or image. The second prompt asks for the students to state whether a claim is supported or unsupported based on the information in the graph or chart.
- **Summarize an Academic Presentation, K through grade twelve:** The student listens to an academic presentation while looking at a related picture(s). The student is prompted to summarize the main points of the presentation using the illustration(s) and key terms of the presentation, if provided.

3.1.4.3 Reading Task Types

Reading task types assessed the ability of an EL to read, analyze, and interpret a variety of grade-appropriate literary and informational texts. The following are descriptions of the nine Reading task types:

- **Read-Along Word with Scaffolding, K:** With scaffolding from the test examiner, the student provides the individual letter names and the initial letter sound for a decodable word. The student then answers a comprehension question about the word.
- **Read-Along Story with Scaffolding, K through grade one:** The student listens and follows along as the test examiner reads aloud a literary text accompanied by three pictures for context and support. The student then answers a series of comprehension questions about the story.
- **Read-Along Information, K through grade one:** The student listens and follows along as the test examiner reads aloud an informational text accompanied by three pictures for context and support. The student then answers a series of comprehension questions about the information.

- **Read and Choose a Word, grades one through two:** The student reads three words and chooses the word that matches a picture.
- **Read and Choose a Sentence, grades one through twelve:** The student reads three or four sentences and chooses the sentence that best describes a picture.
- **Read a Short Informational Passage, grades one through twelve:** The student reads a short informational text and answers MC questions related to the text.
- **Read a Student Essay, grades three through twelve:** The student reads an informational essay presented as if written by a peer and answers a set of MC questions related to the essay.
- **Read a Literary Passage, grades one through twelve:** The student reads a literary text and answers MC questions related to the text.
- **Read an Informational Passage, grades one through twelve:** The student reads an informational text and answers MC questions related to the text.

3.1.4.4 Writing Task Types

Writing task types assessed the ability of an EL to write literary and informational texts to present, describe, and explain information. The following are descriptions of the seven Writing task types:

- **Label a Picture—Word, with Scaffolding, K through grade one:** With scaffolding from the test examiner, the student writes labels for objects displayed in a picture.
- **Write a Story Together with Scaffolding, K through grade two:** With scaffolding from the test examiner, the student collaborates with the test examiner to jointly compose a short literary text by adding letters, words, and a sentence to a story.
- **Write an Informational Text Together, grades one through two:** With scaffolding from the test examiner, the student listens to a short informational passage and then collaborates with the test examiner to jointly compose a text about the passage by writing a dictated sentence and an original sentence about the topic.
- **Describe a Picture, grades one through two:** The student looks at a picture and writes a brief description about what is happening.
- **Describe a Picture, grades three through twelve:** The student looks at a picture and is prompted to examine a paragraph written by a classmate about what is happening in the picture. The student is asked to expand, correct, and combine different sentences written by a classmate before completing the final task of writing a sentence explaining what the students will do next.
- **Write About an Experience, grades three through twelve:** The student is provided with a common topic, such as a memorable classroom activity or event, and is prompted to write about the topic.
- **Write About Academic Information, grades three through twelve:** The student interprets academic information from a graphic organizer created for a group project and answers two questions about it.
- **Justify an Opinion, grades three through twelve:** The student is asked to write an essay providing a position and appropriate supporting reasons about a school-related topic.

3.1.5. Updates to Item Writing Guidelines

The first pilot of the ELPAC items provided a wealth of experience with new ELPAC task types that informed subsequent item-writer training and item-development efforts. ETS assessment specialists used data from the pilot to refine task types and develop descriptions of the ELPAC task types in the *Item Writing Guidelines for the ELPAC* (CDE, 2016a). These guidelines were used to train California educators to develop additional items for the ELPAC item pool at the Item-Writer Training for California Educators from February 22, 2016, through February 25, 2016.

3.2. Item Review Process

3.2.1. Overview

In partnership with SCOE, ETS convened ELPAC item-writer trainings and item review panels to develop test items for both the Initial ELPAC and the Summative ELPAC. Select California educators were trained to write new items for the ELPAC. In addition, ETS trained a small group of experienced contractors to draft ELPAC items. After the items went through ETS internal and CDE reviews, California educators reviewed the items during Content Review Panel and Bias and Sensitivity Review Panel meetings. This subsection describes how California educators were selected and the process used to develop items for the ELPAC.

3.2.2. Selection of Item Writers

California educators were recruited through email communications and by letter. To ensure broad representation, an email message and letter announcing the opportunities to write items and to review items were sent by the CDE to the following groups:

- The CDE's ELPAC listserv (includes California English Language Development Test District Coordinators and Title III county leads)
- The Bilingual Coordinators Network
- The CDE's California Assessment of Student Performance and Progress Coordinator listserv
- The CDE's All Assessment listserv
- The ELPAC Technical Advisory Group

The email and letter directed applicants to fill in an online application in SurveyMonkey, a third-party, online survey provider. The application allowed California educators to apply for any or all of the events. The information from the application was loaded into a database that was used for the review and selection process.

During the selection process, applications were selected from current and retired California educators who had the following minimum qualifications:

- Bachelor's degree
- Expertise in language acquisition or experience teaching ELs in K through grade twelve
- Knowledge of and experience working with the 2012 ELD Standards

Additional desirable qualifications included:

- A teaching credential authorization for English language development, specially designed academic instruction in English, or content instruction delivered in the primary language (e.g., Cross-cultural, Language, and Academic Development Certificate; or Bilingual, Cross-cultural, Language, and Academic Development Certificate)
- Specialized teaching certification in reading (e.g., Reading Certificate or Reading and Language Arts Specialist Certificate)
- Experience writing or reviewing test items for standardized tests, especially tests for ELs in K through grade twelve
- Recent experience administering the CELDT

Selections were made to ensure representation from different cultural and linguistic groups, various-sized local educational agencies (LEAs) and county offices of education, and different geographical regions of the state, and with regard to the travel budget allowable in the contract. ETS and SCOE made preliminary selections, which were reviewed by the CDE, adjusted as needed, and then approved. Forty-two educators were selected for item-writer training, along with 14 alternates. Forty-two educators were selected for Content Review Panels, along with 14 alternates. Ten educators were selected for Bias and Sensitivity Review Panels, along with three alternates.

SCOE contacted and invited the participants and contacted the alternates as necessary. Once all participants confirmed, SCOE notified those who were not selected.

3.2.3. Item Writing by Educators

Item-writer training was divided into two sets of meetings, each of which lasted two days. Twenty-four educators from K through grade five were trained on Monday and Tuesday, February 22 and 23, 2016. Eighteen educators from grades six through twelve were trained on Wednesday and Thursday, February 24 and 25, 2016. The educators represented a mix of rural, suburban, and urban LEAs.

3.2.3.1 Introduction to Item Writing

During each of the two-day meetings, educators received training and then drafted ELPAC items. At the start of day one, a PowerPoint (PPT) presentation was used to provide information to the educators about topics regarding the ELPAC and item development. Topics covered during the presentation included an overview of the ELPAC, general principles of item development, a review of the 2012 ELD Standards, the overall item development process, and the process for drafting and submitting items. After the PPT presentation, ETS trainers provided educators with examples of task types that are shared across grade levels and grade spans.

ETS trainers facilitated brainstorming sessions, during which educators listed topics that served as a basis for item development. Educators were asked to propose topics for item content that is covered during prior grades to ensure that topics were appropriate. After brainstorming, educators worked as a whole group to assign topics to appropriate grade levels or grade spans. Educators then split up into grade-level groups to draft items corresponding to the topics from their brainstorming session. This pattern was followed for all domains (Listening, Speaking, Reading, and Writing).

3.2.3.2 Process

After educators divided into their grade-level groups, ETS trainers provided them with *Item Writing Guidelines for the ELPAC* (CDE, 2016a), sample items, and item templates. The *Item Writing Guidelines for the ELPAC* provided details about the type of information that is required when drafting items, such as the length of any Listening stimuli or Reading passages, the number of items within the set, and the types of English language knowledge, skills, and abilities to be assessed by the items.

The sample items were developed by ETS assessment specialists to serve as examples of the task types to be developed. The item templates were Word files that contained areas for entering information. The item templates assured that items were drafted in a standardized manner and that all needed item information was entered. ETS trainers used the *Item Writing Guidelines for the ELPAC*, sample items, and item templates as training materials to provide clear expectations regarding the information needed when drafting each task type, as well as the level of quality that was expected.

3.2.3.3 Assignment

ETS trainers gave educators item writing assignments to be completed during the two-day training. ETS trainers remained within the training rooms when educators were drafting items to answer questions and to provide feedback regarding initial drafts of items.

Educators were also given the opportunity to take an item writing assignment to be completed in the weeks after the two trainings. They were provided with the printed training materials needed to complete the assignment and given two weeks to complete their assignments. Educators were required to return all secure printed training materials at the time their assignments were submitted. Secure printed training materials were returned via secure express delivery.

To submit assignments, educators saved their assignments in password-protected files and copied them to a secure ETS server. After ETS confirmed receipt of the files, educators were prompted to delete the files from their personal devices.

3.2.4. Item Writing by Contractors

ETS assessment specialists worked with five contractors (i.e., outside item writers) who are fully trained, experienced item writers with a record of developing quality items for other ETS English language assessments. Since there was a limited amount of time to train California educators to develop Listening and Reading sets, ETS contractors developed the Listening task types with relatively long stimuli and the Reading task types with relatively long passages. The focus of the contractors was to develop the following task types:

- *Listening—Listen to a Story*
- *Listening—Listen to an Oral Presentation*
- *Reading—Read a Literary Passage*
- *Reading—Read an Informational Passage*

The contractors delivered all items to a secure ETS server. After ETS confirmed receipt of the files, contractors were prompted to delete the files from their personal devices.

3.2.5. Item Review Panels

Before ELPAC items were designated as field-test ready, the draft versions underwent a thorough ETS internal review process, including two content reviews, a fairness review, and an editorial review; external reviews by item review panels; and a CDE review and final

approval. This section describes the reviews conducted by two sets of item review panels: the Content Review Panels and the Bias and Sensitivity Review Panels.

To help establish content validity for the ELPAC and to develop test materials that are fair to all students, ELPAC test items were reviewed by a Content Review Panel and a Bias and Sensitivity Review Panel during the week of August 1, 2016, through August 5, 2016. Content Review Panels reviewed items to ensure that items were aligned with the 2012 ELD Standards, items were appropriate for the grade level or grade span, items addressed the construct being tested, and selected-response items had one and only one correct answer. Bias and Sensitivity Review Panels reviewed items to ensure that they did not contain content that would result in bias to identified groups or that is offensive.

3.2.5.1 Meeting Plan and Training

The Content Review Panel meeting began on August 1, 2016, and finished on August 5, 2016. The Bias and Sensitivity Review Panel meeting began on August 3, 2016, and finished on August 5, 2016.

The ETS technical proposal in response to the ELPAC request for proposals stated that there would be enough time between the two panel reviews to allow revisions from the Content Review Panel to be applied to items before the Bias and Sensitivity Review Panel review. However, the plan to apply revisions to the items between the panel reviews was revised after work on the 2015–18 ELPAC contract was delayed during a 71-day period during which the contract award was protested.

After the protest period ended, the CDE and ETS agreed to hold the two panel meetings on an overlapping schedule within a single week. This approach allowed ELPAC items to be developed on time for stand-alone sample field testing in 2016–17 while ensuring that appropriate procedures were followed to produce a high-quality pool of items.

Two trainings for the panel participants were conducted during the meetings and prior to the item reviews: educators serving on the Content Review Panel were trained on Monday, August 1, 2016, and educators serving on the Bias and Sensitivity Review Panel were trained on Wednesday, August 3, 2016.

3.2.5.2 Process

The Bias and Sensitivity Review Panel members reviewed items as revised by the Content Review Panel. Members of the Bias and Sensitivity Review Panels needed to read and understand the comments of the Content Review Panel before providing comments on bias and sensitivity issues. Facilitators were responsible for transferring comments from the Content Review Panels to the Bias and Sensitivity Review Panels during designated times. Notetakers projected the Content Review Panel comments on a screen to allow members of the Bias and Sensitivity Review Panels to read them.

Facilitators monitored the progress of the panel reviews to ensure all items were reviewed by the last day of the panel meetings. As planned, the Content Review Panel members finished their reviews by noon on Friday, August 5. This allowed the Bias and Sensitivity Review Panel facilitators to retrieve the final Content Review Panel comments during the lunch break on August 5. Bias and Sensitivity Review Panel members then completed their reviews of the final items by the end of their sessions on August 5.

3.2.5.3 Outcome

Educators at both the Content Review Panel meeting and the Bias and Sensitivity Review Panel meeting had the option of making one of three decisions regarding each stimulus and item: approve as is, approve with revisions, or reject.

Table 3.1 provides the status of the stimuli and items after the 2016 item review panel meetings.

Table 3.1 Status of Stimuli and Items After the 2016 Item Review Panel Meetings

Grade Level or Grade Span	Approve As Is	Approve with Revisions	Reject
Kindergarten	202	91	0
Grade 1	213	87	0
Grade 2	240	90	5
Grade span 3–5	284	90	1
Grade span 6–8	216	132	0
Grade span 9–10	253	118	2
Grade span 11–12	270	88	12
Totals:	**1,678**	**696**	**20**

After the item review panel meetings, the CDE reviewed the proposed revisions to items, made any adjustments needed, and then approved the revisions.

3.3. Item Banking

The ETS Item Banking Information System (IBIS) was used as the database of record throughout the item-development process. IBIS was used to store item text, graphics, scripts for audio recordings, scoring information, and metadata. After ETS assessment development staff drafted and reviewed items in IBIS, the CDE used IBIS to review items in preparation for item review panels. After the CDE approved proposed revisions from the item review panel meetings, CDE staff confirmed the items in IBIS to ensure that revisions were implemented correctly before the items were approved for field testing.

References

California Department of Education. (2014). *California English Language Development Standards: Kindergarten through grade 12*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

California Department of Education. (2016a). *Item writing guidelines for the English Language Proficiency Assessments for California* (Unpublished report). Sacramento, CA: California Department of Education.

California Department of Education. (2015). *Proposed test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/ta/tg/ep/documents/elpacblueprts.pdf

California Department of Education. (2016b). *Report on the First Pilot of Items for the ELPAC* (Unpublished report). Sacramento, CA: California Department of Education.

Chapter 4: Test Development

4.1. Test Design

This chapter describes the development of the 2017–18 Summative English Language Proficiency Assessments for California (ELPAC) forms, including the revisions to the Summative ELPAC test blueprints based on the field test results, the rules for item selection, the structure of the test forms, and the development of the test materials. Each form of the Summative ELPAC assesses the four domains of Listening, Speaking, Reading, and Writing. All items included on the 2017–18 Summative ELPAC were administered first in a stand-alone field test. Refer to *Chapter 11: Field Testing* for more details about the fall 2017 stand-alone field test.

4.1.1. Revision of the Test Blueprints

All items included on the 2017–18 Summative ELPAC were administered in a stand-alone field test. After the administration of the stand-alone field test, items went through statistical analysis and the *Proposed Test Blueprints for the ELPAC* (California Department of Education [CDE], 2015) were revised. Based on the statistical performance of the items in the stand-alone field test, Educational Testing Service (ETS) adjusted the number of items in the test blueprint.

The State Board of Education (SBE) had adopted the *Proposed Test Blueprints for the ELPAC* on November 4, 2015, which was prior to the first pilot test of items. Revisions to the test blueprints from the first pilot of items and the stand-alone field test were compiled and then presented to the SBE for review. The SBE approved and adopted the updated *Summative Assessment Test Blueprints for the ELPAC* on September 14, 2017 (CDE, 2017b).

The next two subsections provide an overview of the analyses performed in making the decisions for the final blueprint.

4.1.1.1 Statistical Analysis

The Summative ELPAC stand-alone field test was held from March 6, 2017, through April 14, 2017. After the administration, all items from the stand-alone field test underwent statistical item analysis. The ETS statistical analysis team used student responses to compile item statistics and flagged any items that fell outside of acceptable parameters. Assessment specialists reviewed each flagged item and made one of three recommendations:

1. Keep the flagged item as is and classify it as operationally ready
2. Revise the flagged item and classify it as field-test ready for a future form
3. Reject the flagged item and discontinue using it

After the field test items went through statistical item analysis, ETS delivered the item analysis results to the CDE.

As part of the test design process, an evidence-centered design (ECD) approach was used. ECD is a principled framework that “ensures that the way in which evidence is gathered and interpreted bears on the underlying knowledge and purposes the assessment is intended to address” (Mislevy, Steinberg, & Almond, 1999, p. 1). Through this approach the performance of the Summative ELPAC task types was reviewed. Those task types that

were most appropriate for use in the upcoming operational assessment were retained at each grade.

4.1.1.2 Analysis of Results

The overall number of Summative ELPAC task types remained at 27, but adjustments were made to the number of task types and items at each grade level or grade span. This was particularly the case in the Writing domain at kindergarten (K), grade one, and grade two. The Summative ELPAC test blueprints for Writing were adjusted to include fewer Writing items than were included in the stand-alone field test. The reasons for this were as follows:

- Avoid a Writing domain that is overly burdensome on students and test examiners
- Ensure that the Writing domain elicits appropriate evidence of students' skills in relation to the 2012 *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards), reflecting information learned about each task type from the field test
- Ensure that the Writing domain contributes appropriately to valid and reliable score reporting

4.2. Item Selection

The development of the Summative ELPAC necessitated fulfilling the requirements of the test blueprints as well as meeting the statistical and psychometric criteria specified, as described in this section.

4.2.1. Test Development Specifications

The development of the 2017–18 Summative ELPAC began with the creation of test development specifications. ETS created the test development specifications that the CDE reviewed and approved after revision. The test development specifications for the 2017–18 Summative ELPAC described the goals of the assessment, the content criteria for selecting items, the psychometric criteria for selecting items, the test development process, and a timeline for major activities.

The 2017–18 Summative ELPAC consisted of newly constructed tests of previously field-tested items that were approved as operationally ready and included one form at each of the seven grade levels and grade spans. Each form assessed all four domains of Listening, Speaking, Reading, and Writing. Items from the 2016–17 stand-alone field test were used, as mentioned previously.

The 2017–18 Summative ELPAC did not include any embedded field test items because the 2016–17 stand-alone field test provided enough operationally ready items to allow for a 30 percent refresh of items during the development of the 2018–19 Summative ELPAC.

4.2.2. Content Criteria

Test validity requires that content coverage adheres to test blueprints. The blueprints specify the number of items from each task type to include in each domain and which 2012 ELD Standards are assessed in each domain. ETS assessment specialists used the *Summative Assessment Test Blueprints for the ELPAC* (CDE, 2017b) as the basis to select task types and items for the 2017–18 Summative ELPAC. Assessment specialists selected items that covered a variety of content areas and topics to ensure that balanced forms were created.

ETS assessment specialists used IBIS to develop form planners for the 2017–18 Summative ELPAC. A form planner is an Excel spreadsheet that contains information about each of the items included in a test form. The form planners include information such as the item's accession number (i.e., the unique item identification code), grade, domain, correct answer (for multiple-choice items), score scale (for constructed-response items), and alignment to the 2012 ELD Standards. After form planners were created, ETS reviewed them internally. An ETS assessment specialist who did not participate in test assembly performed a full review of each test form to ensure that an appropriate set of items was selected. After this review was completed, the form planners were delivered to ETS psychometricians for review.

4.2.3. Statistical and Psychometric Criteria

The statistical specifications provided guidelines for selecting items and developing tests with appropriate psychometric properties. Since this was the first operational administration, statistics from the 2016–17 stand-alone field test were used to inform the development of the 2017–18 Summative ELPAC.

Each ELPAC conformed with the following psychometric criteria:

- Individual items had *p*-values—a measure of item difficulty—that ranged from 0.20 to 0.95.
- The collection of items within each domain represented an overall difficulty level with average *p*-values from 0.5 to 0.7.
- Point-biserial correlations—a measure of reliability—for each item was greater than 0.15.
- Differential item functioning (DIF) analyses were conducted to detect possible test bias and locate items for which one group of students performed significantly better than another group of students of similar ability. Items flagged for DIF may be measuring something other than the intended construct. Items with C-level DIF flags, the most severe, were not used.

ETS assessment specialists assembled the 2017–18 Summative ELPAC test forms based on the classical statistics obtained from the 2016–17 stand-alone field test. ETS psychometricians then reviewed the composition of the test forms and compiled distribution tables, which showed the distribution of items according to difficulty, to ensure that correct numbers and distributions of items were selected. Having a broad distribution of item difficulties ensured that there was reasonable measurement power across the range of difficulty.

Although form evaluation and approval were based on classical item statistics, item response theory (IRT) calibrations were conducted after the dimensionality study that was conducted in summer 2017. The two-parameter logistic (2PL) and generalized partial credit (GPC) IRT model combination was selected for item calibration because it offers flexibility in its ability to estimate a range of item discriminations (Smarter Balanced, 2015). Flexible estimation is useful for vertical scaling and avoids the stability problems encountered by a three-parameter logistic model (Holland, 1990). The 2PL/GPC IRT model combination was used in the calibration, and item parameters were used to construct the number-correct-to-scale-score tables so that preequated scoring tables were available for the 2017–18 Summative ELPAC.

Finally, after ETS psychometricians reviewed the composition of the 2017–18 Summative ELPAC test forms, ETS assessment specialists revised the composition of the test forms based on psychometric review, as needed.

4.2.4. CDE Review of Item Selection

After revisions were made to the form planners during internal ETS reviews, the form planners and distribution tables were delivered to the CDE for review. CDE staff had access to item content and metadata via IBIS, through which they reviewed the item content, form planners, and distribution tables.

The CDE made recommendations for replacing items within the test forms. ETS adjusted the form planners as needed and then submitted the revised form planners to the CDE for review and approval.

4.3. Forms Development

This section describes the development of the paper-based test materials, including the production of audio recordings for Listening and Speaking items and the development of the large-print and braille versions as well as the breach edition.

4.3.1. Developing Paper-Based Test Materials

This subsection describes the development of audio recordings and paper-based test materials for the 2017–18 Summative ELPAC.

4.3.1.1 Audio Recordings

ETS worked with professional recording studios and voice actors to develop all audio recordings used at grades three through twelve for Listening items and *Speaking—Summarize an Academic Presentation* items. All audio recordings for Listening operational items were developed prior to the field test administration according to both the quality standards established during the development of a demonstration reel and the CDE confirmation of the field test recordings.

The item-level audio recordings from the field test administrations were used to develop the 2017–18 Summative ELPAC Listening test forms for grades three through twelve. Based on feedback from the field test, professional audio recordings of the grades three through twelve *Speaking—Summarize an Academic Presentation* items were also developed according to the same standards that were established during the development of the demonstration reel.

After the form planners for each test form of the 2017–18 Summative ELPAC were approved, test-length audio files were developed for the Listening domain at grades three through twelve. To prepare for the development of the test (domain)-level audio files, item-level scripts were compiled to create test-length scripts, including section directions, task type directions, practice items, and operational items. The test-length scripts and item-level audio recordings were delivered to a professional studio for compilation.

After the studio compiled the item-length recordings into test-length recordings, ETS proofed the test-length audio recordings to ensure they were compiled accurately. The grades three through twelve *Speaking—Summarize an Academic Presentation* audio files were compiled separately because they were the only audio files for the Speaking domain.

4.3.1.2 Paper-based Test Materials

After the form planner for each 2017–18 Summative ELPAC test form was approved, ETS assessment specialists delivered the form planners and item content to the ETS production team. ETS production staff used the instructions provided by the assessment specialists to compile the item content and create the paper-based test materials. The collaboration between the two teams resulted in the development of all paper-based test materials, including seven *Examiner's Manuals* (all grade levels and grade spans), seven Test Books (all grade levels and grade spans), and four Answer Books (grade spans three through five, six through eight, nine and ten, and eleven and twelve).

After the ETS production teams composed the paper-based test materials, the materials were subject to internal ETS reviews before they were delivered to the CDE for review and approval.

4.3.2. Developing Special Version Forms

4.3.2.1 Braille

The goal of the ELPAC braille versions of the forms is to provide valid and reliable measurement of ELP for students who use braille by including scoring tables with the same performance level cut scores as the standard version of the assessment. This subsection describes the development of the braille forms used during the 2017–18 Summative ELPAC administration.

ETS assessment specialists collaborated with members of the ETS Alternate Test Form (ATF) team to develop the braille forms.

4.3.2.1.1. Criteria

A foundational step in developing the braille forms was to review the ELPAC task types at a high level and determine which task types were amenable to braille, which needed to be revised to become amenable to braille, and which were not amenable to braille. Each ELPAC task type was analyzed for suitability for administration in a braille form. Solutions were proposed at the task-type level.

In developing the proposed solutions, the ETS team endeavored to minimize changes needed to task types to make them accessible to ELs with visual impairment. Any necessary adaptations were designed to preserve the target construct and measure the same ELP standards and targeted performance level descriptors.

ETS staff analyzed the Summative ELPAC task types and documented the process to be used to develop braille and large-print versions in the *Process for Development of Special Test Versions* (CDE, 2017a). The CDE reviewed and approved the document before ETS began development of the braille versions.

To begin, ETS reviewed individual items to ensure that item content was sensitive to the experiences of ELs with visual impairment. Reviews of individual items were also needed to confirm that the cognitive load of the braille item remained comparable with the original item. Once items were selected for the 2017–18 Summative ELPAC, ETS staff reached agreement on the adaptations needed for the braille forms. For those items needing adaptations, variants of the items were created in IBIS, adapted, and then reviewed in IBIS to confirm accuracy.

4.3.2.1.2. Process

After items were adapted for the braille forms, ETS provided the braille vendor with the information needed to produce the braille forms. Before the braille forms were produced, ETS communicated with the braille vendor to confirm the exact specifications for the final deliverable according to the *Rules for Unified English Braille* (2013) set forth by the Round Table on Information Access for People with Print Disabilities Inc. and International Council on English Braille, a collective that includes the Braille Authority of North America. The braille vendor was responsible for ensuring the quality of the braille forms. The quality control measures included two proofs of all test materials.

In addition to reviewing the ELPAC task types, instructions for test examiners, such as test administration and domain-specific procedures, were reviewed and adapted as needed for the braille administration. Test developers, in consultation with the ATF team, reviewed and adapted the directions by grade level and grade span in parallel with item-level evaluation. Final directions were provided to the CDE for approval prior to certification.

4.3.2.2 Large-Print

The goal of the ELPAC large print versions of forms is to provide valid and reliable measurement of ELP for students who use large-print materials. This subsection describes the development of the large-print forms used during the 2017–18 Summative ELPAC administration.

ETS assessment specialists collaborated with members of the ETS ATF team to develop the large-print forms.

4.3.2.2.1. Criteria

Form 1 of the 2017–18 Summative ELPAC was selected for use as the large-print ELPAC. ETS assessment specialists worked with the ATF group to agree upon the content to be enlarged.

Marked-up content was delivered to page composers, who created the large-print versions. The assessment specialists and ATF team then reviewed the large-print forms, requesting any revisions needed before the materials were delivered to the large-print vendor. The vendor produced proofs, which ETS then reviewed. The vendor made any requested adjustments before the large-print forms were submitted to the CDE for review.

4.3.2.2.2. Process

All student-facing test content was enlarged to develop the large-print forms. Most of the student-facing content was found in the Test Books and Answer Books, although some was in a grade-level *Examiner's Manual*. Any student-facing content that was in an *Examiner's Manual* was enlarged and placed in the Answer Book at K through grade two and in the Test Book at grades three through twelve.

ETS enlarged all student-facing text to 18- to 20-point font in grades two through twelve. In K and grade one, where the font size is already 18 points or larger, the font size was increased by four points, (e.g., 18-point text was increased to 22 points in the large-print forms).

4.3.2.3 Breach Edition

After the 2017–18 Summative ELPAC was developed, a breach form was developed for each grade level and grade span for use in a breach edition. The breach edition was developed as a print-ready edition that could be printed and distributed in the case of large-scale, public exposure of an operational ELPAC form. Items that were approved as

operationally ready during the item analysis of the 2016–17 stand-alone field test were used to populate the forms in the breach edition. To use the ELPAC item pool efficiently, approximately 30 percent of the items were shared between the 2017–18 Summative ELPAC and the breach edition.

As there were no security breaches during the 2017–18 Summative ELPAC administration, the breach form was not used.

4.3.3. CDE Review of Assembled Forms

After revisions were made to the form planners during internal ETS reviews, the form planners and distribution tables were delivered to the CDE for review. Using the metadata available in IBIS, the CDE reviewed the item content, form planners, and distribution tables. The CDE made recommendations for replacing items within the test forms. ETS adjusted the form planners as needed and then submitted the revised form planners to the CDE for review and approval.

References

California Department of Education. (2014). *California English Language Development Standards: Kindergarten through grade 12*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

California Department of Education. (2017a). *Process for development of special test versions*. Sacramento, CA: California Department of Education.

California Department of Education. (2017b). *Summative assessment test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from https://www.cde.ca.gov/ta/TG/ep/documents/elpacadoptedblueprints.pdf

California Department of Education. (2015). *Proposed test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/ta/tg/ep/documents/elpacblueprts.pdf

Holland, P. W. (1990). *The Dutch identity: A new tool for the study of item response models.* Psychometrika, 55, 5–18.

Mislevy, R. J., Steinberg, L. S., & Almond, R. (1999). *Evidence-centered assessment design*. Princeton, NJ: Educational Testing Service.

Round Table on Information Access for People with Print Disabilities Inc. and International Council on English Braille. (2013). *The rules of Unified English Braille second edition 2013*. Retrieved from http://www.iceb.org/Rules%20of%20Unified%20English%20Braille%202013%20(linked).pdf

Smarter Balanced Assessment Consortium (2016). *Smarter Balanced Assessment Consortium: 2013–14 Technical Report.* Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from https://portal.smarterbalanced.org/library/en/2013-14-technical-report.pdf

Chapter 5: Test Administration

5.1. Procedures to Maintain Standardization

To maintain standardization during the 2017–18 English Language Proficiency Assessments for California (ELPAC) administration, local educational agencies (LEAs) and ELPAC staff were provided with several forms of communication and training. Educational Testing Service (ETS) produced and provided the *ELPAC Test Administration Manual*, which detailed the process and policies for a secure and standardized administration, as well as other quick-reference guides describing various aspects of ELPAC administration. Additionally, the Sacramento County Office of Education (SCOE) provided several trainings across the state to site ELPAC coordinators and ELPAC test administrators. These trainings provided a hands-on opportunity for participants to learn about and ask questions regarding ELPAC administration. SCOE also provided training for test examiners who administered the Speaking and Listening sections of the ELPAC.

5.1.1. LEA ELPAC Coordinator

An LEA ELPAC coordinator was designated by the district superintendent at the beginning of the 2017–18 school year. LEAs include public school districts, statewide benefit charter schools, State Board of Education–authorized charter schools, county office of education (COE) programs, and direct funded charter schools.

LEA ELPAC coordinators were responsible for ensuring the proper and consistent administration of the ELPAC. In addition to the responsibilities set forth in *California Code of Regulations (CCR)*, Title 5, Section 11518.40, their responsibilities included

- adding ELPAC site coordinators and test examiners into the Test Operations Management System (TOMS);
- reporting test security incidents (including testing irregularities) to the CDE;
- ensuring that correct testing procedures were followed;
- ensuring that test materials were distributed to the schools and kept in a locked, secure area at all times
- ensuring that all test examiners were trained and certified to administer the Summative ELPAC;
- ordering test materials, pre-identification (Pre-ID) labels, and supplemental test materials in TOMS;
- ensuring adequate test materials were on hand and redistributed throughout the LEA during the testing window as needed;
- shipping all materials back for scoring; and
- requesting rescores through TOMS.

The 2017–18 LEA ELPAC coordinator was required to sign the ELPAC *Test Security Agreement* (5 *CCR* 11518.50[b]).

5.1.2. Site ELPAC Coordinator

A site ELPAC coordinator was trained by the LEA ELPAC coordinator for each test site (5 *CCR* Section 11518.40[b][7]). The 2017–18 site ELPAC coordinator was required to sign both the ELPAC *Test Security Agreement* and the ELPAC *Test Security Affidavit* (5 *CCR* Section 11518.45[b][3]).

In addition to the responsibilities set forth in 5 *CCR* Section 11518.45, their responsibilities included

- identifying test examiners and ensuring that they signed ELPAC *Test Security Affidavits*;
- adding test examiners into TOMS;
- managing ELPAC testing at the school;
- ensuring the proper administration of all testing procedures;
- maintaining the security of all test materials at the site; and
- assuring the proper packing and return of test materials to the LEA ELPAC coordinator.

5.1.3. Test Examiner

Test examiners were identified by ELPAC site coordinators as individuals who administered the Summative ELPAC and were an employee or contractor of an LEA. A test examiner was proficient in English with complete command of pronunciation, intonation, and fluency, and certified that he or she had completed training in the administration and scoring of the ELPAC. Proctors assisted test examiners during group administration.

Prior to handling testing materials, a test examiner and any other individual handling 2017–18 Summative ELPAC testing materials was required to sign a *Test Security Affidavit* (5 *CCR* Section 11518.50[d]), which was provided at the Administration and Scoring Training workshop and also available on the ELPAC Forms web page at https://www.elpac.org/test-administration/forms/.

A test examiner's duties may have included

- ensuring the physical conditions of the testing room met the criteria for a secure test environment;
- viewing student information in their local student information system prior to testing to ensure that the students' English Language Acquisition Status was EL;
- reporting all test security incidents to the ELPAC site coordinator and LEA ELPAC coordinator in a manner consistent with ELPAC, state, and LEA policies; and
- fully complying with all directions provided in the *Examiner's Manual*.

5.1.4. Instructions for Test Administration

5.1.4.1 *Examiner's Manuals*

These were grade-level or grade-span manuals that described the standardized testing procedures used by test examiners to administer the 2017–18 Summative ELPAC to students. Test examiners were required to follow the procedures in the manuals so that all students were given an equal opportunity to demonstrate their English language proficiency.

The *Examiner's Manuals* provided directions and guidelines for filling in student demographic information on each student Answer Book prior to the test if the LEA did not use the Pre-ID service.

During the test, test examiners read, word-for-word, the directions and scripts for administration. Test examiners also used the Speaking rubrics and anchor samples in the *Examiner's Manual* to evaluate students' responses and assign scores. At grades three through twelve, where recorded audio is played for the Listening domain and for *Speaking—Summarize an Academic Presentation*, the manuals described procedures for playing the recorded audio.

5.1.4.2 *Summative ELPAC Test Administration Manual*

The *Summative ELPAC Test Administration Manual* (CDE, 2017a) contained information and instructions on overall procedures and guidelines for all LEA and test site staff involved in the administration of the ELPAC. Sections included the following topics:

- Dates for ordering materials and testing
- Roles and responsibilities of those involved with ELPAC testing
- Test administration resources
- Test security
- Administration preparation and planning
- General test administration
- Instructions for steps to take before, during, and after testing
- Guidelines for handing materials

5.1.4.3 TOMS Guide for the ELPAC

TOMS is a web-based application that allowed LEA ELPAC coordinators to add and manage users and order materials for the Summative ELPAC. In 2017–18, test examiners used TOMS to play the audio recordings used during the Listening and Speaking portions of the ELPAC in grades three through twelve.

TOMS modules used for Summative ELPAC administration that are described in the *TOMS Guide for the ELPAC* included the following (CDE, 2017c):

- **Adding and Managing Users—**Allowed LEA ELPAC coordinators to add ELPAC test site coordinators and test administrators to TOMS so that the designated user could administer, monitor, and manage the ELPAC
- **Ordering Test Materials—**Allowed LEA ELPAC coordinators to approve orders, view summary orders, view and track orders, and place supplemental orders within specified windows
- **Ordering Pre-ID Labels—**Allowed LEA ELPAC coordinators to request Pre-ID labels that were affixed to Answer Books and used to track student testing and assign results within specified windows
- **Playing Audio Modules—**Allowed test examiners access to the audio files that were part of the Listening and Speaking portions of the ELPAC in grades three through twelve

5.2. Training

SCOE provided several trainings across the state to site ELPAC coordinators and ELPAC test administrators. These trainings provided a hands-on opportunity for participants to learn about and ask questions regarding ELPAC administration. SCOE also provided training for test examiners who administered the Speaking and Listening sections of the ELPAC.

5.2.1. General Test Administration

The online Moodle Training Site was developed as a restricted site that could be accessed only by LEA trainers and others requiring general training in the administration of the ELPAC. (Moodle is a free learning management open-source software.) The site contained all resources needed to conduct a training such as training presentations along with the presenters' scripts.

5.2.2. Scoring Training of Trainers Workshops

All LEAs in California were required to send a trainer to the all-day California Department of Education (CDE)–sponsored statewide 2017–18 ELPAC Summative Assessment Administration and Scoring Training (AST), which employed the "training-of-trainers" model.

5.2.2.1 Goals

The goals of the 2017–18 Summative ELPAC AST were to do the following:

1. Standardize the administration of the ELPAC at all domains (i.e., Listening, Speaking, Reading, and Writing)
2. Train test examiners to score the Speaking items accurately and reliably
3. Train LEA trainers to train other qualified persons locally to administer and score the ELPAC
4. Inform LEAs about resources that are available to support optional local scoring of the ELPAC, including resources for scoring Writing responses locally

The training covered the test administration of all grade levels and grade spans as well as all domains. However, most of the training day was spent on the administration and scoring of the Speaking domain. Extensive training was provided because Speaking scores were given "in the moment" by test examiners, so the standardization of the scoring is critical. Refer to subsection *7.4 Constructed-Response Scoring for Speaking* for details about this aspect of the training.

5.2.2.2 Locations

The Summative ELPAC AST trainings were held at 24 locations throughout California from October 2017 through December 2017. All participants completing the Summative ELPAC AST were emailed certificates of completion. A total of 2,967 educators attended, representing a total of 1,382 LEAs (refer to table 5.1).

Table 5.1 2017 AST Training

2017 Date	Location	Attended
October 24	Sacramento	142
October 25	Regional Training	24
October 26	Redding	91
October 31	San Diego	151
November 1	Montebello	165
November 2	Burbank	164
November 3	Santa Barbara	105
November 6	Monterey	100
November 7	San Jose	133
November 8	Redwood City	89
November 9	Santa Rosa	114
November 14	Bakersfield	115
November 15	Fresno	153
November 16	Merced	77
November 17	Stockton	108
November 28	Burbank	198
November 29	Costa Mesa	112
November 30	Torrance	101
December 1	Pomona	113
December 5	Anaheim	120
December 6	Oceanside	134
December 7	Palm Springs	77
December 8	Riverside	104
December 12	Concord	130
December 13	Sacramento	147
Total:	**-**	**2,967**

An additional 320 LEAs were trained at COE-sponsored regional trainings. There were 20 regional trainings held by 14 COEs throughout the state. SCOE sold training materials on a cost-recovery basis to these county offices for their regional trainings to standardize all trainings.

One hundred and ninety-two LEAs had no participation data available, indicating they did not attend one of the scheduled training sessions.

5.2.2.3 Availability of Materials

The online Moodle Training Site was developed as a restricted site that could be accessed only by LEA trainers and test examiners. The site contained all resources needed to

conduct an LEA test examiner training session, such as downloadable training manuals, training presentations, training videos, scoring rubrics, as well as training and calibration quizzes for Speaking scoring. LEA trainers downloaded materials to prepare for their training sessions and shared access to the site with the test examiners within the LEA. Test examiners used the site to review training materials and to calibrate in preparation for Speaking scoring.

5.2.3. Scoring Rubrics

Scoring rubrics provide guidance to the raters who evaluate student responses. The *Speaking Rubrics for the ELPAC* (CDE, 2017b) and the *Writing Rubrics for the ELPAC* (CDE, 2018b) are essential components in the design of the ELPAC Speaking and Writing items.

5.2.3.1 Creation

Draft rubrics for scoring responses to the ELPAC Speaking and Writing items were designed in tandem with the design of task types for the ELPAC. The draft rubrics were designed to be used to score responses to several task types. As part of the 2014–15 pilot of ELPAC task types, the draft rubrics were used to evaluate student responses. After being modified as a result of further study after their first use, the revised rubrics were used to support the 2015–16 ELPAC item writing effort in which the item pool for the stand-alone field tests was developed.

During the item writing effort, the rubrics were further refined. The most significant change was that the rubrics were revised to be specific to each task type. This change was made based on the judgment that the use of task-specific rubrics, rather than generic rubrics, would increase the ease of internalization and usability by raters and help support efficient and reliable scoring.

5.2.3.2 Range Finding and Approval

After the item pool for the Summative ELPAC field test was developed, the *Speaking Rubrics for the ELPAC* (CDE, 2017b) and the *Writing Rubrics for the ELPAC* (CDE, 2018b) were reviewed during meetings held in Sacramento in 2016. CDE, ETS, and SCOE staff practiced scoring student responses from the pilot to evaluate the usability of the rubrics. After revisions were applied, the rubrics were approved for use during Speaking range finding meetings held in October 2016 and Writing range finding meetings held in May 2017.

The purpose of the Speaking range finding meetings and Writing range finding meetings was to select sample responses that were used to train raters and to calibrate them prior to scoring. During the range finding meetings, educators reviewed the rubrics and refined them. CDE staff reviewed and approved the revisions to the rubrics and selected samples that aligned with the rubrics while the range finding meetings were in session.

The approved rubrics were used to score student responses to the Summative ELPAC field test and the 2017–18 Summative ELPAC operational assessment.

5.3. Testing Students with Disabilities

The ELPAC provided a number of accessibility resources to enable all students to participate in the 2017–18 Summative ELPAC administration. ETS produced large-print test books as well as braille test books in contracted and uncontracted braille. The CDE's Matrix Four provided LEAs with guidance on available accessibility resources and accommodations (CDE, 2018a). Because the 2017–18 Summative ELPAC was a paper-

pencil assessment, embedded resources typically available on a computer-based assessment, such as text-to-speech, closed captioning, American Sign Language videos, and embossed braille, were not available.

5.3.1. Alternate Assessments

Individualized education program (IEP) teams may have determined that a student was unable to participate in one or more domains of the ELPAC, even with accommodations, due to short- or long-term disabilities. In this instance, the student may have been tested with a locally determined alternative assessment per the student's IEP.

A version of the Summative ELPAC for students with the most significant cognitive disabilities is not currently available.

5.4. Test Security and Confidentiality

For the 2017–18 Summative ELPAC administration, every person who worked with the assessments, communicated test results, or received testing information was responsible for maintaining the security and confidentiality of the tests, including CDE staff, ETS staff, ETS subcontractors, LEA ELPAC coordinators, site ELPAC coordinators, and test examiners.

ETS' Code of Ethics requires that all test information, including tangible materials (e.g., test items and test books), confidential files (e.g., those containing personally identifiable student information), processes related to test administration (e.g., the packing and delivery of test materials), and activities are kept secure. ETS has systems in place that maintain tight security for test items, test books, and test results, as well as for student data. To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI), which is described in the next subsection.

All tests within the ELPAC system, as well as the confidentiality of student information, should be protected to ensure the validity, reliability, and fairness of the results. As stated in *Standard 7.9* of the the *Standards for Educational and Psychological Testing*, "The documentation should explain the steps necessary to protect test materials and to prevent inappropriate exchange of information during the test administration session" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

5.4.1. ETS' Office of Testing Integrity

The OTI is a division of ETS that provides quality assurance services for all testing programs managed by ETS; this division resides in the ETS legal department. The Office of Professional Standards Compliance at ETS publishes and maintains *ETS Standards for Quality and Fairness* (ETS, 2014), which supports the OTI's goals and activities. The *ETS Standards for Quality and Fairness* provides guidelines to help ETS staff design, develop, and deliver technically sound, fair, and beneficial products and services and to help the public and auditors evaluate those products and services.

The OTI's mission is to

- minimize any testing security violations that can impact the fairness of testing,
- minimize and investigate any security breach that threatens the validity of the interpretation of test scores, and
- report on security activities.

The OTI helps prevent misconduct on the part of students and administrators, detects potential misconduct through empirically established indicators, and resolves situations involving misconduct in a fair and balanced way that reflects the laws and professional standards governing the integrity of testing. In its pursuit of enforcing secure practices, the OTI strives to safeguard the various processes involved in a test development and administration cycle.

5.4.2. Test Delivery

Because the 2017–18 Summative ELPAC was a paper-pencil assessment, there were logistics involved to ensure the timely delivery of test materials to LEAs across the state. To manage the materials ordering process, ETS used TOMS, a secure website that permitted ELPAC users to perform a number of tasks for the ELPAC program. Through TOMS, users could perform the following activities:

- Confirm or update an LEA shipping address, add a score report shipment address, and indicate whether an LEA can receive pallet shipments
- Order test materials, including braille and large-print forms, in either Round 1 or Round 2, and order additional test materials as needed, in the Supplemental window
- Add site ELPAC coordinators and test examiners
- Order Pre-ID labels
- Administer the Listening domain and the *Speaking—Summarize Academic Presentations* item for grades three through twelve

The ETS warehouse team prepared shipments based on orders submitted by each LEA. Materials were tracked using closed-loop tracking and United Parcel Service tracking methods to ensure timely delivery of ELPAC test materials. Shipping notices were included in each delivery. These notices provided LEAs with an inventory of the number of Test Books, Answer Books, and other materials included in the shipment. Additionally, LEAs were provided with return materials that included Group Identification Sheets, which were precoded, scannable forms facilitating identification of materials when they were received at ETS; and shipping labels that allowed tracking of materials that were returned to ETS for scoring.

5.4.3. Security of Electronic Files Using a Firewall

A firewall is software that prevents unauthorized entry to files, email, and other organization-specific information. All ETS data exchanges and internal email remain within the ETS firewall at all ETS locations, ranging from Princeton, New Jersey, to San Antonio, Texas, to Concord and Sacramento, California.

All electronic applications that are included in TOMS remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student information processed by TOMS, the firewall plays a significant role in maintaining assurance of confidentiality among the users of this information.

5.4.4. Transfer of Scores via Secure Data Exchange

Due to the confidential nature of test results, ETS currently uses secure file transfer protocol (SFTP) and encryption for all data file transfers; test data is never sent via email. SFTP is a method for the reliable and exclusive routing of files. Files reside on a password-protected server that only authorized users can access. ETS shares an SFTP server with the CDE.

On that site, ETS posts Microsoft Word and Excel files, Adobe Acrobat PDFs, or other document files for the CDE to review; the CDE returns reviewed materials in the same manner. Files are deleted upon retrieval.

The SFTP server is used as a conduit for the transfer of files; secure test data is stored only temporarily on the shared SFTP server. Industry-standard secure protocols are used to transfer test content and student data from the ETS internal data center to any external systems. For the 2017–18 Summative ELPAC, ETS entered information about the deliverable in a web form on a SharePoint website when a file was posted. A CDE staff member checked this log throughout the day for updates on the status of deliverables and downloads and deleted the file from the SFTP server when its status shows it has been posted.

5.4.5. Data Management

ETS currently maintains a secure database to house all student demographic data and assessment results. Information associated with each student has a database relationship to the LEA, school, and grade codes as the data is collected during operational testing. Only individuals with the appropriate credentials can access the data. ETS builds all interfaces with the most stringent security considerations, including interfaces with data encryption for databases that store test items and student data. ETS applies best and up-to-date security practices, including system-to-system authentication and authorization, in all solution designs.

All stored test content and student data is encrypted. ETS complies with the Family Educational Rights and Privacy Act (20 *United States Code [USC]* § 1232g; 34 *Code of Federal Regulations* Part 99) and the Children's Online Privacy Protection Act (15 *USC* §§ 6501–6506, P.L. No. 105–277, 112 Stat. 2681–1728).

In TOMS, staff at LEAs and test sites were given different levels of access appropriate to the role assigned to them.

5.4.6. Statistical Analysis on Secure Servers

The 2017–18 Summative ELPAC results were scanned or entered by human raters. After scoring constructed-response items, the Information Technology team at ETS loaded data files from the SFTP site and then loaded them into a database. The ETS Data Quality Services staff extracted the data from the database and performed quality-control procedures before passing files to the ETS statistical analysis group. The statistical analysis group kept the files on secure servers. All staff members involved with the data adhered to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access to data.

5.4.7. Student Confidentiality

To meet the requirements of the Every Student Succeeds Act as well as state requirements, LEAs must collect demographic data about students' ethnicity, disabilities, parent/guardian education, and so forth. ETS took every precaution to prevent any of this information from becoming public or being used for anything other than testing purposes. These procedures were applied to all documents in which student demographic data appeared, including reports and the Pre-ID files and response booklets used in paper-pencil testing.

5.4.8. Student Test Results

5.4.8.1 Types of Results

Printed Student Score Reports (SSRs) were produced for each student administered the Summative ELPAC. Score reports were produced within six to eight weeks of the receipt and scoring of student responses. LEAs received two copies of the SSR: one for the student and a copy for the student's cumulative file.

Additionally, ETS produced aggregate data files containing ELPAC test result data for schools.

5.4.8.2 Security of Results Files

ETS took measures to protect files and reports that showed students' scores and ELPAC levels. ETS is committed to safeguarding all secure information in its possession from unauthorized access, disclosure, modification, or destruction. ETS has strict information security policies in place to protect the confidentiality of both student and client data. ETS staff access to production databases was limited to personnel with a business need to access the data. User IDs for production systems were person-specific or for systems use only.

ETS has implemented network controls for routers, gateways, switches, firewalls, network tier management, and network connectivity. Routers, gateways, and switches represent points of access between networks. However, these do not contain mass storage or represent points of vulnerability, particularly for unauthorized access or denial of service.

ETS has many facilities, policies, and procedures to protect computer files. Software and procedures such as firewalls, intrusion detection, and virus control are in place to provide for physical security, data security, and disaster recovery. ETS is certified in the BS 25999-2 standard for business continuity and conducts disaster recovery exercises annually. ETS routinely backs up all data to either disks through deduplication or to tapes, all of which are stored off site.

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that can only be unlocked by the badges of personnel who have functional responsibilities within its secure perimeter. Authorized personnel accompany visitors to the ETS Computer Processing Center at all times. Extensive smoke detection and alarm systems, as well as a preaction fire-control system, are installed in the Center.

5.4.8.3 Security of Individual Results

ETS protects individual students' results on both electronic files and paper reports during the following events:

- Scoring
- Transfer of scores by means of secure data exchange
- Reporting
- Analysis and reporting of erasure marks
- Posting of aggregate data
- Storage

In addition to protecting the confidentiality of testing materials, ETS' Code of Ethics further prohibits ETS employees from financial misuse, conflicts of interest, and unauthorized appropriation of ETS property and resources. Specific rules are also given to ETS

employees and their immediate families who may be administered a test developed by ETS (e.g., an ELPAC). The ETS OTI verifies that these standards are followed throughout ETS. This verification is conducted, in part, by periodic on-site security audits of departments, with follow-up reports containing recommendations for improvement.

5.4.9. Security and Test Administration Incident Reporting Process

The LEA ELPAC coordinator was responsible for reporting all testing incidents and security breaches immediately.

If an irregularity or security breach occurred at the school, the test examiner was required to report the incident to the LEA ELPAC coordinator. Testing irregularities relate to incidents that occur during the administration of the ELPAC that were likely to impact the reliability and validity of the test.

Testing irregularities included *but are not limited to*:

- Cheating by students
- Failing to follow test administration directions
- Rushing students through the test or parts of the test
- Coaching students, for example:
  - Discussing questions with students before, during, or after testing
  - Giving or providing any clues to the answers
- Administering the wrong grade level or grade span test to a student or using mismatched test materials
- Writing on the Answer Book by a test examiner that would cause the Answer Book to be unscorable and therefore need transcription to a new Answer Book
- Leaving instructional materials on walls in the testing room that may assist students in answering test questions
- Allowing students to have additional materials or tools (e.g., books, tables) that are **not** specified in an IEP, Section 504 plan, or approved by the CDE as an allowed testing accommodation

Security breaches included, *but are not limited to:*

- Site ELPAC coordinators, test examiners, proctors, or students using electronic devices such as cell phones during testing
- Posting pictures of test materials on social media sites
- Missing test materials
- Copying or taking a photo of any part of the test materials
- Permitting eligible students access to test materials outside of the testing periods
- Developing scoring keys or reviewing any student responses
- Failing to maintain security of all test materials
- Sharing test items or other secure materials with anyone who has not signed the Test Security Affidavit

- Discussing test content or using test materials outside of training and administration
- Allowing students to take the test out of the designated testing area
- Allowing test examiners to take the test home
- Allowing untrained personnel to administer the test

If an incident occurred, the LEA ELPAC coordinator was instructed to notify ETS and the CDE within 24 hours of the incident. Additionally, the coordinator was required to complete the ELPAC *Testing Irregularities and Security Breach Report* form. The CDE and ETS collaborated on defining next steps and providing the LEA ELPAC coordinator with instructions on how to mitigate the incident.

References

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

California Department of Education. (2017a). *ELPAC test administration manual.* Sacramento, CA: California Department of Education. Retrieved from https://www.elpac.org/s/pdf/ELPAC_test-admin-manual-summative.2018.pdf

California Department of Education. (2018a). *Matrix Four: Universal tools, designated supports, and accommodations for the English Language Proficiency Assessments for California*. Retrieved from https://www.cde.ca.gov/ta/tg/ep/documents/elpacmatrix4.docx

California Department of Education. (2017b). *Speaking rubrics for the English Language Proficiency Assessments for California* (Unpublished manuscript). Sacramento, CA: California Department of Education.

California Department of Education. (2017c). *Test Operations Management System guide for the ELPAC*. Sacramento, CA: California Department of Education. Retrieved from https://www.elpac.org/s/pdf/ELPAC.TOMS-users-guide.2017-18.pdf

California Department of Education. (2018b). *Writing rubrics for the English Language Proficiency Assessments for California.* Sacramento, CA: California Department of Education. Retrieved from https://www.elpac.org/s/pdf/ELPAC_Writing-Rubrics.pdf

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/s/about/pdf/standards.pdf

Chapter 6: Standard Setting

This chapter summarizes the standard setting process through which Summative English Language Proficiency Assessments for California (ELPAC) performance levels were established. Included are background of the development of ELPAC, an overview of the standard setting methodology, a summary of the standard setting procedures, the description of the performance level descriptors (PLDs), and the results. The detailed standard setting information for the Summative ELPAC is described in the *Standard-Setting Technical Report for the Summative ELPAC* (Educational Testing Service [ETS], 2018).

6.1. Background

Implementation of the *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards) and the administration of the new Summative ELPAC required a standard setting process to evaluate students' English language proficiency (ELP) against the new expectations.

Figure 6.1 presents the score reporting hierarchy for the Summative ELPAC, approved in September 2017 by the California State Board of Education (SBE). As depicted in this figure, four performance levels must be reported for three composite scores: scale scores and performance levels for the overall, oral language, and written language scores. The oral language scale score branches off into the Listening and Speaking domains, which each have three performance levels. The written language scale score branches off into the Reading and Writing domains, which each have three performance levels.

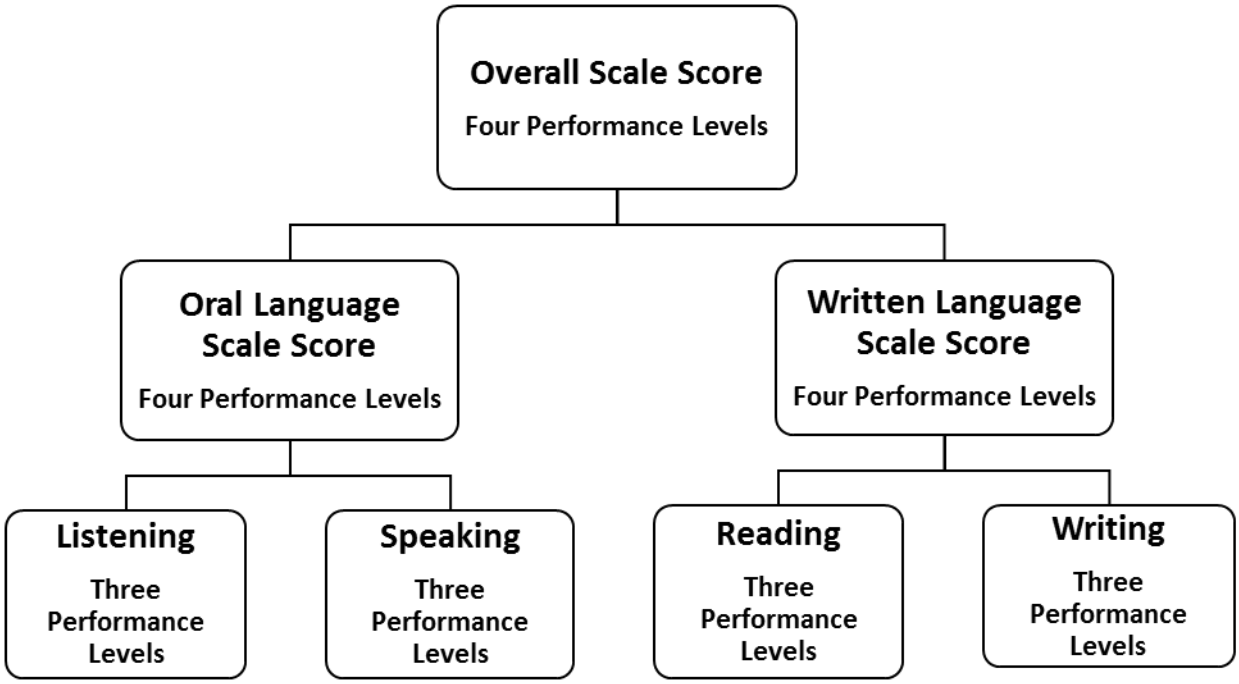

Figure 6.1 Summative ELPAC Score Reporting Hierarchy

To develop threshold score recommendations aligned with the score reporting hierarchy, ETS conducted standard setting workshops in Sacramento, California, for the seven Summative ELPAC grade levels and grade spans on October 17, 2017, through

October 20, 2017 (kindergarten [K], grade one, and grade two), and October 23, 2017, through October 26, 2017 (grade spans three through five, six through eight, nine and ten, and eleven and twelve). Standard setting for K through grade two was conducted in week one and for grades three through twelve in week two. All four domains and the overall score were considered in the standard setting process.

6.2. Performance Level Descriptors

The Summative ELPAC general (policy) PLDs describe short policy descriptors that convey the expectation at each performance level, across all grades tested (CDE, 2016). They were provided to the panelists for prereading prior to the standard setting workshop.

After the General PLDs were available, a team of educators familiar with both students taking the Summative ELPAC and the 2012 ELD Standards reviewed the general PLDs for the ELPAC target population. They developed more detailed grade- and content-specific PLDs for the range of expectations at each performance level (Range PLDs). Panelists referenced the SBE-approved general PLDs and the Range PLDs as part of the standard setting process.

6.3. Standard Setting Methodologies

Standard setting refers to a class of methodologies by which one or more performance threshold scores are used to determine performance levels. The purpose of the standard setting process for the Summative ELPAC was to collect recommendations from California educators for the placement of the threshold scores for review by the CDE, with final determination and approval by the SBE.

ETS conducted standard setting workshops in fall 2017, following the field test administration of the Summative ELPAC. The overall approach used for setting standards for the Summative ELPAC aligned with the 2012 ELD Standards, which reflect the interdependence of the language domains.

By design, the Summative ELPAC and the standard setting methodologies explicitly support a treatment of skills such as Speaking and Listening in combination, rather than as isolated skills. Educators working on standard setting panels considered the assessment by domain, articulated skills that are expected in Listening, Speaking, Reading, and Writing, and made final threshold score recommendations by considering the interdependence of these skills.

The Bookmark method (Lewis, et al., 1998; Mitzel, et al., 2001) was applied to the Reading and Listening domains; a Performance Profile approach was applied to the Writing and Speaking domains (Baron & Papageorgiou, 2014; Tannenbaum & Cho, 2014; Tannenbaum & Baron, 2010; Wan, Bay, & Morgan, 2017). In the final round, panelists were instructed to think holistically across the four domains and consider consequence data when they made the overall threshold score recommendations.

6.3.1. Bookmark Method (Reading and Listening Domains)

The Summative ELPAC standard setting process employed the Bookmark method for the seven grade levels and grade spans (K, grades one and two, and grade spans three through five, six through eight, nine and ten, and eleven and twelve) for the Reading and

Listening domains, which consisted of dichotomously scored multiple-choice items[2]. This portion of the workshop resulted in recommendations for threshold scores for these two domains.

The Bookmark method has its basis in item response theory (IRT) analysis. IRT is used to estimate item difficulties. These estimates are used to order the items from easiest to hardest in a booklet known as an ordered item booklet (OIB) and to place item difficulty estimates on the score scale. One benefit of this approach is that, once panelists make judgments in the OIB, the difficulty values associated with each item have a built-in relationship to scale scores through theta (the ability parameter in IRT), which allows results to be provided to score users and policy makers on the familiar metric of scale score. Panelists completed two rounds of Bookmark judgments for Reading and Listening for their assigned grade level or grade span. Then, the panelists began work on the Speaking and Writing domains for the same assigned grade level or grade span.

6.3.2. Performance Profile Method (Speaking and Writing Domains)

The Summative ELPAC standard setting process employed the Performance Profile method for the Speaking and Writing domains, which consisted of constructed-response items. This portion of the workshop resulted in recommendations for threshold scores for these two domains.

The Performance Profile method is a holistic method that requires panelists to make decisions or judgments based on an examinee's score profiles, or overall performance, rather than on each separate test item or task. This method has been used in standard setting studies for English learner (EL) assessments and other types of K–12 statewide assessments throughout the United States (e.g., Baron & Papageorgiou, 2014; ETS, 2014).

In this approach, panelists reviewed actual samples of student responses across multiple tasks, such as Speaking video samples of student performance on the Speaking tasks, and multiple Writing responses. Item scores for a student's set of responses to the items form a profile; panelists considered the performance at each total score represented by the profiles of responses across tasks. Writing profiles were sampled from field test responses, and speaking profiles were sampled from scorer-training videos developed by the Sacramento County Office of Education in June 2017. Profiles were selected to represent the full range of scores and the most frequently occurring score patterns.

In each of two rounds of judgments, all panelists independently selected total scores associated with score profiles and marked the score representing the expected knowledge and skills at the threshold of each performance level, using the definitions of borderline students. The instruction to the panelists was to base decisions about which total score aligns best with the definition of the borderline student on the full set of evidence provided across all test items in Speaking. (The same process was followed for Writing.)

Panelists recorded their Round 1 recommended Speaking or Writing total score for each threshold score. After Round 1, each panelist's individual cut-score recommendations were shared with the panel and discussed; panel judgments were summarized and discussed prior to the next round of judgments.

[2] Grade two included two items that were multipoint items. These items appeared twice in the ordered item booklet (OIB), according to the RP67 theta value associated with each score point.

6.4. Standard Setting Procedures

This subsection describes what occurred prior to and during the standard setting workshop.

6.4.1. Panelists

Prior to the standard setting, panelists were recruited to include a diverse, representative group of California educators with both experience in the education of students who will take the ELPAC and familiarity with the 2012 ELD Standards. An additional goal was to recruit subject-area teachers working with these students in grades six and above, because these teachers provide a perspective on content-specific learning goals for the students taking the ELPAC. Educators were selected using the following criteria:

- Educators who are working with ELs in the grade level(s) assigned to the panel
- English-language specialists
- Educators teaching any or all of the subject areas of mathematics, science, and social studies

The final decision on the panelists selected for the workshops was made by the CDE.

For the Summative ELPAC, there were six panels of educators. Three panels—kindergarten and grades one and two—met during the first week of the workshop. Three panels—grade spans three through five, six through eight, and nine through twelve—met in the second week. There were 71 panelists; the number in each panel ranged from 8 to 11.

Because standard setting is based on expert judgment—informed by performance data—it was important that panelists collectively reflected the diversity of the educators working with students who take the assessment. Special efforts were made to assemble panels that were representative of the geographic and socioeconomic diversity of California in general and the ELPAC educator population in particular. Panels included a sample across genders, ethnic and racial backgrounds, and geographical regions in California. A majority of the educators indicated they had more than five years' experience working with ELs. Most panels included educators with experience teaching mathematics, science, social studies, and English.

6.4.2. Materials

All panelists, regardless of the standard setting methodology, were provided the following materials prior to the workshop:

- A letter describing the purpose and procedures of the standard setting workshop
- A preworkshop assignment specific to their panel assignment
- A notetaking form for the assignment
- A link to the SBE-approved general PLDs
- The domain and grade- and grade-span-specific PLDs for the tests the panelists would be reviewing

At the standard setting workshop, panelists received training materials, a set of operational materials, and evaluation forms. The definitions of borderline students were developed by the panelists themselves during the workshop.

The operational materials panelists received at the workshop included the following:

- The OIB (for the Bookmark method)
- An item map
- Judgment recording forms
- Performance samples for Speaking and Writing
- Rubrics used operationally to score student responses

At the end of the training for each method, and at the end of the workshop, panelists completed evaluation forms. Evaluations included questions about training, understanding the tasks, the influence of different aspects of the standard setting process, and panelists' beliefs about the final recommended threshold scores.

6.4.2.1 Descriptions of Materials

6.4.2.1.1. Materials for the Bookmark Method

The OIB is a booklet of all items included in the standard setting judgments, ordered by difficulty on the basis of student performance. For each item, the page of the OIB shows the item, along with any short passage or graphic, the possible responses, and the correct answer. For the items that are associated with a passage, a separate passage booklet was included with the OIB for panelists to reference for items associated with a passage.

The item map is a summary document displaying relevant information regarding each item. It shows the ordered item number, the original item number in the test, the correct answer, a difficulty value, and the passage title and score-level scale. The item map provided was ordered by difficulty in the same manner as the OIB.

6.4.2.1.2. Materials for the Performance Profile Method

Performance Profile samples are complete student responses to the Speaking and Writing tests. For Speaking, video files of students responding to all tasks were displayed; each student score was known to the panelists, allowing them to visualize a sample of students across the range of performance. All student videos played for each panel showed students taking the same items.

For Writing, copies of students' Writing responses to the full set of Writing tasks were provided in a booklet of Writing samples. The Writing sample book included the prompt and written response for a range of Writing domain scores. More than one score sample was displayed when available for both Speaking and Writing.

Panelists also used the scoring rubrics for Speaking and Writing in their discussions and in their individual judgments

6.4.3. Process

The workshop process included a general session, where all panelists were provided an overview of the purpose of the meeting, their role and the roles of facilitators, and an explanation of the two approaches used in the standard setting for the ELPAC. Educators were then guided to grade span-specific panel rooms, where they completed the training and judgment process (Baron & Papageorgiou, 2016; Morgan, 2004).

6.4.3.1 Training

Training was provided on the following topics:

- Test familiarization
- Development of Borderline Student Definitions
- Standard setting judgment process for both bookmark and performance profile

  - Training and practice prior to the first round of judgments
- Review of ordered items and practice in method of bookmark placement
- Review of speaking videos for performance profile judgments
- Feedback and discussion and Round 2 judgments for each domain
- Round 3 integrated holistic judgments on the overall score

6.4.3.2 Judgments and Feedback

The Reading and Writing section scores were combined into the written composite score, and the Speaking and Listening section scores were combined into the oral composite score. The feedback to the panelists after Round 2 judgments were complete included each of the four domain score recommendations as well as the recommended threshold scores for the two composites. Panelists received training on how the domain scores were combined and how to consider the data provided for the domains and composites in the Round 3 integrated judgments.

Panelists made recommendations for three threshold scores on the Summative ELPAC overall score and were instructed to consider all of the information provided and then make a recommendation for the overall score performance level expectations.

6.5. Standard Setting Results

Results from the ELPAC standard setting after Round 3 included a recommended threshold score for each composite (oral and written) and the overall composite for each test (kindergarten, grades one and two, and grade spans three through five, six through eight, nine and ten, and eleven and twelve). The *Standard-Setting Technical Report for the Summative ELPAC* (ETS, 2018) presents details about the following results from the standard setting workshops:

- The median threshold score recommendations for each domain at the end of each round
- Standard errors of judgment, scale scores, and conditional standard errors of measurement in the Bookmark metric for Reading and Listening
- Standard errors of judgment, scale scores, and conditional standard errors of measurement in the Performance Profile metric for Speaking and Writing

Table 6.1 through table 6.7 show the projected percentage of students statewide who would be placed at this performance level on the basis of the results of the 2016–17 field test administration of the Summative ELPAC. The threshold scale score is the minimum standard setting scale score needed to achieve a performance level.

Scales provided in these tables were presented and used in the standard setting process and are more user-friendly than scores in the theta metric. However, it should be noted that the scores presented are not the ELPAC-reported scale scores. The scale was created, based on the 2016–17 field test data, for standard setting prior to the approval of the official scale for the Summative ELPAC and was used as a tool for the standard setting process.

Table 6.1 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Kindergarten

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	10.6
Level 2	338	20.4
Level 3	380	35.8
Level 4	428	33.2

Table 6.2 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade One

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	9.1
Level 2	381	14.0
Level 3	411	21.3
Level 4	441	55.5

Table 6.3 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade Two

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	5.1
Level 2	389	10.2
Level 3	424	33.6
Level 4	475	51.1

Table 6.4 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade Span Three Through Five

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	8.1
Level 2	441	21.4
Level 3	490	52.8
Level 4	569	17.7

Table 6.5 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade Span Six Through Eight

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	6.5
Level 2	451	24.6
Level 3	516	40.7
Level 4	577	28.3

Table 6.6 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade Span Nine and Ten

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	16.7
Level 2	484	25.6
Level 3	544	31.9
Level 4	607	25.9

Table 6.7 Projected Distribution of 2017 Students Based on Round 3 Recommendations: Grade Span Eleven and Twelve

Performance Level	Threshold Scale Score	Percentage
Level 1	NA	13.4
Level 2	486	24.3
Level 3	547	36.4
Level 4	618	25.8

Results presented in the *Standard-Setting Technical Report for the Summative ELPAC* are based on the standard setting workshop and panel-recommended threshold scores at the end of the workshop. Following the standard setting workshop, the SBE reviewed both the panel recommendations and the State Superintendent of Public Instruction's recommendations for threshold scores.

References

Baron, P. A., & Papageorgiou, S. (2014). *Mapping the TOEFL® Primary™ test onto the Common European Framework of Reference* (Research Memorandum RM-14-05). Princeton, NJ: Educational Testing Service.

Baron, P. and Papageorgiou, S. (2016). *Setting language proficiency score requirements for English-as-a-second-language placement decisions in secondary education*. (Research Report 16-17). Princeton, NJ: Educational Testing Service. Retrieved from https://files.eric.ed.gov/fulltext/EJ1124774.pdf

California Department of Education. (2014). *2012 California English language development standards: Kindergarten through grade 12*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

California Department of Education. (2016). *Summative ELPAC general PLDs.* Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/ta/tg/ep/elpacgpld.asp

Educational Testing Service. (2014). *Standard-setting technical report; Proficiency Assessments for Wyoming Students [PAWS] and Student Assessment of Writing Skills [SAWS])* (Unpublished report). Princeton, NJ: Educational Testing Service.

Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998). *The Bookmark standard setting procedure: Methodology and recent implementations.* Paper presented at the 1998 annual meeting of the National Council on Measurement, San Diego, CA.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.). *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 249–81). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Morgan, D.L. (2004, June). *The performance profile method (PPM): A unique standard setting method as applied to a unique population.* Presented at the annual meeting of the Council of Chief State School Officers in Boston, MA.

Tannenbaum, R. J., & Baron, P. A. (2010). *Mapping TOEIC® test scores to the STANAG 6001 language proficiency levels* (Research Memorandum RM-10-11). Princeton, NJ: Educational Testing Service.

Tannenbaum, R. J., & Cho, Y. (2014). Critical factors to consider in evaluating standard-setting studies to map language test scores to frameworks of language proficiency. *Language Assessment Quarterly*, 11, 233–249.

Wan, L., Bay, L., & Morgan, D. (2017) *Validity evidence for the Performance Profile standard setting method.* Presentation at the National Council for Measurement in Education Annual Conference, San Antonio, TX.

Chapter 7: Scoring and Reporting

7.1. Procedures for Maintaining and Retrieving Individual Scores

The local educational agency (LEA) English Language Proficiency Assessments for California (ELPAC) coordinator was responsible for returning all materials to Educational Testing Service (ETS) for scoring. When materials were received at the ETS warehouse, several quality checks were implemented. These include verifying there was no damage to the Answer Books prior to scanning as well as capturing issues such as double marks and inconsistencies between the pre-identification label and the marked information.

Once received, all Answer Books were scanned and writing items were routed to trained raters at ETS for scoring. Once student responses were scored, a Student Score Report (SSR) was produced.

7.2. Multiple-Choice Scoring

After the certification of student records for scoring, ETS transferred the records to the scoring management system. These records contained all relevant response data and identifying information for matching against the correct scoring keys. The ETS scoring engine then processed the records and produced the multiple-choice (MC) raw scores before permanently storing the results in the students' records.

7.3. Constructed-Response Scoring for Writing

Prior to operational use, for all ELPAC Writing items, a range of professionals that included California educators carefully developed and vetted the rubrics, benchmark sample responses, and rater training materials over the course of field testing and additional reviews.

7.3.1. Scorer Training

It is critical for the success of the ELPAC constructed-response (CR) scoring to have well-defined scorer recruitment, training, and certification processes with staff in place to control scoring quality.

7.3.1.1 Procedures

The procedures ETS used in training ELPAC scorers included the following:

- **Rigorous Training for the Scoring Leaders.** ETS developed training materials and helped select benchmark and training samples during range finding for the purpose of training scoring leaders and scorers. ETS hired scoring leaders with experience and familiarity in scoring similar programs at ETS. Scoring leaders were given materials to study independently.
- **Extensive Training of Scorers.** Scorers were trained to properly apply the appropriate rubric for scoring each task type, following generic sample responses that exemplified the quality required for each score point. This ensured that every prompt was scored using the same general criteria. The ETS Online Network for Evaluation (ONE) scoring system supported scorer training with a full-service menu of options, including orientation materials, program-specific information, and training on how to use the platform, as well as interactive training that included practice scoring for both potential and qualified scorers.

There were two types of training sets offered within ONE:

- **Feedback Sets** provided users with feedback after each response. Users could also access the overall results at the end of the set.
- **Practice Sets** mimicked the actual scoring, and users do not have access to any score results until they have completed the set.

ETS provided role-based training modules for using ONE. For example, ETS expected scoring leaders to review and study more modules (on topics such as monitoring) than scorers. Scoring leaders and scorers were required to review PDFs as well as training videos that covered the critical functions required for their individual scoring roles.

7.3.1.2 Certification

Certification occurred after training and was intended to determine how well scorers could adopt and apply the scoring standards. Scorers' ratings on certification responses were compared to predetermined correct scores to ascertain whether scorers successfully applied the scoring standards reflected in the rubrics (scoring guides).

As part of the initial qualification for scoring ELPAC prompts, every scorer had to successfully complete training and pass a certification test consisting of a set of prescored responses. If a scorer was unsuccessful on the first certification attempt, that scorer was retrained prior to making a second attempt. If a scorer was unsuccessful at the second attempt, that scorer was not added to the potential ELPAC scorer pool. This process increased the likelihood of securing a highly proficient scorer pool.

7.3.1.3 Regular Calibration

Calibration is a short test of reader accuracy that occurs regularly at the beginning of a scoring session to determine whether scorers are ready to begin scoring the assessment. Calibration is a proven method to mitigate scoring drift and promote the quality of scoring over time.

Before calibration, scorers were directed to review relevant training materials (rubrics and benchmarks). During calibration, scorers assigned scores to a prescored set of responses to determine their ability to accurately apply scores for a particular task type.

As with the certification process described previously, scorers had two opportunities for correct calibration. If a scorer was unsuccessful on his or her first calibration attempt, that scorer conferred with his or her scoring leader for advice and guidance. The scoring leader had access to the scorer's performance results and could mentor the scorer on specific areas of scoring inaccuracy. The scoring leader advised the scorer to refer to training content and read over practice responses prior to a second attempt at calibration. Scorers who were unsuccessful after two attempts at calibration were not allowed to score that particular prompt on that day.

7.3.1.4 Temporary Samples

To recruit and train enough scoring staff to complete scoring according to schedule, training and certification were conducted before samples were selected at the Writing range finding meetings. As a result, Writing samples from other testing programs were used on a temporary basis to train and certify staff to score the Summative ELPAC Writing field test only. The Writing samples were drawn from task types similar to the ELPAC task types. For example, Writing samples were drawn from the California English Language Development

Test *Writing—Short Compositions* prompts, which are similar to the ELPAC prompts in *Writing—Write About an Experience*.

ETS assessment specialists selected the temporary samples and scored them using the ELPAC Writing rubrics. The samples were used to train and certify raters to use the ELPAC Writing rubrics.

7.3.2. Scoring Practices

ETS adhered to the following scoring practices and procedures:

- New scorers had to demonstrate their accuracy by passing a certification test before they were scheduled to score the ELPAC. Scorers had to then pass a shorter, more focused calibration test before each scheduled scoring session. Certification and calibration were described in the previous subsection.
- Scorers underwent training in appropriately applying the rubric for each specific task type, following the generic sample responses that exemplified the quality required for each score point. This ensured that every prompt was scored using the same general criteria.
- If scorers disagreed on a score, the decision moved up to a scoring leader and, if needed, the content scoring leader. Scoring leaders provided adjudications of discrepant scores (i.e., scores more than one score point apart). A scorer discovered to be scoring inaccurately was additionally monitored and might have been required to have additional training. In some cases, if scoring inaccuracies were a problem, the scorer was dismissed.
- ETS trained all levels of scoring leadership, not only on the prompts, rubrics, and related scoring material, but also on how best to monitor the quality of the scoring.
- Scoring leaders read behind and monitored scorers. Scoring leaders also had the option of evaluating responses a scorer previously scored, with or without the knowledge of the score given (“informed” versus “blind” back-rating). Scoring leaders read behind up to 10 percent of a scorer’s responses.
- ETS used a double scoring percentage of 10 percent to monitor and verify interrater reliability.
- ONE provided operational data on scorers and teams who were reading at unusually slow or rapid rates, allowing scoring leadership to investigate and provide counseling and guidance, if warranted.
- During each scoring session, highly skilled content scoring leaders monitored scoring leaders and their virtual teams by reviewing interrater agreement rates as well as back-scoring agreement rates between scorers and scoring leaders. Content scoring leaders adjudicated any discrepant scores that arose and provided feedback to scorers and scoring leaders as needed.
- ETS assessment specialists and top scoring leadership analyzed interrater reliability statistics to verify that scorers were scoring consistently and at levels that meet professional psychometric standards.

7.3.3. Managing Scoring

ETS invited applicants to score the Summative ELPAC Writing domain largely from its existing rater pool of more than 30,000 experienced raters. Raters who accepted the invitation met the rater qualifications and demonstrated their scoring accuracy by passing a certification test before being selected to score for the ELPAC.

7.3.4. Monitoring Scoring

There are proven processes in place for monitoring ELPAC scoring. During the 2017–18 operational scoring, raters passed a regular calibration test that measured the rater's ability to accurately apply scores to responses for a particular prompt or task. Scoring leadership mentored the raters with feedback during shifts and completed back-rating of rater scores.

ETS had communication channels in place among raters, leadership, and ETS staff to share information related to operational scoring and personnel concerns. ETS staff monitored these communications and investigated all scoring accuracy and personnel concerns.

7.4. Constructed-Response Scoring for Speaking

7.4.1. Scorer Training for Speaking

Participants in the Summative ELPAC Administration and Scoring Training (AST), described in subsection *5.2.2 Scoring Training of Trainers Workshops*, received training on the administration and scoring of the Speaking domain. The training agenda primarily focused on Speaking task types.

7.4.1.1 Methods of Approaching Item Types

Workshop trainers presented each of the six Speaking task types using the following strategies:

- Video of student being administered the task type
- 2012 *California English Language Development Standards, Kindergarten Through Grade 12* associated with the task type
- Test administration procedures
- Rubric overview
- Scoring and prompting guidelines
- Anchors
- Practice scoring

7.4.1.2 Agenda

What follows is the agenda used during the scoring training:

- Section 1—Introduction
- Section 2—Test Administration
  - Test Administration
  - Grades 3–12 Listening video
  - Grades 3–12 Reading video
  - Grades 3–12 Writing video
  - Moodle Training Site

- Section 3—Speaking Overview
  - Speaking Overview
  - Full Speaking video
- Section 3—Talk About a Scene (K–12)
- Section 4—Speech Functions (2–12)
- Section 5—Support an Opinion (K–2)
- Section 5—Support an Opinion (3–12)
- Section 6—Retell a Narrative (K–5)
- Section 7—Present and Discuss Information (6–12)
- Section 8—Summarize an Academic Presentation (K–5)
- Section 8—Summarize an Academic Presentation (6–12)
- Section 9—Full Speaking video (K–5)
- Section 9—Full Speaking video (6–12)
- Section 10—K–1 Administration of Reading, Writing, and Listening

7.4.1.3 Training Materials

To establish consistency in statewide local training, training materials were developed and provided to all LEAs. Each person attending training received a printed training binder with access to a PDF provided on the Moodle training website. Participants were also provided with administration training videos and training presentations, with scripts posted on the Moodle training website for LEA trainers to use for their local training of test examiners. The training materials were primarily focused on scoring the Speaking task types. Training materials are described in the next subsections.

7.4.1.3.1. Training Binder

A Summative ELPAC AST binder was provided to participants in the training. Each binder contained the following sections:

- **Section 1—Introduction**
  - Overview of the program
  - Contact information
  - Program resources
- **Section 2—Test Administration**
  - Overview of use of TOMS for streaming Listening domain and Speaking
  - Group test administration
  - Accommodation matrix
  - Other logistics
- **Section 3—Talk About a Scene**
  - Prompting and scoring guidelines
  - Rubrics
  - Each scene
  - Anchor charts

- **Section 4—Speech Functions**
  - Prompting and scoring guidelines
  - Rubrics
  - Anchors with more than 96 audio tracks as samples for training and calibration
- **Section 5—Support an Opinion**
  - Prompting and scoring guidelines
  - Rubrics
  - Anchors with more than 127 audio tracks as samples for training and calibration
- **Section 6—Retell a Narrative**
  - Prompting and scoring guidelines
  - Rubric
  - Anchors with more than 96 audio tracks as samples for training and calibration
- **Section 7—Present and Discuss Information**
  - Prompting and scoring guidelines
  - Rubrics
  - Anchors with more than 78 audio tracks as samples for training and calibration
- **Section 8—Summarize an Academic Presentation**
  - Prompting and scoring guidelines
  - Rubrics
  - Anchors with more than 168 audio tracks as samples for training and calibration
- **Section 9—Video Scoring Practice**
  - Seven full Speaking videos for scoring practice of an entire administration by grade level or grade span
- **Section 10—K–1 Administration**
  - Narrated training video
  - PowerPoint talking points slides

7.4.1.3.2. Training Videos

Five test administration videos were created and presented during statewide training; these were made available with the other training materials. Videos used are listed in table 7.1:

Table 7.1 Available Scoring Training Videos

Topic	Description
Kindergarten and grade one	The video, which includes narration, presents a kindergarten student being administered all four domains and a grade one student being administered the Writing domain.
Reading, grades three through twelve	The narrated video was recorded with high school students being administered the Reading domain in a group setting.

Table 7.1 (*continued)*

Topic	Description
Writing, grades three through twelve	The narrated video was recorded with high school students being administered the Writing domain in a group setting.
Grade two administration	This narrated grade two video incorporates small-group directions and testing of grade two Reading, Writing, and Listening.
Listening	This narrated video includes test administration practices and was recorded with middle school students taking the Listening domain with audio streamed through TOMS.
Video scoring practice	Seven full Speaking administration videos were created; each presents a student in a different grade level or grade span being administered the Speaking test.

7.4.1.3.3. Training Presentations

Ten training presentations were created for LEA ELPAC trainers to use for local training. These training presentations included all of the Speaking video and audio files to be embedded into the presentations. Most of these presentations focused on training and scoring the Speaking task types.

Table 7.2 includes a list of the training presentations available to LEAs.

Table 7.2 Available Training Presentations for Speaking

Binder Section	Training Presentations
Section 1	Introduction Training Presentation
Section 2	Test Administration Training Presentation
Section 3	Talk About a Scene Training Presentation
Section 4	Speech Functions Training Presentation
Section 5	Support an Opinion Training Presentation
Section 6	Retell a Narrative Training Presentation
Section 7	Present and Discuss Information Training Presentation
Section 8	Summarize an Academic Presentation Training Presentation
Section 9	Scoring Video Practice (one presentation per grade level or grade span)
Section 10	K–1 Administration of Reading-Writing-Listening

7.4.1.3.4. Online Training Resources

Moodle provides a password-protected, online platform where course materials can be developed and made available. The ELPAC Moodle Training Site provides California LEAs with necessary training resources to train test examiners to score the ELPAC. There were 20,472 users as of the close of the Summative ELPAC test administration window on May 31, 2018.

To give test examiners an opportunity to refresh and test their knowledge prior to administering the Summative ELPAC, the online training site included more than 53 training and calibration quizzes with more than 400 audio samples.

To access the ELPAC Moodle Training Site, LEA users required individual user accounts. Each LEA had its own district group; the LEA ELPAC coordinator was issued a unique enrollment key for the training course and could view the results of the quizzes taken by test examiners to monitor scoring calibration.

The training quizzes allowed a test examiner to listen to the audio, select a score, and receive feedback. The Moodle quiz provided the correct score, justification, and feedback after the test examiner completed 10 samples.

For items that included artwork, such as *Retell a Narrative* and *Present and Discuss Information*, the picture stimulus was included in the quiz for the test examiner's reference while listening to the audio. A replay feature allowed the test examiner to replay the audio as necessary.

Upon completion of the calibration quiz, the "Pass/Fail" and "Percent correct" notifications were posted for the test examiner.

Table 7.3 shows a list of the training and calibration quizzes by task type created and posted to the Moodle Training Site.

Table 7.3 Training and Calibration Quizzes by Task Type

Task Type	Training Quizzes	Calibration Quizzes
Talk About a Scene	• Kindergarten video quiz • Grade 1 video quiz • Grade 2 video quiz • Grades 3–5 video quiz • Grade 6–8 video quiz • Grade 9–10 video quiz • Grade 11–12 video quiz	[None]
Speech Functions	• Grades 2–12 • Grades 2–5 • Grades 6–8 • Grades 9–12	• Grades 2–12 • Grades 2–5 • Grades 6–8 • Grades 9–12
Support an Opinion	• Grades K–2 • Grades 3–5 • Grades 6–8 • Grades 9–12 • Grades 3–12	• Grades K–2 • Grades 3–5 • Grades 6–8 • Grades 9–12 • Grades 3–12
Retell a Narrative	• Kindergarten • Grade 1 • Grade 2 • Grades 3–5	• Kindergarten • Grade 1 • Grade 2 • Grades 3–5
Present and Discuss Information	• Grades 6–8 • Grades 9–10 • Grades 11–12	• Grades 6–8 • Grades 9–10 • Grades 11–12

Table 7.3 *(continued)*

Task Type	Training Quizzes	Calibration Quizzes
Summarize an Academic Presentation	• Kindergarten • Grade 1 • Grade 2 • Grades 3–5 • Grades 6–8 • Grades 9–10 • Grades 11–12	• Kindergarten • Grade 1 • Grade 2 • Grades 3–5 • Grades 6–8 • Grades 9–10 • Grades 11–12

7.4.2. Scorer Qualifications for Speaking

The Speaking domain was scored by test examiners "in the moment." All test examiners were required to receive the Speaking scoring training from an LEA trainer.

7.5. Types of Scores

7.5.1. Raw Scores

Raw scores for each domain were obtained by summing the number of MC and machine-scorable CR items answered correctly and adding the total number of points obtained on the hand-scored CR items within the Speaking and Writing domains and the kindergarten (K) and grade one Reading domain.

The domain raw scores from Listening and Speaking were summed to compute the oral language skill raw score. The domain raw scores from Reading and Writing were summed to compute the written language skill raw score. The number and percentage of students at each raw score, and the associated level, are reported for each domain in table 7.A.1 through table 7.A.52 in appendix 7.A.

7.5.2. Scale Scores

Raw scores are not directly comparable from administration to administration because each raw score is based on a set of items that may differ in difficulty. Student performance on the ELPAC is reported in terms of scale scores that express student proficiency in terms of a constant metric. Thus, a scale score of 1350 in one language skill area in one administration represents the same level of proficiency as 1350 on the same language skill area in another administration, even though each scale score may represent a different raw score.

ELPAC scale scores are expressed as four-digit numbers that range from 1150 to 1950 across grade levels and grade spans. Lower scores indicate lesser proficiency and higher scores indicate greater proficiency.

7.5.2.1 Scale Score Conversions

For each language skill area, the following steps are used to establish the raw-score-to-scale-score relationship. The process begins by inverting the test characteristic curve (Stocking, 1996) where each possible raw score is mapped to a corresponding theta score. These theta scores represent a student's ability level on a particular language skill and are transformed onto their respective language skill area through a linear transformation as described in equation 7.1.

$$\text{Scale score} = \text{Intercept} + \text{Slope} \times (\text{theta score}) \quad (7.1)$$

Refer to subsection *11.5.6 Developing Summative ELPAC Reporting Scales* for applicable scaling constraints (e.g., slope and intercept terms) for converting theta scores to the oral language and written language scales.

Through this process, raw-to-scale-score conversion tables are established. The complete raw-to-scale score conversion tables for oral and written language skills are presented in the tables in appendix 8.D.

7.5.2.2 Overall Scale Score

The overall scale score is calculated as the weighted average of the scale scores of the oral and written language skills scale scores. For K, the overall scores are calculated as the weighted average scores of the two composite scores as shown in equation 7.2:

$$0.70 \times \text{Oral language skill score} + 0.30 \times \text{Written language skill score} \quad (7.2)$$

For grades one through twelve, the overall scores are calculated as the average scores of the two composite scores as shown in equation 7.3:

$$0.50 \times \text{Oral language skills score} + 0.50 \times \text{Written language skills score} \quad (7.3)$$

The frequency distribution of raw score, scale score, and level for composite language skills are presented in appendix 7.B, in table 7.B.1 through table 7.B.26. Additionally, appendix 7.C provides the overall scale score distribution for each grade.

Refer to subsection *11.5.6 Developing Summative ELPAC Reporting Scales* for more details of regarding how the Summative ELPAC reporting scales were established.

7.5.3. ELPAC Levels

Reporting scales for the Summative ELPAC's two composite language skills classify each student's performance into one of the four levels, which are as follows:

1. Level 1—Beginning stage of developing English skills
2. Level 2—Somewhat developed English skills
3. Level 3—Moderately developed English skills
4. Level 4—Well developed English skills (indicating the highest level of performance)

Appendix 7.D provides a summary of student ELPAC levels for each of the composite language skills. Each table presents the number and percentage of students at each ELPAC level for K through grade twelve.

To guide the interpretation of the scale scores for each domain, the range of possible scale scores for each domain is divided into three levels:

1. Level 1—Beginning
2. Level 2—Somewhat/Moderately Developed
3. Level 3—Well Developed

Appendix 7.E provides a summary of student levels for each language domain. Each table presents the number and percentage of students at each level for K through grade twelve.

The scale score ranges defining the various levels and grade levels or grade spans are presented in table 7.4.

Table 7.4 Composite Language Skills and Overall Reporting Scale Score Ranges for Each Reporting Level by Grade Level and Grade Spans

Grade Level or Grade Span	Test	Level 1	Level 2	Level 3	Level 4
Kindergarten	Overall	1150–1378	1379–1413	1414–1443	1444–1700
Kindergarten	Oral Language	1150–1389	1390–1417	1418–1450	1451–1700
Kindergarten	Written Language	1150–1351	1352–1402	1403–1427	1428–1700
Grade 1	Overall	1150–1414	1415–1436	1437–1466	1467–1700
Grade 1	Oral Language	1150–1411	1412–1432	1433–1461	1462–1700
Grade 1	Written Language	1150–1416	1417–1439	1440–1471	1472–1700
Grade 2	Overall	1150–1419	1420–1446	1447–1488	1489–1700
Grade 2	Oral Language	1150–1409	1410–1436	1437–1476	1477–1700
Grade 2	Written Language	1150–1429	1430–1455	1456–1500	1501–1700
Grade span 3–5	Overall	1150–1458	1459–1489	1490–1538	1539–1800
Grade span 3–5	Oral Language	1150–1438	1439–1465	1466–1511	1512–1800
Grade span 3–5	Written Language	1150–1477	1478–1513	1514–1565	1566–1800
Grade span 6–8	Overall	1150–1472	1473–1510	1511–1553	1554–1900
Grade span 6–8	Oral Language	1150–1435	1436–1477	1478–1531	1532–1900
Grade span 6–8	Written Language	1150–1509	1510–1543	1544–1575	1576–1900
Grade span 9–10	Overall	1150–1486	1487–1530	1531–1580	1581–1950
Grade span 9–10	Oral Language	1150–1446	1447–1497	1498–1549	1550–1950
Grade span 9–10	Written Language	1150–1525	1526–1563	1564–1610	1611–1950
Grade span 11–12	Overall	1150–1487	1488–1529	1530–1587	1588–1950
Grade span 11–12	Oral Language	1150–1445	1446–1480	1481–1541	1542–1950
Grade span 11–12	Written Language	1150–1528	1529–1578	1579–1633	1634–1950

7.6. Overview of Score Aggregation

The summary performance for the two composite language skills and overall scores for selected groups of students are provided in appendix 7.F. In table 7.F.1 through table 7.F.39. Students are grouped by demographic characteristics, including gender, ethnicity, economic status (disadvantaged or not), migrant status, and special education services status. For each student group, the number tested, scale score means, standard deviations, and the percentage of students in each level are reported. To protect student privacy, when the number of students in a student group is 10 or fewer, the summary statistics are not reported and are presented as “NA.”

Table 7.5 provides definitions of demographic student groups.

Table 7.5 Demographic Student Groups Reported

Category	Student Groups
Gender	• Male • Female
Ethnicity	• American Indian or Alaska Native • Asian • Native Hawaiian or Other Pacific Islander • Filipino • Hispanic or Latino • Black or African American • White • Two or more races
Special Education Service Status	• Students not receiving special education services • Students receiving special education services
Economic Status	• Not economically disadvantaged • Economically disadvantaged
Migrant Status	• Eligible for the Title I Part C Migrant Program (Migrant) • Not eligible for the Title I Part C Migrant Program (Nonmigrant)

7.7. Reports Produced and Scores for Each Report

7.7.1. Online Reporting

The Test Operations Management System (TOMS) is a secure website hosted by ETS that permits LEA users to manage aspects of the ELPAC administration and report delivery. This system used a role-specific design to restrict access to certain tools and applications based on the user's designated role. Specific functions of TOMS included the following:

- Manage user access privileges
- Manage test material orders
- Run and download various reports

7.7.2. Special Cases

All students identified as English learners (ELs) were required to take the Summative ELPAC. There were no special cases that excuse a student from receiving a score. In instances where a student's individualized education program or Section 504 plan specified that the student had a disability for which there were no appropriate accommodations for assessment in one or more of the Speaking, Listening, Reading, and Writing domains, the student was assessed in the remaining domains in which it was possible to assess the student, per the *Code of Federal Regulations*, Title 34, Section 200.6.

Note the following about special reporting cases:

- A student may have been assigned an overall score only if assessed in both oral and written language. To be considered as having been assessed in oral language, the student must have been assessed in either Speaking or Listening. To be considered

as having been assessed in written language, the student must have been assessed in either Reading or Writing.

- A valid score could only be provided in those instances where the student tested in at least one of the domains for oral language and written language. In all instances where the overall score resulted in NS, the student counts as tested, without a valid score.

7.7.3. Types of Score Reports

The following is a list of score reports produced for the 2017–18 Summative ELPAC:

- SSR—The SSR was the official score report for the parents or guardians and described the student's results.
- Student data files and corresponding aggregate files—Aggregate files were used for public web reporting through DataQuest and for CDE apportionment. LEA student data files were generated monthly to coincide with the SSRs.
- Production data files—The production data file was the full operational file and included 100 percent of the student scores and eligibility data. This file was provided to the CDE.

7.7.3.1 Student Score Reports

The SSR was the official score report for the parents or guardians and describes the student's results. For the 2017–18 administration, LEAs received two printed versions. These reports were also available as PDFs the LEA could download from TOMS. The SSR included the following:

- Overall score and reporting level
- Oral language score and reporting level
- Written language score and reporting level
- Domain levels

As mentioned previously, overall score, the oral language score and written language score placed a student within one of the four ELPAC reporting levels as Beginning, Somewhat Developed, Moderately Developed, or Well Developed. Each domain score placed a student within one of three proficiency levels as Beginning, Somewhat/Moderately Developed, and Well Developed.

7.7.3.2 School Reports

Schools received SSRs as well as score report labels for the student file. Additionally, site ELPAC coordinators could download a file of student results for the school from TOMS.

7.7.3.3 LEA Reports

LEAs had the option of downloading the following ELPAC reports from TOMS:

- LEA student data files
- LEA-level aggregate files

7.7.4. Score Report Applications

Summative ELPAC results provided parents and guardians with information about their child's progress toward English proficiency. The results were a tool for increasing communication and collaboration between parents or guardians and teachers.

Summative ELPAC results were one of the components schools could use to help make decisions about how best to support student progress. Summative ELPAC results, however, should never be used as the only source of information to make important decisions about a child's education.

7.7.5. Criteria for Interpreting Test Scores

An LEA may use ELPAC results to help make decisions about student placement in programs that support the student's ongoing development toward English proficiency. However, it is important to remember that a single test can provide only limited information. Other relevant information should be considered as well. It is advisable for parents or guardians to evaluate their child's progress by looking at classroom work and progress reports in addition to the child's ELPAC results.

LEAs may use ELPAC results to help make decisions about student placement in EL programs, student exit from EL programs, and student growth in proficiency while in EL programs. The ELPAC, however, is a single measure of student performance and is intended to be used in combination with other relevant information in the decision-making process. Test scores must be interpreted cautiously when making decisions about student or program performance.

2017–18 Summative ELPAC reporting levels represented broad ranges of proficiency with wide gradations between the lowest and highest possible scores in each range that were reflected in student performance. While statistical procedures were carefully applied to ensure a continuous scale throughout the full range of the common scale, ETS recommends using caution in comparing individual student performance across nonadjacent grade spans. Although the common scales have the same general properties across domains, numeric comparisons across domains cannot be made—a student scoring 400 in oral language and 420 in written language is not necessarily doing better in terms of written skills.

7.7.6. Criteria for Interpreting Score Reports

Summative ELPAC scores represented only one view of a child's progress toward language proficiency. It is advisable for parents or guardians to evaluate their child's progress by looking at classroom work and progress reports in addition to the child's ELPAC results before making reclassification decisions.

Because the Summative ELPAC results were vertically scaled, scale scores for a test may be compared to scale scores for the same student or groups of students in different years, as well as for between specific grade levels. This allows users to say that achievement for a given grade was higher or lower one year as compared with another. For example, the grade two Summative ELPAC scale scores in 2017–18 and 2018–19 may be compared, as can the grade five Summative ELPAC scale score in 2017–18 and the grade six Summative ELPAC scale score in 2018–19, because of the vertical scale.

Reference

Stocking, M. L. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics, 21*, 365–89.

Chapter 8: Test Analyses and Results

This chapter summarizes the item- and test-level statistics from the analyses conducted for the 2017–18 operational administration of the English Language Proficiency Assessments for California (ELPAC).

8.1. Background

This chapter provides information on the psychometric analyses of the 2017–18 Summative ELPAC operational data. It describes the data samples used for statistical analyses and presents the results of the item and test analyses, such as classical item analyses and differential item functioning (DIF). It includes explanations for all statistical procedures implemented during the psychometric analyses, including reliability estimates, standard errors of measurement and decision consistency, and accuracy of the performance-level classifications. Information on the procedures designed to ensure the validity of score uses and interpretations is also provided.

8.1.1. Summary of the Analyses

Each of these sets of analyses for the Summative ELPAC is presented in the body of the text and in the listed appendices.

1. **Classical Item Analyses (IA).** Classical item analysis for the Summative ELPAC is discussed in subsection *8.2 Classical Item Analysis Statistics*. Appendix 8.A presents results of the classical item analyses, including item difficulty indices, item-total correlation coefficients, and the omission rates for multiple choice (MC) and constructed-response (CR) items. In addition, the distribution of score points for the CR with multiple score points and the machine-score MC items is provided.
2. **Item Response Theory (IRT) Analyses.** IRT analyses, including calibration, are elaborated in subsection *8.3 Item Response Theory (IRT) Analyses*. Appendix 8.B includes distribution of IRT *a*-values, *b*-values, and item statistics by domain.
3. **Differential Item Functioning (DIF) Analyses.** DIF analysis is described in subsection *8.4 Differential Item Functioning (DIF)*. Table 8.12 presents the results of the DIF analyses for all items of the Summative ELPAC.
4. **Reliability Analyses.** Reliability estimation is illustrated in subsection *8.5 Reliability Analyses*. The following results of the analyses are presented:
    - Appendix 8.C provides results of the reliability analyses of total test scores for the for selected student groups of interest (e.g., gender, ethnicity, etc.).
    - Appendix 8.D presents the raw-score-to-scale-score conversion tables with the conditional standard errors of measurement (CSEM) for the oral and written language skill.
    - Appendix 8.E shows interrater reliability statistics showing the agreement between two raters.
    - Appendix 8.F presents statistics describing the decision accuracy and decision consistency of the performance classifications.

8.1.2. Samples for the Analyses

In general, analyses included in the technical report are based on all valid students' scores in the tested population. The actual data sample used depended on the availability date and content of the data file. Additionally, a student data file was selected to meet an analysis timeline. Students taking the braille version were excluded from these item analyses.

Table 8.1 shows the number of students tested by grade level. The data includes the Summative ELPAC population comprised of students who have been identified as English learners in kindergarten through grade twelve. The N-counts here may not match those in other reports, nor will they always match those shown in other tables and appendices of this report, due to different reporting specifications requiring demographic information that may be missing from some records. Students with an include indicator of "Y" in table 8.1 were used for the chapter 8 analyses. Table 8.1 also presents the number of excluded students using braille at each grade level.

Table 8.1 Number of Students Tested by Include Indicator and Grade Level

Grade Level	Y	T	Total Number Tested	Braille Count
Kindergarten	175,789	737	176,526	2
Grade 1	145,762	754	146,516	4
Grade 2	127,304	772	128,076	6
Grade 3	113,495	794	114,289	7
Grade 4	103,695	852	104,547	8
Grade 5	90,517	820	91,337	7
Grade 6	78,158	820	78,978	2
Grade 7	66,055	776	66,831	7
Grade 8	54,843	824	55,667	1
Grade 9	51,515	806	52,321	1
Grade 10	48,327	740	49,067	1
Grade 11	41,769	693	42,462	1
Grade 12	33,239	1,759	34,998	2

Note: "Y" indicates students who were enrolled during the active testing window and completed the test. "T" indicates students who were receiving special education services and were tested with an alternate assessment for any or all domains.

8.2. Classical Item Analysis Statistics

Many of the statistics that are in common use for evaluating tests, such as *p*-values, point-biserial correlations, DIF, and reliability coefficients arise from classical test theory. These item analyses were conducted for each item across all domains. The students who took the braille version were excluded from these item analyses.

Detailed results of these item analyses are presented in appendix 8.A and are summarized in the tables in this chapter.

8.2.1. Description of Classical Item Analysis Statistics

The classical item analyses include the item difficulty indices and the item-total correlation indices. Flagging rules associated with these statistics identify items that are not performing as expected. The omit rate of each item, the proportion of test takers choosing each distractor, the correlation of each distractor with the total score, and the distribution of students at each score point for the polytomous items are also included in the classical item analyses.

8.2.1.1. Classical Item Difficulty Indices (*p*-value)

For multiple choice (MC) items, item difficulty is indicated by the *p*-value, which is the proportion of students who answer an item correctly. The range of *p*-values is from 0.00 to 1.00. Items with higher *p*-values are easier items; those with lower *p*-values are more difficult items.

The formula for *p*-value for an MC item is:

$$p\text{–}value_{MC} = \frac{\sum X_{ic}}{N_i}, \quad (8.1)$$

where,

X_{ij} is the score received for a given MC item I for student j, and

N_i is the total number of students who were presented with item i.

For CR items, difficulty is indicated by the average item score (AIS). The AIS can range from 0.00 to the maximum total possible points for an item. To facilitate interpretation, the AIS values for CR items or machine-scorable CR items are often expressed as the proportion of the maximum possible score, which is analogous to the *p*-values of dichotomous items.

For CR items, the *p-value* is defined as:

$$p\text{–}value_{CR} = \frac{\sum X_{ij}}{N_i \times Max(X_i)}, \quad (8.2)$$

where,

X_{ij} is the score received for a given CR item i for student j,

$Max\ (X_i)$ is the maximum score for item i, and

N_i is the total number of students who were presented with item i.

The Summative ELPAC *p*-values were generally within the expected range of above 0.20 and below 0.95; most were also in the desired difficulty range of 0.30 to 0.90. These ranges were defined to produce items that discriminate most effectively throughout the range of student proficiency.

Mean item *p*-values are presented in table 8.2.

Table 8.2 Mean *p*-values

Grade Level or Grade Span	Listening Mean *p*-value	Speaking Mean *p*-value	Reading Mean *p*-value	Writing Mean *p*-value
Kindergarten	0.771	0.750	0.694	0.772
Grade 1	0.784	0.807	0.722	0.782
Grade 2	0.776	0.825	0.731	0.787
Grade span 3–5	0.702	0.846	0.572	0.679
Grade span 6–8	0.665	0.833	0.470	0.702
Grade span 9–10	0.649	0.789	0.512	0.691
Grade span 11–12	0.641	0.817	0.492	0.684

8.2.1.2. Item-Total Correlation

An important indicator of item discrimination is the point-biserial correlation (i.e., item-total correlation), defined as the correlation between student scores on an individual item and student "total" scores on the test (after subtracting out the scores of the item in question). They are included in the item analysis tables in appendix 8.A.

To calculate point-biserial correlations by domain, the total scores are, instead, domain scores. In general, the item-total correlation ranges from -1.0 (for a perfect negative relationship) to 1.0 (for a perfect positive relationship). A relatively high positive item-total correlation is desired, as it indicates that students with higher scores on the test tended to perform better on the item than students with lower test scores. A negative item-total correlation typically signifies a problem with the item, because it indicates that students with low scores on the test are getting higher scores on the item than students with high scores on the test.

To avoid artificially inflating the correlation coefficients, the contribution of the item in question was first removed from the total when calculating each of the correlations. Thus, performance on each Listening item was correlated with the total Listening score minus the score on the item in question. Likewise, performance on each Reading item was correlated with the total Reading score minus the score on the item in question, and so on for the Speaking and Writing items. Table 8.3 reports the mean point-biserial correlations by grade span and domain.

Desired values for this correlation are positive and larger than 0.20. Negative item-total correlations indicate that low-ability students obtain higher scores on the item than high-ability students, an indication that the scoring key may be incorrect. Items with item-total correlations below 0.20 were flagged for review.

Table 8.3 Mean Point-Biserial Correlation

Grade Level or Grade Span	Listening Mean Point-Biserial Correlations	Speaking Mean Point-Biserial Correlations	Reading Mean Point-Biserial Correlations	Writing Mean Point-Biserial Correlations
Kindergarten	0.488	0.713	0.510	0.787
Grade 1	0.476	0.659	0.614	0.684
Grade 2	0.430	0.641	0.500	0.672
Grade span 3–5	0.405	0.679	0.439	0.691
Grade span 6–8	0.352	0.709	0.385	0.696
Grade span 9–10	0.426	0.792	0.422	0.708
Grade span 11–12	0.392	0.748	0.383	0.666

8.2.2. Summary of Classical Item Analysis Flagging Criteria

Items are flagged for review if the item analysis yields any of the following results, including both MC and CR items:

1. The *p*-value is above 0.95.
4. The *p*-value is below 0.20.
5. Item-total correlation (point-biserial) is below 0.20.
6. Among the highest-performing students (the top 20 percent), the number of students choosing any distractor is greater than the number choosing the key.
7. The omit rate is above 5 percent.

Educational Testing Service (ETS) psychometric staff and content assessment development staff carefully reviewed each of the items flagged after the 2017–18 Summative ELPAC administration and summarized the results for the California Department of Education (CDE), with recommendations for subsequent analyses. These results were also entered into the item bank and used by the assessment development team for test assembly for future operational administrations.

8.2.3. Classical Item Analysis Results Summary

This subsection presents tables of the classical item analysis results for the 2017–18 test items. Table 8.4 presents *p*-value and item-total correlation information by grade level and grade span as well as the number of unique items in each test.

Detailed results of the item analyses for each item by grade level and grade span are presented in appendix 8.A. The item statistics, including *p*-value, point-biserial correlation, and item type, are included in those tables. The distribution of item scores on each CR item is presented in table 8.A.23 through table 8.A.24.

Table 8.4 Classical Item Statistics for Each Domain

Grade Level or Grade Span and Domain	No. of Unique Items	Mean *p*-value	Minimum *p*-value	Maximum *p*-value	Mean Point-Biserial Correlation
Kindergarten Listening	20	0.771	0.301	0.936	0.488
Kindergarten Speaking	10	0.750	0.537	0.892	0.713
Kindergarten Reading	14	0.694	0.448	0.848	0.510
Kindergarten Writing	8	0.772	0.706	0.878	0.787
Grade 1 Listening	22	0.784	0.557	0.934	0.476
Grade 1 Speaking	10	0.807	0.642	0.971	0.659
Grade 1 Reading	20	0.722	0.475	0.913	0.614
Grade 1 Writing	7	0.782	0.670	0.946	0.684
Grade 2 Listening	22	0.776	0.413	0.964	0.430
Grade 2 Speaking	13	0.825	0.562	0.957	0.641
Grade 2 Reading	26	0.731	0.387	0.933	0.500
Grade 2 Writing	7	0.787	0.701	0.926	0.672
Grade span 3–5 Listening	22	0.702	0.278	0.907	0.405
Grade span 3–5 Speaking	13	0.846	0.695	0.965	0.679
Grade span 3–5 Reading	26	0.572	0.231	0.856	0.439
Grade span 3–5 Writing	6	0.679	0.590	0.752	0.691
Grade span 6–8 Listening	22	0.665	0.386	0.854	0.352
Grade span 6–8 Speaking	13	0.833	0.622	0.970	0.709
Grade span 6–8 Reading	26	0.470	0.202	0.692	0.385
Grade span 6–8 Writing	6	0.702	0.578	0.798	0.696
Grade span 9–10 Listening	22	0.649	0.386	0.813	0.426
Grade span 9–10 Speaking	13	0.789	0.592	0.928	0.792
Grade span 9–10 Reading	26	0.512	0.291	0.766	0.422
Grade span 9–10 Writing	6	0.691	0.562	0.790	0.708
Grade span 11–12 Listening	22	0.641	0.425	0.807	0.392
Grade span 11–12 Speaking	13	0.817	0.681	0.930	0.748
Grade span 11–12 Reading	26	0.492	0.254	0.773	0.383
Grade span 11–12 Writing	6	0.684	0.585	0.769	0.666

8.2.4. Omit Rates

For both MC and CR items, examining item omission is useful for identifying potential problems with test features such as testing time and item or test layout. Omit rates are often useful in determining whether testing times are sufficient, particularly if there is a high rate of items omitted at the end of a test section. In the case of the Summative ELPAC, where speed is not an issue because the test is untimed, high item omit rates may indicate extreme item difficulty.

The Summative ELPAC omit rates tended to be low, with the highest values for students in kindergarten. Omit rates were generally highest for the Writing domain. Table 8.5 reports the mean omit rates by grade span and domain.

Table 8.5 Mean Omit Rates

Grade Level or Grade Span	Listening Mean Percent of Items Omitted	Speaking Mean Percent of Items Omitted	Reading Mean Percent of Items Omitted	Writing Mean Percent of Items Omitted
Kindergarten	0.59	0.68	0.52	4.69
Grade 1	0.45	0.37	2.14	2.50
Grade 2	0.49	0.75	0.97	0.96
Grade span 3–5	0.27	0.80	0.48	1.04
Grade span 6–8	0.18	0.83	0.48	0.95
Grade span 9–10	0.29	1.76	0.63	1.84
Grade span 11–12	0.31	1.26	0.61	1.46

8.3. Item Response Theory Analyses

Item response theory (IRT) is built upon the item response function, which describes the probability of a given response as a function of a test-taker's true ability. IRT can be used to implement item calibrations, link item parameters, scale test scores across different forms or test administrations, evaluate item performance, build an item bank, and assemble test forms.

The two-parameter logistic (2PL) IRT model was used for the Summative ELPAC item calibration. In particular, the generalized partial credit (GPC) model (Muraki, 1992) was applied to both dichotomous and polytomous items. The mathematical formula of the GPCM is the following:

$$P_{ih}(\theta_j) = \begin{cases} \dfrac{\exp(\sum_{v=1}^{h} Da_i(\theta_j - b_i + d_{iv}))}{1 + \sum_{c=1}^{n_i} \exp(\sum_{v=1}^{c} Da_i(\theta_j - b_i + d_{iv}))}, & \text{if score } h = 1, 2, ..., n_i \\ \dfrac{1}{1 + \sum_{c=1}^{n_i} \exp(\sum_{v=1}^{c} Da_i(\theta_j - b_i + d_{iv}))}, & \text{if score } h = 0 \end{cases} \quad (8.3)$$

where,

$P_{ih}(\theta_j)$ is the probability of student with proficiency θ_j obtaining score h on item i;

n_i is the maximum number of score points for item i;

a_i is the discrimination parameter for item i ;

b_i is the location parameter for item i;

d_{iv} is the category parameter for item i on score v; and

D is a scaling constant of 1.7 that makes the logistic model approximate the normal ogive model.

Preequated grade-level or grade-span test forms were administered for the 2017–18 Summative ELPAC. These preequated test forms were based on calibrations and linking analyses conducted during the spring 2016–17 field test (refer to appendix 11.A for the field test IRT data). All IRT analyses results for the 2017–18 operational Summative ELPAC, including calibration and linking, are shown in appendix 8.B.

The overall summary of the IRT *a*-value (discrimination) parameter estimates—refer to equation (8.3)—used on the 2017–18 Summative ELPAC oral language and written language skills tests are shown in table 8.6 and table 8.7. The overall summary of the IRT *b*-value (item difficulty) parameter estimates are shown in table 8.8 and table 8.9 for the Summative ELPAC oral language and written language skills tests. The mean, standard deviation (SD), minimum, and maximum values are presented, in addition to the number of items for each language domain.

Table 8.6 IRT *a*-values (Discrimination Parameter) for 2017–18 Oral Language Skill Tests by Grade Level and Grade Span

Grade Level or Grade Span	Domain	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Kindergarten	Listening	20	0.56	0.15	0.26	0.88
Kindergarten	Speaking	10	1.10	0.20	0.84	1.44
Grade 1	Listening	22	0.61	0.17	0.21	0.86
Grade 1	Speaking	10	1.04	0.22	0.83	1.41
Grade 2	Listening	22	0.56	0.23	0.17	1.30
Grade 2	Speaking	13	0.74	0.13	0.52	0.96
Grade span 3–5	Listening	22	0.43	0.12	0.20	0.63
Grade span 3–5	Speaking	13	0.70	0.14	0.46	1.02
Grade span 6–8	Listening	22	0.24	0.11	0.06	0.46
Grade span 6–8	Speaking	13	0.75	0.22	0.50	1.24
Grade span 9–10	Listening	22	0.23	0.07	0.07	0.32
Grade span 9–10	Speaking	13	0.63	0.15	0.45	0.89
Grade span 11–12	Listening	22	0.21	0.07	0.09	0.38
Grade span 11–12	Speaking	13	0.58	0.10	0.40	0.71

Table 8.7 IRT *a*-values (Discrimination Parameter) for 2017–18 Written Language Skill Tests by Grade Level and Grade Span

Grade Level or Grade Span	Domain	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Kindergarten	Reading	14	0.52	0.30	0.27	1.31
Kindergarten	Writing	8	1.69	0.34	1.15	2.27
Grade 1	Reading	20	0.92	0.23	0.52	1.60
Grade 1	Writing	7	0.92	0.42	0.57	1.75
Grade 2	Reading	26	0.81	0.30	0.29	1.33
Grade 2	Writing	7	0.76	0.21	0.45	1.03
Grade span 3–5	Reading	26	0.60	0.23	0.15	1.05
Grade span 3–5	Writing	6	0.60	0.10	0.42	0.71
Grade span 6–8	Reading	26	0.40	0.13	0.13	0.59
Grade span 6–8	Writing	6	0.58	0.08	0.47	0.64
Grade span 9–10	Reading	26	0.39	0.12	0.14	0.65
Grade span 9–10	Writing	6	0.54	0.06	0.45	0.61
Grade span 11–12	Reading	26	0.35	0.15	0.04	0.63
Grade span 11–12	Writing	6	0.52	0.10	0.41	0.65

Table 8.8 IRT *b*-values (Item Difficulty Parameter) for 2017–18 Oral Language Skill Tests by Grade Level and Grade Span

Grade Level or Grade Span	Domain	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Kindergarten	Listening	20	-3.42	1.04	-5.01	-0.38
Kindergarten	Speaking	10	-2.77	0.54	-3.40	-1.81
Grade 1	Listening	22	-2.80	0.75	-4.59	-1.65
Grade 1	Speaking	10	-2.42	0.75	-3.86	-1.56
Grade 2	Listening	22	-2.33	1.30	-3.98	1.30
Grade 2	Speaking	13	-2.25	0.59	-3.66	-1.03
Grade span 3–5	Listening	22	-1.82	1.49	-3.98	2.56
Grade span 3–5	Speaking	13	-2.17	0.62	-3.43	-1.19
Grade span 6–8	Listening	22	-1.62	1.70	-3.44	3.88
Grade span 6–8	Speaking	13	-1.80	0.85	-3.05	-0.21
Grade span 9–10	Listening	22	-1.52	1.87	-3.93	5.10
Grade span 9–10	Speaking	13	-1.56	0.79	-3.29	0.28
Grade span 11–12	Listening	22	-1.52	1.02	-4.00	0.81
Grade span 11–12	Speaking	13	-1.65	0.84	-3.22	-0.33

Table 8.9 IRT *b*-values (Item Difficulty Parameter) for 2017–18 Written Language Skill Tests by Grade Level and Grade Span

Grade Level or Grade Span	Domain	Number of Items	Mean	Standard Deviation	Minimum	Maximum
Kindergarten	Reading	14	-4.27	1.23	-6.46	-1.86
Kindergarten	Writing	8	-3.87	0.35	-4.43	-3.56
Grade 1	Reading	20	-2.41	0.48	-3.45	-1.48
Grade 1	Writing	7	-2.90	0.71	-4.14	-2.20
Grade 2	Reading	26	-1.87	0.69	-3.05	0.16
Grade 2	Writing	7	-2.38	0.56	-3.21	-1.88
Grade span 3–5	Reading	26	-0.49	1.13	-3.51	1.68
Grade span 3–5	Writing	6	-1.04	0.82	-2.16	-0.27
Grade span 6–8	Reading	26	1.01	1.39	-0.72	5.01
Grade span 6–8	Writing	6	-0.72	0.80	-1.75	0.17
Grade span 9–10	Reading	26	0.81	0.79	-1.08	2.23
Grade span 9–10	Writing	6	-0.53	0.95	-1.81	0.63
Grade span 11–12	Reading	26	2.07	3.70	-0.33	18.94
Grade span 11–12	Writing	6	-0.47	0.88	-1.54	0.72

The distributions of the IRT *a*-values and *b*-values for all operational items appearing on the 2017–18 test forms are provided in table 8.B.1 through table 8.B.4 in appendix 8.B. In addition, table 8.B.5 through table 8.B.11 provide the IRT discrimination, difficulty, and step parameter estimates at the item level for each grade level or grade span for both oral language and written language skills.

8.4. Differential Item Functioning (DIF)

In addition to the classical item analyses, gender DIF analyses were conducted for the Summative ELPAC. The sample size requirements for the DIF analyses were 700 in the combined focal and reference groups and 300 in the smaller of the two groups.

If an item performs differentially across identifiable student groups—e.g., gender—when students are matched on ability, the item may be measuring something else other than the intended construct (i.e., possible evidence of DIF). It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills between groups (i.e., impact) or statistical Type I error, which might falsely find DIF in an item. As a result, DIF analysis is used mainly as a statistical tool to identify *potential* item bias. Subsequent reviews by content experts and bias and sensitivity experts are required to determine the source and meaning of performance differences.

There are many possible reasons for DIF. The wording of an item, for example, may be such that one group interprets the question differently than the other, or the reading demands of the items are such that, although reading is not being measured (e.g., in a mathematics test), reading differences between the groups lead to differential outcomes on the item.

The Summative ELPAC DIF procedures used were the Mantel-Haenszel (MH) procedure (1959) for the MC items and the standardized mean difference (SMD) procedure (Dorans, 1989) for the CR items.

8.4.1. Multiple-choice Items

The MH-DIF statistic was calculated for MC items (Mantel & Haenszel, 1959; Holland & Thayer, 1985). Using the total domain raw score as the criterion score, students in each domain score category in the focal group (e.g., females) are compared with examinees in the same theta score category in the reference group (e.g., males).

For the MH-DIF, the examinees are split into a focal group, which is typically of prime interest, and a reference group. Each group is then further divided into K matched ability groups, often on the basis of total test raw score. That is, all examinees obtaining a raw score of 10 represented one matched ability group, for example. Then for an item, j, the data from the kth level of reference and focal group members can be arranged as a 2 × 2 table as shown in figure 8.1.

Group	Item j Correct	Item j Incorrect	Total
Reference Group	A_k	B_k	n_{Rk}
Focal Group	C_k	D_k	n_{Fk}
Total Group	R_k	W_k	n_{Tk}

Figure 8.1 MH Data Structure

The MH odds ratio estimate, αMH, for item j compares the two groups in terms of their odds of answering the item correctly and is given as follows:

$$\alpha_{MH} = \frac{\sum_k \frac{A_k D_k}{n_{Tk}}}{\sum_k \frac{B_k C_k}{n_{Tk}}} \tag{8.4}$$

To facilitate the interpretation of MH results, the common odds ratio is frequently transformed to the delta scale using the following formula (Holland & Thayer, 1988):

$$\Delta_{MH} = -2.35 \ln\left[\alpha_{MH}\right] \tag{8.5}$$

Δ_{MH} is negative when the item is more difficult for members of the focal group than it is for the comparable members of the reference group. Δ_{MH} is positive when the item is more difficult for members of the reference group than it is for the comparable members of the focal group.

MC items are assigned one of three DIF classifications shown in table 8.10.

Table 8.10 DIF Categories for MC and CR with 1-point Score Item

DIF category	Criteria
C (large)	Δ_{MH} is at least 1.5 and is significantly greater than 1.0.
B (moderate)	Δ_{MH} is at least 1.0 and is significantly greater than 0.0.
A (negligible)	Otherwise

Items with a "C" classification will not be used in the creation of future forms. In these cases, the items were not originally flagged with "C" DIF during field test item analyses but are now flagged with "C" DIF because the underlying student populations changed. During form construction, items with a "B" classification are used only when necessary to meet test specifications.

8.4.2. Constructed-Response Items

The standardization DIF (Dorans & Schmitt, 1993; Zwick, Thayer, & Mazzeo, 1997; Dorans, 2013), in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959), is used to identify polytomous items with DIF. The SMD compares the item means of the two groups after adjusting for differences in the distribution of students across the values of the matching variable, using the total domain raw score as the criterion score. The SMD statistic is computed using the following formula:

$$SMD = \frac{\sum_{m=1}^{M} N_{fm} \times (E_f(Y \mid X = m) - E_r(Y \mid X = m))}{\sum_{m=1}^{M} N_{fm}} = \frac{\sum_{m=1}^{M} N_{fm} \times D_m}{\sum_{m=1}^{M} N_{fm}} \quad (8.6)$$

where,

X = the criterion score,

Y = the item score,

M = the number of score categories on X,

N_{fm} = the number of students in the focal group in score category m,

E_r = the expected item score for the reference group,

E_f = the expected item score for the focal group, and

D_m = the expected item score difference between the focal group and the reference group in score category m.

These indexes are indicators of the degree to which members of one gender group perform better or worse than expected on each CR item or each machine-scorable CR item.

CR items and machine-scorable CR items are also assigned one of three DIF classifications.

A positive SMD value means that, conditional on the criterion score, the focal group has a higher mean item score than the reference group. In contrast, a negative SMD value means

that, conditional upon the criterion score, the focal group has a lower mean item score than the reference group.

CR items are assigned one of three DIF classifications shown in table 8.11.

Table 8.11 DIF Categories for CR or Polytomous Items

DIF category	Criteria
C (large)	$p_{x^2_{MH}}$ is less than .05, and $\frac{SMD}{sd}$ is greater than .25
B (moderate)	$p_{x^2_{MH}}$ is less than .05, and $\frac{SMD}{sd}$ is greater than .125.
A (negligible)	Otherwise

These classifications were defined to be in alignment with the MC classifications in terms of stringency (Zwick, Thayer, and Mazzeo, 1997). Items with a "C" classification will not be used in the creation of future forms, and items with a "B" classification will be used only when necessary to meet test specifications.

8.4.3. Classification

Based on the DIF statistics and significance tests, items were classified into three categories and assigned values of A, B, or C. Category A items contained negligible DIF, Category B items exhibited slight to moderate DIF, and Category C items possessed moderate to large DIF. Items with a Category C will not be used in the creation of future forms.

Table 8.12 presents the summary of the DIF analysis and shows that there is no item with Category C by DIF by gender for each domain.

Table 8.12 Gender DIF Classification

Grade Level or Grade Span and Domain	Category C+	Category B+	Category A	Category B-	Category C-	Total Number of Items
Kindergarten Listening	0	1	19	0	0	20
Grade 1 Listening	0	0	21	1	0	22
Grade 2 Listening	0	1	20	1	0	22
Grade span 3–5 Listening	0	1	21	0	0	22
Grade span 6–8 Listening	0	0	22	0	0	22
Grade span 9–10 Listening	0	0	22	0	0	22
Grade span 11–12 Listening	0	0	22	0	0	22
Kindergarten Speaking	0	0	10	0	0	10
Grade 1 Speaking	0	0	10	0	0	10
Grade 2 Speaking	0	0	13	0	0	13
Grade span 3–5 Speaking	0	0	13	0	0	13
Grade span 6–8 Speaking	0	0	13	0	0	13

Table 8.12 *(continued)*

Grade Level or Grade Span and Domain	Category C+	Category B+	Category A	Category B-	Category C-	Total Number of Items
Grade span 9–10 Speaking	0	0	13	0	0	13
Grade span 11–12 Speaking	0	0	13	0	0	13
Kindergarten Reading	0	0	14	0	0	14
Grade One Reading	0	0	20	0	0	20
Grade Two Reading	0	0	26	0	0	26
Grade span 3–5 Reading	0	1	24	1	0	26
Grade span 6–8 Reading	0	0	26	0	0	26
Grade span 9–10 Reading	0	0	26	0	0	26
Grade span 11–12 Reading	0	0	26	0	0	26
Kindergarten Writing	0	0	8	0	0	8
Grade 1 Writing	0	0	7	0	0	7
Grade 2 Writing	0	0	7	0	0	7
Grade span 3–5 Writing	0	0	6	0	0	6
Grade span 6–8 Writing	0	0	6	0	0	6
Grade span 9–10 Writing	0	0	6	0	0	6
Grade span 11–12 Writing	0	0	6	0	0	6

8.5. Reliability Analyses

The reliability for a particular group of students' test scores estimates the extent to which the scores would remain consistent if those same students were retested with another parallel version of the same test. If the test includes CR items, reliability extends to an evaluation of the extent to which the students' scores would remain consistent if both the items and the scorers were changed.

8.5.1. Internal Consistency Reliability

The reliability coefficient cannot, in fact, be computed directly unless the student actually takes two parallel versions of the same test. However, with some reasonable assumptions, reliability can be estimated from the students' responses to a single version of the test.

Like other statistics, the reliability coefficient can vary substantially from one group of students to another. It tends to be larger in groups that are more diverse in the ability measured by the test and smaller in groups that are more homogeneous in the ability measured.

The Summative ELPAC test reliabilities were evaluated for each domain and the composite scores by the coefficient alpha (Cronbach, 1951) index of internal consistency, which is calculated as

$$\hat{\alpha} = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k}\hat{\sigma}_i^2}{\hat{\sigma}_X^2}\right] \quad (8.7)$$

where,

k is the number of items on test form,

$\hat{\sigma}_i^2$ is the estimated variance of item i, and

$\hat{\sigma}_X^2$ is the estimated total test variance.

Table 8.13 presents reliability coefficients for each domain and composite score of the test by grade level or grade span. The reliability coefficients ranged from 0.70 to 0.94. For grades three through twelve, the oral language skill had higher reliability coefficients than the written language skill. However, for the lower grade levels, the written language skill had slightly higher reliability coefficients than the oral language skill.

Table 8.13 Reliability Coefficient of Domains and Composite Scores

Grade Level or Grade Span	Listening: Reliability-Coefficient Alpha	Speaking: Reliability-Coefficient Alpha	Reading: Reliability-Coefficient Alpha	Writing: Reliability-Coefficient Alpha	Oral: Reliability-Coefficient Alpha	Written: Reliability-Coefficient Alpha
Kindergarten	0.828	0.876	0.790	0.913	0.900	0.902
Grade 1	0.837	0.847	0.915	0.822	0.894	0.926
Grade 2	0.798	0.878	0.886	0.802	0.893	0.910
Grade span 3–5	0.775	0.897	0.842	0.808	0.893	0.885
Grade span 6–8	0.699	0.903	0.784	0.816	0.882	0.851
Grade span 9–10	0.816	0.944	0.830	0.841	0.927	0.883
Grade span 11–12	0.801	0.938	0.804	0.827	0.919	0.867

The reliabilities of each domain and composite scores were also examined for various student groups from the population. Table 8.C.1 through table 8.C.7 present the reliabilities for the student groups based on gender, ethnicity, economic status, migrant status, and students receiving special education services status.

8.5.2. Standard Error of Measurement (SEM)

The standard error of measurement (SEM) is a measure of how much students' scores would vary from the scores they would earn on a perfectly reliable test. If it were possible to compute the error of measurement for each student's score in a large group of students, these errors of measurement would have a mean of zero. These standard errors of measurement would be an indication of how much the errors of measurement are affecting the students' scores. This statistic is the SEM.

The SEM is expressed in the same units as the test score, whether the units are in raw score or scale score points. In a large group of students, approximately two-thirds of the students will earn scores within one SEM of the scores they would earn on a perfectly reliable test.

The SEM is the square root of the error variance in the scores, that is, the standard deviation of the distribution of the differences between students' observed scores and their true scores. The SEM is calculated by:

$$SEM = SD\sqrt{1-\alpha} \tag{8.8}$$

where,

a is the reliability estimated in equation 8.8 for two composite scores of oral and written, and

SD is the standard deviation of the total score of oral score or composite scores (either theta or scale score).

For grades one through twelve, the SEM for the overall score is calculated according to the formula

$$SEM_{overall} = \sqrt{.5^2 SEM_{Oral}^2 + .5^2 SEM_{Written}^2} \tag{8.9}$$

and for K,

$$SEM_{overall} = \sqrt{.7^2 SEM_{Oral}^2 + .3^2 SEM_{Written}^2} \tag{8.10}$$

These SEM values are shown in table 8.14. The range of raw score standard errors for the Summative ELPAC were between 1.18 and 3.17 points across all grade levels and domains. In general, this translated into an error band of about two raw score points in most domains. For example, if a student received a raw score of 25 with a standard error of 2.00 points, upon retesting, the student would be expected to obtain a score between 23 and 27 about two-thirds of the time.

Table 8.14 SEM based on Classical Test Theory

Grade Level or Grade Span	SEM—Listening Raw Score	SEM—Speaking Raw Score	SEM–Reading Raw Score	SEM—Writing Raw Score	SEM—Oral Raw Score	SEM—Written Raw Score	SEM—Overall Raw Score
Kindergarten	1.592	1.796	1.653	1.182	2.581	2.174	1.921
Grade 1	1.688	1.725	1.582	1.577	2.537	2.333	1.723
Grade 2	1.658	1.846	1.913	1.509	2.625	2.527	1.822
Grade span 3–5	1.872	1.818	2.169	1.591	2.782	2.836	1.986
Grade span 6–8	2.038	1.941	2.311	1.508	3.019	2.918	2.099
Grade span 9–10	2.008	1.962	2.279	1.553	3.173	2.944	2.164
Grade span 11–12	2.055	1.932	2.297	1.528	3.139	2.927	2.146

It is important to remember that assessments are not perfectly reliable and only offer an estimate of what the student is capable of in a specified domain. As table 8.15 shows, the SEM scale score values for oral and written language skills averaged about 23 scale score points and 17 scale score points for overall.

Table 8.15 SEM Based on Scale Score

Grade Level or Grade Span	SEM—Oral	SEM—Written	SEM—Overall
Kindergarten	19.733	26.449	15.930
Grade 1	19.948	21.988	14.844
Grade 2	19.026	19.299	13.550
Grade span 3–5	20.730	17.974	13.719
Grade span 6–8	26.863	22.500	17.521
Grade span 9–10	30.467	25.405	19.834
Grade span 11–12	30.940	30.809	21.832

8.5.3. Conditional Standard Error of Measurement (CSEM)

Classical test theory assumes that the standard error of a test score is constant throughout the score range. While the assumption is probably reasonable in the midscore ranges, it is less reasonable at the extremes of the score distribution. IRT expands the concept by providing estimates of the standard error at each score point on the distribution.

The item response theory, or conditional SEM for scale scores, is defined as

$$\text{CSEM}(SS) = a\frac{1}{\sqrt{\text{I}\left(\hat{\theta}\right)}} \tag{8.11}$$

where,

$SS = a \times \theta + b$,

CSEM(SS) is the conditional standard of measurement on the scale score scale, and

a and b are the scaling constants (the slope and intercept) needed to transform theta to the scale score metric.

$I(\hat{\theta})$ is the test information function at ability level $\hat{\theta}$. For student j, test information is calculated as

$$I(\theta_j) = \sum_{i=1}^{n} I_i(\theta_j) \tag{8.12}$$

where,

$I_i(\theta_j)$ is the item information of item i for student j.

Item information is calculated as

$$I_i(\theta_j) = [s_{i2}(\theta_j) - s_i^2(\theta_j)] \tag{8.13}$$

where,

$S_i(\theta_j)$ is the expected item score for item i on a theta score θ_j calculated as

$$s_i(\theta_j) = \sum_{h=0}^{n_i} hp_{ih}(\theta_j), \tag{8.14}$$

and

$$s_{i2}(\theta_j) = \sum_{h=0}^{n_i} h^2 p_{ih}(\theta_j) \tag{8.15}$$

where,

$P_{ih}(\theta_j)$ is the probability of an examinee with θ_j getting score h on item i, the computation of which is shown in equation 8.3, and

n_i is the maximum.

The item response theory's SEM has an inverse normal distribution in which SEM values decrease as scores move toward the center of the range. Conditional SEM values are

reported as part of the raw-score-to-scale-score conversion tables presented in appendix 8.D for the oral and written language skills.

CSEMs varied across the scale, and are typically smaller in scale score units toward the center of the scale where more items are located and typically larger at the extreme ends of the scale.

8.5.4. Writing Score Reliability

Rater consistency is critical to the scores of ELPAC writing items and their interpretations. When two trained raters independently assign the same score (or rating) to an item response, there is evidence that the scoring standard is being applied consistently. Double scoring substantially increases the reliability of the scoring process. When used to monitor and evaluate the accuracy of rating, 15 percent of the responses are rated twice, by two independent raters. Interrater reliability is evaluated empirically by computing the percentage of exact agreement between two raters.

Evidence that the raters' scores are consistent helps to support the inference that the scores have the intended meaning. The exact agreement data collected is used to evaluate interrater agreement. Table 8.16 presents the range of exact agreement by grade level or grade span.

Table 8.16 The Range of Exact Agreement by Grade Level or Grade Span

Grade Level or Grade Span	Minimum of Percent Exact	Maximum of Percent Exact	Average of Percent Exact
Kindergarten	0.92	0.98	0.96
Grade 1	0.67	0.99	0.84
Grade 2	0.66	0.99	0.85
Grade span 3–5	0.57	0.86	0.72
Grade span 6–8	0.56	0.90	0.72
Grade span 9–10	0.59	0.89	0.73
Grade span 11–12	0.57	0.89	0.72

Table 8.E.1 in appendix 8.E provides interrater agreement statistics for each Writing domain item on the 2017–18 Summative ELPAC.

8.5.5. Decision Classification Analyses

While the reliabilities of performance-level classifications, which are criterion referenced, are related to the reliabilities of the test scores on which they are based, they are not identical. Glaser (1963) was among the first to draw attention to this distinction, and Feldt and Brennan (1989) extensively reviewed the topic. While test reliability evaluates the consistency of test scores, decision classification reliability evaluates the consistency of classification.

Consistency in classification represents how well two versions of an assessment with equal difficulty agree in their classification of students (Livingston & Lewis, 1995). This is

estimated by using actual response data and total test reliability from an administered form of the assessment from which two parallel versions of the assessment are statistically modeled and classifications compared. Decision consistency, then, is the extent to which the test classification of examinees into mastery levels agrees with classifications based on a hypothetical parallel test. The examinees' scores on the second form are modeled statistically.

Note that the values of all indexes depend on several factors, such as the reliability of the actual test form, distribution of scores, number of cut scores, and location of each cut score. The probability of a correct classification is the probability that the classification the examinee received is consistent with the classification that the examinee would have received on a parallel form. This is akin to the exact agreement rate in interrater reliability. The expectation is that this probability would be high.

Decision accuracy is the extent to which the test's classification of examinees into levels agrees with the examinees' true classification. The examinees' true scores—and, therefore, true classification—are not known but can be modeled. Consistency and accuracy are important to consider in concert. The probability of accuracy represents the agreement between the observed classification based on the actual test form and true classification, given the modeled form.

Commonly used indexes for decision consistency and accuracy include (a) decision consistency and accuracy at each cut score, (b) overall decision consistency and accuracy across all cut scores, and (c) coefficient kappa.

Cohen's kappa (Fleiss and Cohen, 1973) represents the agreement of the classifications between two parallel versions of the same test, taking into account the probability of a correct classification by chance. It measures how the test contributes to the classification of examinees over and above chance classifications. In general, the value of kappa is lower than the value of the probability of correct classification because the probability of a correct classification by chance is larger than zero.

The methodology used for estimating the reliability of classification decisions described in Livingston and Lewis (1995) is implemented using the ETS-proprietary computer program RELCLASS-COMP (Version 4.14).

Overall decision accuracy and consistency—that is, classification across all cut scores—are reported in table 8.17. Cohen's kappa statistics were substantially lower than accuracy and consistency.

Table 8.17 Classification Consistency and Accuracy for Composite Language

Grade Level or Grade Span	Oral Accuracy	Oral Consistency	Oral Kappa	Written Accuracy	Written Consistency	Written Kappa
Kindergarten	0.761	0.680	0.546	0.773	0.703	0.575
Grade 1	0.785	0.715	0.545	0.780	0.704	0.593
Grade 2	0.818	0.750	0.565	0.780	0.698	0.580
Grade span 3–5	0.756	0.675	0.533	0.728	0.633	0.491
Grade span 6–8	0.754	0.670	0.519	0.689	0.595	0.455
Grade span 9–10	0.780	0.709	0.597	0.719	0.630	0.479
Grade span11–12	0.784	0.712	0.583	0.728	0.636	0.491

Results of classification consistency and accuracy are reported in appendix 8.F by grade level or grade span and composite language skills.

References

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 2, 217–33.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd edition) (pp. 105–46). New York: Macmillan.

Fleiss, J. L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613–19.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist,* 18. 519–32

Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty* (Research Report 85–43). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 129–45). Hillsdale, NJ: Erlbaum.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement, 32*, 179–97.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analyses of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–48.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*(4), 321–44.

Chapter 9: Validity

9.1. Validity of the ELPAC Test Design

The Summative English Language Proficiency Assessments for California (ELPAC) was developed in accordance with the criteria for test development, administration, and use described in the *Standards for Educational and Psychological Testing* (2014) adopted by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

Test validation is an ongoing process, beginning at initial conceptualization and continuing throughout the lifetime of the assessment. Every aspect of an assessment provides evidence in support of its validity (or evidence to the contrary), including design, content requirements, item development, and psychometric quality. "Validity refers to the degree to which evidence and theory support the interpretations made from test scores. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound, scientific basis for the proposed score interpretations" (AERA, APA, & NCME, 2014, p. 9).

9.1.1. Purpose of the ELPAC

The Summative ELPAC was designed and developed to provide scores representing English language proficiency performance levels for required educational decision making as defined by the test purposes in the California *Education Code (EC)* Section 313. The primary inferences from the test results, in general, include (a) the proficiency level of individual students and (b) English language development (ELD) program effectiveness based on the results of groups of students.

Progress can be tracked over years and grades. The results can be used to analyze the strengths and weaknesses of students' growth in the four domains measured and to report progress to parents. The results can also be used as one body of evidence in making administrative decisions about ELD program effectiveness, class grouping, needs assessment, and placement in EL programs.

9.1.2. The Constructs to be Measured

The Summative ELPAC is designed to show how well students perform relative to the *California English Language Development Standards, Kindergarten Through Grade 12* (2012 ELD Standards) (California Department of Education [CDE], 2014). The standards describe the English language proficiency (ELP) knowledge, skills, and abilities that students are expected to acquire at each grade. The Summative ELPAC test blueprints describe the assessment task types that the students perform, the number of items per task type, and the alignment of the items to the 2012 ELD Standards (Educational Testing Service [ETS], 2017).

EC Section 60810 specifies that the state ELP assessment shall measure the language domains of Listening, Speaking, Reading, and Writing. The test blueprints describe the assessment task types and the number of items that are used to assess students' ELP in each language domain.

The Summative ELPAC provides three scale scores and placement within one of four levels. The oral language scale score and reporting level are drawn from the Listening and Speaking results. The written language scale score and reporting level are drawn from the

Reading and Writing results. The overall scale score and reporting level are derived from the oral language composite (Listening and Speaking) and the written language composite (Reading and Writing).

At grades one through twelve, the overall scale score is derived from the equal weighting of the oral language composite and the written language composite. At kindergarten, where students are developing foundational literacy skills, the overall scale score is derived from differential weighting in which 70 percent of the overall scale score comes from the oral language composite (Listening and Speaking) and 30 percent of the overall scale score comes from the written language composite (Reading and Writing).

In addition, assessment results are used to place students within one of three levels in each of the four domains of Listening, Speaking, Reading, and Writing.

9.1.3. The Interpretations and Uses of the Scores

Student scores were delivered to LEAs and used as one criterion for considering whether a student will be reclassified as fluent English proficient. *EC* Section 313(f) describes four criteria that are used to establish reclassification policies and procedures:

1. Assessment of language proficiency using an objective assessment instrument, including, but not limited to, the ELPAC
2. Teacher evaluation, including, but not limited to, a review of the student's curriculum mastery
3. Parental opinion and consultation
4. Comparison of student performance in basic skills against an empirically established range of performance in basic skills based upon the performance of English proficient students of the same age.

Summative ELPAC data also is used to calculate the English Language Progress Indicator for the California School Dashboard. ELPAC results are now used for federal accountability as required by Title I.

9.1.4. The Intended Population

The ELPAC is the required state test for ELP that must be given to students whose primary language is a language other than English. The test-taking population for the Summative ELPAC includes students who have been formally identified as English learners (ELs) in kindergarten through grade twelve based upon the results from the initial assessment, which was the California English Language Development Test during the 2017–18 academic year. ELs continue to take the Summative ELPAC each year to monitor their ELP until they are reclassified as fluent English proficient.

Students with disabilities who cannot take one or more domains of the ELPAC with allowed universal tools, designated supports, or accommodations take an alternate assessment(s), as noted in their individualized education program.

9.2. Validity of the ELPAC Test Content

9.2.1. Description of the State Standards

The 2012 ELD Standards were developed and approved by the California State Board of Education in 2012 and then published in 2014. The 2012 ELD Standards describe the key knowledge, skills, and abilities that students who are learning English need to access, engage with, and achieve in grade-level academic content. The 2012 ELD Standards provide a framework to guide the development of ELD assessment systems that help California educators ensure that all ELs make progress in the English language knowledge, skills, and abilities needed to become college- and career-ready.

9.2.2. Item Writing Guidelines

Item Writing Guidelines were developed to define the task types and content of the items. They were used as a key reference document during item-writer training to provide guidance to item writers and drive consistency and efficiency in item development. The Item Writing Guidelines were intended to facilitate the development of comparable items that measure appropriate skills and content aligned with the 2012 ELD Standards.

9.2.3. Test Blueprints

Test blueprints describe the content of the Summative ELPAC and include four tables with information about the task types in each of the four language domains of Listening, Speaking, Reading, and Writing. Task types are individual items or sets of items that require a student to perform an activity to elicit information about the student's ELP.

The test blueprints provide information about the number of items and points that were administered per task type within each grade level and domain. The test blueprints also provide the alignment of task types with the 2012 ELD Standards (CDE, 2017).

9.2.4. Form Assembly Process

The form assembly process began with the creation of test development specifications, which described the content characteristics, the psychometric characteristics, and the quantity of items to be used in the 2017–18 Summative ELPAC. ETS created the test development specifications that the CDE then reviewed and approved.

After the test development specifications were approved, ETS assessment specialists assembled the tests in the ETS Item Banking Information System (IBIS) according to the specifications. IBIS was then used to generate form planners, which are spreadsheets that contain essential item information such as the number of items, the alignment of items according to the 2012 ELD Standards, and the keys to multiple-choice items. ETS assessment specialists and psychometricians reviewed the form planners before they were delivered to the CDE. CDE staff reviewed the form planners. After ETS made any necessary edits, the CDE approved the form planners. After approval, the form planners were used as the basis for developing the test materials needed to administer the Summative ELPAC: Examiner's Manuals, Test Books, Answer Books, and audio recordings.

9.3. Validity of the ELPAC's Internal Structure

Internal structure evidence evaluates the strength or salience of the major dimensions underlying an assessment using indices of measurement precision such as fairness and differential item functioning (DIF) analysis, test reliability, and reliability of performance classifications.

9.3.1. Fairness and Differential Item Functioning

9.3.1.1 Bias and Sensitivity Reviews

To develop test materials that are fair and unbiased to all students, ELPAC test items underwent reviews by Bias and Sensitivity Review Panels from August 3, 2016, through August 5, 2016. Eighteen California educators reviewed the text and artwork of more than 2,000 newly developed items. Items were approved as is, approved with revisions, or rejected. As described in section *3.2.5 Item Review Panels*, the educators added value to the item pool by revising items to make them fair and unbiased measures of ELP.

9.3.1.2 Differential Item Functioning

DIF analyses were conducted to identify differences in item performance by student gender. There were no items identified as having significant levels of DIF for any domain. Refer to subsection *8.4 Differential Item Functioning (DIF)* for a description the DIF analyses and for the results of the DIF analyses performed on Summative ELPAC items.

9.3.2. Reliability

9.3.2.1 Overall Reliability Estimates

The results of reliability analyses on the four domains and two composite scores are presented in table 8.13. The results indicate that the reliability estimates for each domain of the test were moderately high, ranging from 0.70 to 0.94 across grade level or grade span. For the oral and written composite scores, the reliability estimates were high, ranging from 0.85 to 0.93 across grade level or grade span.

9.3.2.2 Subgroup Reliability Estimates

The reliabilities are also computed for various student groups. The student groups considered were based on gender, ethnicity, economic status, migrant status, and special education service status. Reliability estimates for each domain and composite scores are reported for each student group in table 8.B.1 through table 8.B.7.

9.3.2.3 Interrater Reliability

Interrater reliability is evaluated by computing the percentage of exact agreement between two raters. Refer to subsection *8.5.4 Writing Score Reliability* for a description of agreement analysis and to appendix 8.E, where the results of the analyses are reported.

9.3.2.4 Reliability of Performance Classifications

The methodology used for estimating the reliability of classification decisions is evaluated with the decision classification analyses in subsection *8.5.5 Decision Classification Analyses*. The results of these analyses are presented in appendix 8.F.

9.3.3. Other Validity Evidence

Convergent and discriminant validity evidence can also be established through a pattern of high correlations among scales that purport to measure domains that are known to be closely related, and lower correlations among scales that purport to measure dissimilar domains. The pattern of correlations within the Summative ELPAC provides preliminary validity evidence by showing that the correlations among oral and written language skills are positive and reasonably high. These correlations for each domain and composite score by grade level or grade span are presented in appendix 9.A.

Table 9.1 provides the correlations between composite scores and overall scores.

Table 9.1 Correlation Among Two Composites and the Overall Score

Grade Level or Grade Span	Composite	Written	Overall
Kindergarten	Oral	0.579	0.943
Kindergarten	Written	1.000	0.818
Grade 1	Oral	0.514	0.829
Grade 1	Written	1.000	0.906
Grade 2	Oral	0.583	0.878
Grade 2	Written	1.000	0.901
Grade span 3–5	Oral	0.656	0.926
Grade span 3–5	Written	1.000	0.893
Grade span 6–8	Oral	0.654	0.935
Grade span 6–8	Written	1.000	0.880
Grade span 9–10	Oral	0.693	0.951
Grade span 9–10	Written	1.000	0.882
Grade span 11–12	Oral	0.713	0.944
Grade span 11–12	Written	1.000	0.905

References

California Department of Education. (2014). *California English language development standards: Kindergarten through grade 12*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

Educational Testing Service. (2017). *Summative Assessment test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from https://www.cde.ca.gov/ta/tg/ep/documents/elpacadoptedblueprints.pdf

Educational Testing Service. (2015). *Proposed test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/ta/tg/ep/documents/elpacblueprts.pdf

Chapter 10: Quality-Control Procedures

10.1. Quality Control of Item and Test Development

The California Department of Education (CDE) and Educational Testing Service (ETS) implemented rigorous quality control procedures throughout the test development, administration, scoring, analyses, and reporting processes for the Summative English Language Proficiency Assessments for California (ELPAC). As part of this effort, ETS staff worked with the ETS Office of Professional Standards Compliance, which publishes and maintains the *ETS Standards for Quality and Fairness* (ETS, 2014). These *Standards* support the goals of delivering technically sound, fair, and useful products and services; and assisting the public and auditors in evaluating those products and services. Quality control procedures are outlined in this chapter.

10.1.1. Quality Control of Item Writing

After the CDE approved newly developed items for field testing, ETS performed a final review of the items in the ETS Item Banking System called final content review. During this review, an assessment specialist who was familiar with the Summative ELPAC task types performed an independent review of each item to ensure that the item content, metadata, graphics, and audio files were all accurate. The assessment specialist also reviewed comments that were made during previous reviews to ensure that they were implemented. All items were reviewed and approved at final content review before field testing.

10.1.2. Quality Control of Item Selection

Both ETS assessment development staff and statistical analysis staff checked the prior use of items to ensure that items of the appropriate status were used as equating items, operational items, and field test items. Cross checks were also performed to ensure that none of the items placed on an operational form appeared in a public-facing document, such as a practice test.

10.2. Quality Control of Test Materials

After the CDE approved all printed test materials (i.e., Examiner's Manuals, Test Books, and Answer Books), ETS performed a final certification check of the test materials. For each test form, ETS staff ensured that the various test materials worked together and that all cross references regarding page numbers and question numbers were accurate. The test-length audio files were also checked as part of this process. In each case, the final certification check was completed and any needed revisions were applied before the Examiner's Manuals, Test Books, and Answer Books were delivered to the printers for reproduction, and before the test-length audio files were uploaded into the ELPAC Test Operations Management System.

10.2.1. Test Administration Manuals

ETS staff consulted with internal subject-matter experts and conducted validation checks to verify that test instruction manuals accurately matched the test booklets and testing processes. Copy editors and content editors reviewed each document for spelling, grammar, accuracy, and adherence to CDE style. Manuals such as the *2017–18 Summative ELPAC Test Administration Manual* were approved by the CDE before they were published to the ELPAC website at https://www.elpac.org/. Only nonsecure documents were posted to this website.

10.2.2. Processing Test Materials

Upon receipt of the test materials, ETS personnel examined each shipment for a number of conditions, including physical damage, shipping errors, and omission of materials. The number of students recorded on the Group Information Sheet (GIS)—the precoded identification sheet that accompanied the grade-level test materials for a school—was compared to the number of Answer Books returned to ETS.

ETS' image-scanning process, which captured security information electronically and compared scorable material quantities reported on the GIS to actual documents scanned, was used when processing returned Answer Books. LEAs were contacted by phone if there were any missing shipments or the quantity of materials returned was less than expected.

10.3. Quality Control of Test Delivery

ETS used several methods to manage and monitor the security of the ELPAC paper-based test materials. First, all secure test materials were coded with an individual label that identified the item and the number of materials being packed in a shipment, thus allowing ETS to track materials from the time they left the warehouse until they were returned for scoring.

Materials were shipped using United Parcel Service (UPS) or, for larger orders, they were shipped via freight. In either case, tracking numbers were used to track these shipments until they were securely delivered at the LEA's warehouse.

10.3.1. Quality Control of Test Assignment

State and federal law (California *Education Code* sections 313 and 60810 and federal law Titles I and Ill of the Every Student Succeeds Act) require that all students whose primary language is other than English be assessed for English language proficiency.

Local educational agencies (LEAs) have a role in ensuring students identified as English learners (ELs) are administered the Summative ELPAC annually. Students who were ELs in spring 2018 as a result of taking the California English Language Development Test were required to take the Summative ELPAC. Those ELs must be administered the Summative ELPAC annually until they are reclassified as fluent English proficient.

Proficiency classifications (e.g., initial fluent English proficient, EL) are found in the California Longitudinal Pupil Achievement Data System—the data system used to main student data—in the English Language Acquisition Status (ELAS) field. Students who take the Summative ELPAC have an ELAS of "EL."

10.3.2. Quality Control of Test Administration

During the Summative ELPAC administration, every person who either worked with the assessments, communicated test results, or received testing information was responsible for maintaining the security and confidentiality of the tests, including CDE staff, ETS staff, ETS subcontractors, LEA ELPAC coordinators, site ELPAC coordinators, and teachers.

ETS' Code of Ethics requires that all test information, including tangible materials (e.g., test items and test books), confidential files (e.g., those containing personally identifiable student information), and processes related to test administration (e.g., the packing and delivery of test materials) are kept secure. For the 2017–18 Summative ELPAC, ETS had systems in place that maintained tight security for test items, test books, and test results, as well as for student data.

To ensure security for all the tests that ETS develops or handles, ETS maintains an Office of Testing Integrity.

10.3.3. Quality Control of Machine Scoring Procedures

The quality control of paper-pencil tests is ensured by an independent group at ETS that signs into the ETS Enterprise Score Key Management (eSKM) system and checks scoring keys. This group must sign off and approve the keys before scoring for the administration can begin. This team also creates scoring stencils to be used during the administration to overlay on top of a student's Answer Book to verify the score computed by eSKM is accurate. These quality control procedures were followed during the 2017–18 Summative ELPAC administration.

10.3.4. Quality Control of Hand Scoring Procedures for Writing

Rater qualifications, rater certifications, and daily rater calibrations are all processes used to control the reliability of constructed-response (CR) scoring. For the Summative ELPAC, raters were led through a training period by trained assessment development staff, content scoring leaders, group scoring leaders, and scoring leaders for an assigned grade level and specific prompt types prior to the scoring period. In the training period, raters were trained to appropriately apply the rubrics by using the ELPAC benchmark sample papers.

Trained raters were scheduled to score in four- or eight-hour shifts. Scoring leaders were qualified raters who have the responsibility of providing feedback to raters in order to provide additional content support and offer corrective mentoring for struggling raters.

Each rater was assigned a secure user ID and password to log on to the scoring system and was required to sign a confidentiality agreement. System access for the rater was restricted to the hours that he or she was scheduled to work.

Prior to starting a shift, a rater passed a calibration test that demonstrated sufficient training in ELPAC scoring criteria and an ability to score accurately. Ten percent of responses were scored twice (i.e., "read behind") in order to check agreement among the raters. Scoring leaders read behind the raters throughout a shift and entered their own scores on responses that raters read.

Refer to subsection *7.3 Constructed Response (CR) Scoring for Writing* for details about these processes.

10.4. Quality Control of Psychometric Processes

10.4.1. Development of Scoring Specifications

A number of measures were taken to ascertain that the scoring keys were applied to the student responses as intended and the student scores were computed accurately. ETS built and reviewed the scoring system models based on scoring specifications developed by ETS and approved by the CDE. Machine-scored item responses and demographic information were collected from the Answer Books by ETS. Human-scored item responses were sent electronically to the ETS Online Network for Evaluation for scoring by trained, qualified raters. Record counts were verified against the counts obtained during security check-in from the document processing staff to ensure all students were accounted for in the file.

Once the record counts were reviewed, the machine-scored item responses were scored against the appropriate answer key. In addition, the student's original response string was stored for data verification and auditing purposes.

The scoring specifications contained detailed scoring procedures, along with the procedures for determining whether a student attempted a test and whether that student response data should be included in the statistical analyses and calculations for computing summary data. Standard quality inspections were performed on all data files, including the evaluation of each student data record for correctness and completeness. Student results were kept confidential and secure at all times.

10.4.2. Development of Scoring Procedures

The ETS eSKM scoring system utilizes scoring procedures specified by psychometricians and provides scoring services. Following scoring, a series of quality-control checks were carried out by ETS psychometricians to ensure the accuracy of each score.

10.4.2.1 Enterprise Score Key Management System Processing

ETS developed two independent and parallel scoring structures to produce students' scores: the eSKM[3] scoring system, which collected, scored, and delivered individual students' scores to the ETS reporting system; and the parallel scoring system developed by ETS Technology and Information Processing Services (TIPS), which scored individual students' responses. The two scoring systems independently applied the same scoring algorithms and specifications.

ETS psychometricians verified the eSKM scoring by comparing all individual student scores from TIPS and resolving any discrepancies. This process redundancy is an internal quality-control step and is in place to verify the accuracy of scoring. Students' scores were reported only when the two parallel systems produce identical results.

If scores did not match, the mismatch would have been investigated by ETS' Psychometrics, Statistics, and Data Science and eSKM teams and resolved. The mismatch could be a result of a CDE decision not to score an item because a problem was identified in a particular item or rubric. In cases of a mismatch, ETS applied the problem item notification (PIN) not to score the item through the systematic process in eSKM; the mismatch would be possible if TIPS was still in the process of applying the PIN in the parallel system when the student score was being compared. This real-time scoring check was designed to continually detect mismatches and track remediation.

Finally, data extracts were sent to ETS' Data Quality Services for data validation. Following validation, the student response statistical extracts were made available to the psychometricians.

10.4.2.2 Psychometric Processing

Psychometricians verified the eSKM scoring by comparing the parallel scoring programs, conducting extensive analyses to resolve any discrepancies, and verifying the accuracy of all student scores and reported results. In particular, psychometricians checked variables such as total scale scores, levels, and number of scored items. To investigate discrepancies, theta scores and completeness were also checked.

All scores complied with the ETS scoring specifications and the parallel scoring process to ensure the quality and accuracy of scoring and to support the transfer of scores into the database of the student records scoring system before student reports were generated.

[3] The eSKM system produced the ETS scores of record.

10.5. Quality Control of Reporting

To ensure the quality of Summative ELPAC test results, for both individual student and summary reports, four general areas were evaluated:

1. Report formats were compared with input sources from the CDE-approved samples.
2. Report data was validated through quality-control checks performed by ETS' Data Quality Services and Resolutions teams. Additionally, all Student Score Reports (SSRs) were run through ETS' patented QC Integrator software;
3. Production of printed reports was evaluated by verifying the print quality and comparing the number of report copies, sequence of report order, and offset characteristics to the CDE requirements.
4. Quality check and production reports were proofread by the CDE and ETS prior to any LEA mailings.

All reports were required to include a single, accurate LEA code, a charter school number (if applicable), an LEA name, and a school name. All elements conformed to the CDE's official county/district/school (CDS) code and naming records. From the start of processing through scoring and reporting, the CDS Master File was used to verify and confirm accurate codes and names. CDE provided a revised LEA Master File to ETS throughout the year as updates become available.

After the reports were validated against the CDE's requirements, a set of reports for quality-control (QC) districts were provided to the CDE and ETS for review and approval. Paper reports were sent on the actual report forms, organized as they were expected to look in production. The CDE and ETS reviewed and approved the report package after a thorough examination.

Upon the CDE's approval of the reports generated for the QC districts, ETS proceeded with the first batch of report production. The first production batch was selected to validate a subset of LEAs that contained key reporting characteristics and demographics representative of the state. The first production batch incorporated CDE-selected LEAs and provided the final check prior to generating all reports and mailing them to the LEAs.

10.5.1. Exclusion of Student Scores from Summary Reports

Students who were identified as ELs were required to take the Summative ELPAC. Students who, for medical reasons, were unable to sit through an administration received a lowest obtainable scale score. There were no exclusions on the Summative ELPAC.

10.5.2. End-to-End Testing for Operational Administration

ETS conducted end-to-end testing prior to the start of the test administration. The purpose of this testing was to verify that all systems, processes, and resources were ready for the operational administration.

To begin the quality-control process for paper-pencil test administration, the ETS resolutions team completed response documents by marking responses on Answer Books for fictitious students in selected schools and across several LEAs. They marked Answer Books with answers that were all correct, all incorrect, and other test response combinations. These response combinations were the expected results across levels and score ranges. The response booklets were sent for processing, batching, and scanning. Once released from scanning, the test results were sent through the system for scoring and reporting. SSRs

were created, along with data files for subject matter experts in the teams to review and verify.

Individual SSRs were generated based on the fictitious students when 100 percent quality control was demonstrated by ETS' Resolution staff.

Reference

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/s/about/pdf/standards.pdf

Chapter 11: Field Testing

The focus of this chapter is the results and any specific details of the field testing. Because administration of the operational Summative English Language Proficiency Assessments for California (ELPAC) mirrored the field test, this chapter describes details unique to the field test.

11.1. Purpose of the Field Testing

The purposes of the Summative ELPAC field tests were

- to produce a sufficient number of items with appropriate statistical quality to populate one operational form and one breach form per grade/grade span for the first operational administration (the 2017–18 edition);
- to establish a score scale for each grade level and grade span and place the operational and breach forms on the same scale; and
- to establish a common (i.e., vertical) scale across all seven grade/grade spans to facilitate federal accountability reporting.

11.2. Design

The Summative ELPAC field test design featured four forms that reflected the test blueprint as closely as possible, with common linking items robust enough to provide solid linking across and within grade levels and grade spans. The field test forms needed to reflect the test blueprint in test length and content as closely as possible such that the student performance data (impact data) was a good approximation of student performance in future operational administrations. Additionally, a sufficient number of items needed to be used as common items so in the event some items failed to satisfy the statistical requirements, enough linking items survived to allow robust horizontal and vertical linking.

11.2.1. Content

Four Summative ELPAC field test forms were developed using items identified for the assessment. Several guiding principles were used in designing the stand-alone Summative ELPAC field test:

- The length of the field test forms—measured in the number of items—adhered as closely as possible to the length of the operational assessment as indicated by the *Proposed Test Blueprints for the ELPAC* (CDE, 2015). All four field test forms had the same number of items with the exception of the Reading test for grade span six through eight, where form 4 had two fewer items compared to forms 1 through 3.
- The field test forms included the same task types as the operational test described by the *Proposed Test Blueprints for the ELPAC* (CDE, 2015).
- Task types that were judged to be more difficult in target grades were included as vertical linking items in the adjacent higher grade level or grade span to ensure that items were sufficiently difficult for students at the grade level or grade span above.

Linking items were used for both horizontal and vertical linking designs to construct these four field test forms.

11.2.1.1 Horizontal Linking Items

All four forms contained horizontal linking items at all grade levels and grade spans. Horizontal linking items are common items that are shared across forms within a grade level or grade span.

In most instances, horizontal linking items were shared across two of the four field test forms for a given grade level or grade span. In some cases, horizontal linking items were shared among three or even four field test forms to place items targeted for each grade level or grade span onto the same scale.

11.2.1.2 Vertical Linking Items

For grades one through twelve, two of the four field test forms contained vertical linking items from the grade level or grade span below it. Vertical linking items are items from the grade level or grade span below the one of interest. They are used to link grade level or grade span assessments together and create a single common scale across all the grade level or grade span assessments.

In the kindergarten (K) assessment, the vertical linking design took into account the "opportunity to learn," where K students were not expected to be familiar with expectations of students in grade one. K and grade one field test forms were vertically linked by K items provided to students in grade one who took grade one field test forms.

Refer to subsection *11.5.5. Vertical (Common Scale) Scaling* for additional information about vertical linking in the ELPAC field test forms.

11.2.2. Composition of Forms

There were a total of 1,169 unique items administered on the four forms of the Summative ELPAC 2016–17 field test. Table 11.1 breaks down this number by grade level or grade span and domain.

Table 11.1 Number of Unique Items Administered in the 2016–17 Summative ELPAC Field Test

Grade Level or Grade Span	Listening	Speaking	Reading	Writing
Kindergarten	48	30	37	24
Grade 1	41	29	52	30
Grade 2	48	35	64	28
Grade span 3–5	53	33	72	24
Grade span 6–8	52	36	61	24
Grade span 9–10	55	34	60	25
Grade span 11–12	55	34	60	25
Totals:	**352**	**231**	**406**	**180**

11.3. Administration

11.3.1. Sampling Plan

This section summarizes the field test form assignment and distribution to schools and LEAs for the ELPAC. The Summative ELPAC field tests took place in March and April 2017 and targeted a large sample of English learners (ELs) and a small sample of English only (EO)/ reclassified fluent English proficient (RFEP) students. The EO/RFEP student data was collected only for the standard setting study and did not use either classical item analysis or item response theory (IRT) analysis.

Table 11.2 shows the total number of students classified as EL, EO, and RFEP and initial fluent English proficient (IFEP) who took the field test on which the sampling was performed:

Table 11.2 Number of Students for Sampling

Grade Level or Grade Span	EL	EO	RFEP and IFEP
K	5,369	898	47
Grade 1	6,388	450	62
Grade 2	6,408	435	116
Grade span 3–5	6,118	816	324
Grade span 6–8	5,814	406	420
Grade span 9–10	6,323	262	361
Grade span 11–12	5,343	258	371

To accomplish the goals of the field test successfully, the following three-step student sampling and form-distribution procedures were conducted.

Step 1. Identify eligible local educational agencies (LEAs). The first step was to identify the number of schools needed for the grade levels and grade spans, so as to establish the target number for Summative ELPAC field test sampling. For the Summative ELPAC sample, Educational Testing Service (ETS) used the California English Language Development Test 2015–16 individual data as the sampling frame to recruit students from schools that represented the diverse demographic makeup and geography of California. Enrollment counts for each school at each grade level and grade span were computed and used in sampling criteria. Only LEAs that had enrollment of 30 or more students per grade level or grade span were invited to participate in the field test due to limited in-person training opportunities.

Step 2. Develop EL student rosters for LEAs. ETS then developed student rosters for those LEAs that agreed to participate in the field testing. For the individual EL student sampling, ETS used the county/district/school code to match the accepted school information and California Longitudinal Pupil Achievement Data System student information to generate student rosters for the LEAs. While developing the student roster, ETS included ELs with disabilities that were suitable for the field test. However, ETS excluded ELs with three specific disability types—intellectual disability, visual impairment, and deaf-blindness—that were not appropriate for the field test due to specific accommodations needed.

Step 3. Develop EO/RFEP student rosters for LEAs. Because a small sample of EO/RFEP students was required for the field test, a careful selection of EO/RFEP students was rostered to ensure these students were considered similarly as their EL counterparts. For each accepted school, a number of total EO/RFEP students and a number of EO/RFEP students with valid California Assessment of Student Performance and Progress (CAASPP) Smarter Balanced English language arts/ literacy (ELA) achievement-level information were counted by grade level and grade span. Because CAASPP Smarter Balanced was provided at grades three through eight and grade eleven during the March through May 2016 administration, the EO/RFEP students from grades four through nine and grade twelve had valid ELA performance-level information in December 2016. Other grades (i.e., K through grade three, grades ten and eleven) did not include ELA achievement-level information to participate in the EO/RFEP sample.

For selecting ELs with valid Smarter Balanced ELA achievement levels, the percentage of achievement levels on each grade level from the 2015–16 CAASPP Smarter Balanced summative assessments were used to select a similar proportion of EO/RFEP students with four achievement levels on the student roster.

The assigned EO/RFEP student sampling results showed that the percentage of achievement levels in each grade level and grade span matched with the targeted percentages. Note that in cases where schools provided students from multiple grades, the total count was smaller than the sum of each grade level.

Table 11.3 provides a summary of sampled EO/RFEP students with pre-identification (Pre-ID) information.

Table 11.3 Selected EO/RFEP Students with Pre-ID

Grade Level	Total N	Random Selection	CAASPP Level 1	CAASPP Level 2	CAASPP Level 3	CAASPP Level 4
Transitional Kindergarten	376	376	0	0	0	0
Kindergarten	614	614	0	0	0	0
Grade 1	493	493	0	0	0	0
Grade 2	497	497	0	0	0	0
Grade 3	312	312	0	0	0	0
Grade 4	323	0	116	84	65	58
Grade 5	321	0	125	67	68	61
Grade 6	216	0	73	47	59	37
Grade 7	265	0	75	78	77	35
Grade 8	260	0	81	65	83	31

Table 11.3 *(continued)*

Grade Level	Total N	Random Selection	CAASPP Level 1	CAASPP Level 2	CAASPP Level 3	CAASPP Level 4
Grade 9	255	0	66	74	84	31
Grade 10	250	250	0	0	0	0
Grade 11	255	255	0	0	0	0
Grade 12	265	0	53	64	87	61
Totals:	**4,702**	**2,797**	**589**	**479**	**523**	**314**

Table 11.4 provides a comparison of the grade-level populations in California who have CAASPP Smarter Balanced achievement levels with sampled EO/RFEP students who have CAASPP Smarter Balanced achievement levels. Note that K and grades two, three, ten, and eleven were not included in table 11.4 since there was no CAASPP achievement level information from the prior year's results.

Table 11.4 Comparison of Population and Sampled EO/RFEP Students with Smarter Balanced ELA Achievement Levels

Grade Level	CAASPP Level 1	CAASPP Level 2	CAASPP Level 3	CAASPP Level 4	Sampled EO/ RFEP Level 1	Sampled EO/ RFEP Level 2	Sampled EO/ RFEP Level 3	Sampled EO/ RFEP Level 4
Grade 4	36%	26%	20%	18%	36%	26%	20%	18%
Grade 5	39%	21%	21%	19%	39%	21%	21%	19%
Grade 6	34%	22%	27%	17%	34%	22%	27%	17%
Grade 7	28%	29%	29%	13%	28%	29%	29%	13%
Grade 8	31%	25%	32%	12%	31%	25%	32%	12%
Grade 9	26%	29%	33%	12%	26%	29%	33%	12%
Grade 12	20%	24%	33%	23%	20%	24%	33%	23%
Totals:	**31%**	**25%**	**28%**	**16%**	**31%**	**25%**	**28%**	**16%**

11.3.2. Testing Window

The testing window for the Summative ELPAC field test was March 6, 2017 through April 14, 2017. The field test was administered to samples of students across the state of California.

11.3.3. Field Test Administration Procedures

The Summative ELPAC field test administration mirrored the procedures of an operational administration. LEA ELPAC coordinators and ELPAC test examiners were trained on the administration of the ELPAC, including administration of the Listening and Speaking domains. Security protocols and procedures outlined in subsection *5.3 Test Security and Confidentiality* were implemented and followed.

11.3.4. Field Test Accessibility

Special test versions of the field test forms were not developed. However, LEAs were directed to review the CDE ELPAC Matrix Four (CDE, 2018) to determine the non-embedded resources their students would need in order to participate in the field test.

11.4. Scoring

ETS coordinated the scoring of student responses for the Summative ELPAC field test. After administration, the Listening, Reading, and Speaking responses that were marked in the response circles in the Answer Books were scanned and scored.

Trained test examiners scored Speaking responses “in the moment” and then marked the appropriate response circles in the Answer Book. To create the rubrics that were used, a Speaking range finding was conducted prior to the field test.

Handwritten responses from the Writing section were scanned and scored by human scorers within the ETS Online Network for Evaluation.

Information on the ETS constructed-response (CR) scoring process can be found in chapter 7.

11.5. Test Analyses and Results

For the ELPAC, ETS undertook multiple steps for test analyses. First, classical test analyses were conducted to ensure that test item keys were correct. Second, differential item function (DIF) analyses were conducted to ensure items were not biased against male or female students. After these two steps, items with poor qualities (e.g., extremely low item-total correlation or biased against a certain gender group) and items reviewed and found by the ETS content team to be of a poor quality were excluded from further analyses. The third step involved dimensionality analyses, which examined empirical evidence of how vertical scales should be developed for the Summative ELPAC. Once the two vertical scales were identified, ETS conducted calibration and scaling analyses to develop the final Summative ELPAC reporting scales.

11.5.1. Classical Item Analyses

The processes described in subsection *8.2.1 Description of Classical Item Analysis Statistics* were used to conduct classical item analyses for the field test item pool. Table 11.5 shows the number of items that were not calibrated due to poor classical statistics. Note that the Speaking and Writing domains consisted of all CR items that did not involve any score keys. No Speaking or Writing items were flagged for problematic statistics that required removal from calibration.

Table 11.5 Number of Items Not Calibrated Due to Poor Classical Statistics

Grade Level or Grade Span	Listening: N Items Not Calibrated	Listening: N Items in the Pool	Reading: N Items Not Calibrated	Reading: N Items in the Pool
K	0	48	0	37
Grade 1	0	57	0	67
Grade 2	1	62	0	81
Grade span 3–5	1	66	5	96
Grade span 6–8	0	65	2	91
Grade span 9–10	0	70	3	93
Grade span 11–12	0	70	4	93

11.5.2. Differential Item Function Analyses

This subsection describes the procedures used for DIF analyses. Table 11.6 shows the number of items flagged for C-level DIF. The items with C-level DIF flag were not used in the operational test form. No items were flagged for DIF in either the Speaking or Writing domain.

Table 11.6 Number of Items with C-DIF flag

Grade Level or Grade Span	Listening	Listening: N Items in the Pool	Reading	Reading: N Items in the Pool
K	1	48	0	37
Grade 1	3	57	0	67
Grade 2	0	62	0	81
Grade span 3–5	1	66	1	96
Grade span 6–8	1	65	1	91
Grade span 9–10	0	70	0	93
Grade span 11–12	0	70	1	93

11.5.3. Dimensionality Study

The main purposes of the Summative ELPAC dimensionality analyses were to investigate the factor structure of the Summative ELPAC and, in turn, to inform decisions for IRT scaling and score reporting (ETS, 2019). Practical considerations in test length, as well as ease in score scale maintenance over future administrations, were factors in model evaluation. The study indicated that the combination oral and written model showed reasonable fit across K through grade twelve and also supported continuous (i.e., vertical) scaling.

Results of the dimensionality study provided empirical evidence that oral and written language skills are, to some degree, distinct, and can be considered separately. It also provided validity evidence in support of the scoring hierarchy, used for reporting Summative ELPAC scores and approved by the California State Board of Education in September 2017.

11.5.4. Item Response Theory Analyses

Based on the dimensionality study, two unidimensional IRT scales were developed for each grade level or grade span assessment in the IRT calibration stage: the oral language scale, which comprised the Listening and Speaking tests; and the written language scale, which comprised the Reading and Writing tests. The two-parameter logistic model was used to calibrate dichotomous items, and the generalized partial credit model was used to calibrate polytomous items.

Calibrations and linking analyses conducted during the spring 2016–17 field test provided the basis for the preequated grade level or grade span test forms that were administered for the 2017–18 Summative ELPAC. Refer to appendix 11.A for the field test IRT *a*-value and *b*-value parameter estimates for the field test. Presented for both the oral language and

written language skills tests are the mean, standard deviation, minimum, and maximum values, in addition to the number of items for each language domain.

Refer to appendix 8.B for IRT analyses results for the 2017–18 operational Summative ELPAC. Refer also to subsection *8.3 Item Response Theory Analyses* for a description of IRT analyses that were conducted.

Table 11.7 shows the number of items, score points, and students available to support IRT calibration for the item pool. These item pools supported both operational and breach form assembly.

Table 11.7 Number of Items, Score Points, and Students for IRT Analysis

Language Skill	K	Grade 1	Grade 2	Grade Span 3–5	Grade Span 6–8	Grade Span 9–10	Grade Span 11–12
Number of Oral Items	78	99	105	111	114	118	118
Number of Written Items	61	109	116	124	125	125	124
Maximum Score Points, Oral	113	149	154	167	170	175	175
Maximum Score Points, Written	81	140	154	175	184	186	186
Number of Oral Students	5,360	6,376	6,353	6,040	5,675	6,090	5,181
Number of Written Students	5,339	6,374	6,369	6,077	5,762	6,163	5,199

11.5.5. Vertical (Common Scale) Scaling

Field test forms for all grade spans included horizontal linking items common to pairs of field test forms in all four language modalities, which allowed ETS to place the Summative ELPAC item pool per grade level and grade span on the same scale using concurrent calibration.

In addition, vertical linking items in two of the four field test forms were included to allow linkage of assessments of adjacent grade level and grade spans for grades one through twelve. The vertical linking items used were only from a grade level or grade span below the set grade level or grade span so as to accommodate the “opportunity to learn” issue—where it is not anticipated that students in a lower grade level or grade span will be able to perform skills necessary for the adjacent, higher grade level or grade span because they have not yet been exposed to those standards at school. This is especially true for the literacy skills. Therefore, the K field test forms did not include any vertical linking items, only horizontal linking items.

Figure 11.1 shows the horizontal linking design for a select (i.e., non-K) grade level or grade span for the Summative ELPAC, where form 1 and form 2 share a block of common items, form 2 and form 3 share a different block of common items, and form 4 and form 1 share a block of common items. These common items within a grade level or grade span allowed ETS to calibrate the item pool for each grade span concurrently so that the item parameter estimates for each grade level or grade span are on the same scale.

Figure 11.1 also illustrates how vertical linking is implemented where form 3 and form 4 each include a block of items (i.e., vertical linking items) from a grade level or grade span below the grade of interest. ETS used these vertical linking items to place adjacent grade levels and grade spans on the same scale.

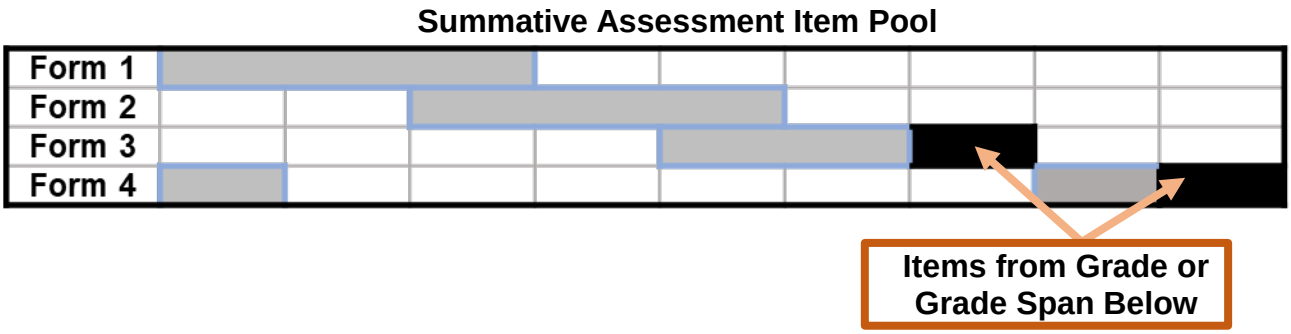

Figure 11.1 Horizontal Linking Design for Summative Assessment, One Selected Grade Level

The ETS team had two basic approaches to vertical scaling: separate and concurrent calibration. First, ETS concurrently calibrated items in all field test forms, including both the vertical and horizontal linking items for a given grade level or grade span, as shown in figure 11.2.

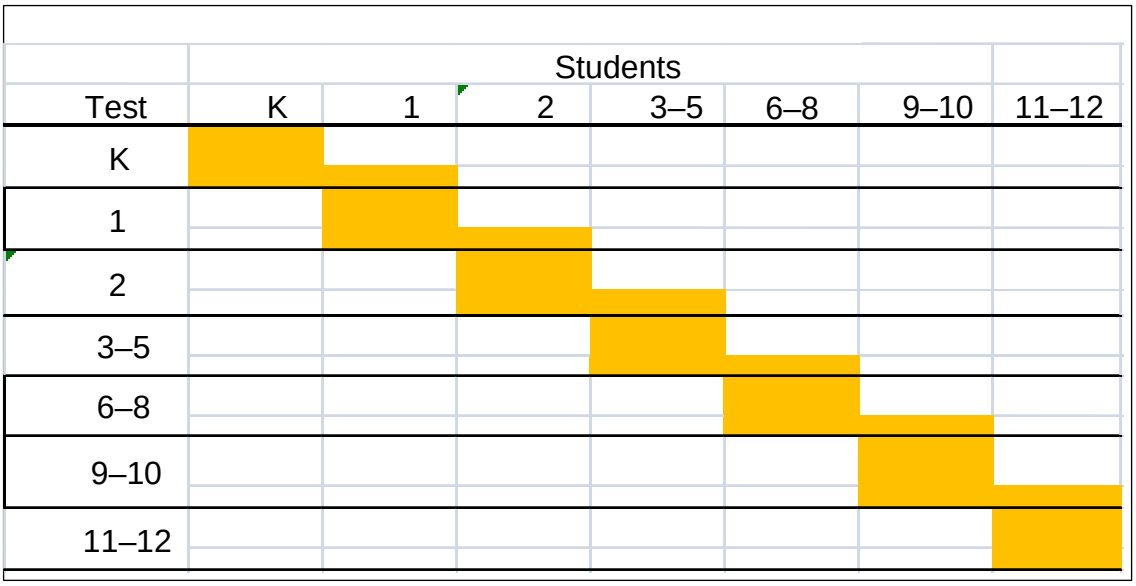

Figure 11.2 Vertical Linking Design Across All Grade Levels and Grade Spans

ETS then separately calibrated items administered to students at each grade level or grade span and effectively conducted seven separate concurrent calibrations. These concurrent calibrations allowed ETS to place items in each grade level or grade span on their own scale through the horizontal linking items.

ETS linked the adjacent grade level or grade span scales through vertical linking items using the Stocking and Lord (1983) test characteristic curve (TCC) linking procedure. Using grade span three through five as the baseline, a linking chain was used to link parameter estimates for higher and lower grade levels or grade spans onto the common vertical scale.

For example, grade two was linked to grade span three through five using the vertical linking items that were common between the grade levels and grade spans, grade one was linked to grade two using the vertical linking items that were common between the two grades, and K was linked to grade one using the vertical linking items that were common between the two grades. A similar process was followed to link grade span six through eight to grade span three through five, to link grade span nine and ten to grade span six through eight, and to link grade span eleven and twelve to grade span nine and ten. This chained linking approach optimizes the comparability of test scores within a grade span, as well as of test scores at adjacent grades.

Once item parameter estimates were obtained after placing all grade levels and grade spans on the same scale, ETS calculated proficiency scores for all students who participated in the Summative ELPAC stand-alone field test. These test summary scores were then used to evaluate the common scale constructed by the field test data.

Plots of TCCs for adjacent grade levels and grade spans for written scales are shown in figure 11.3. The curves in figure 11.3 are derived from the data in table 11.10.

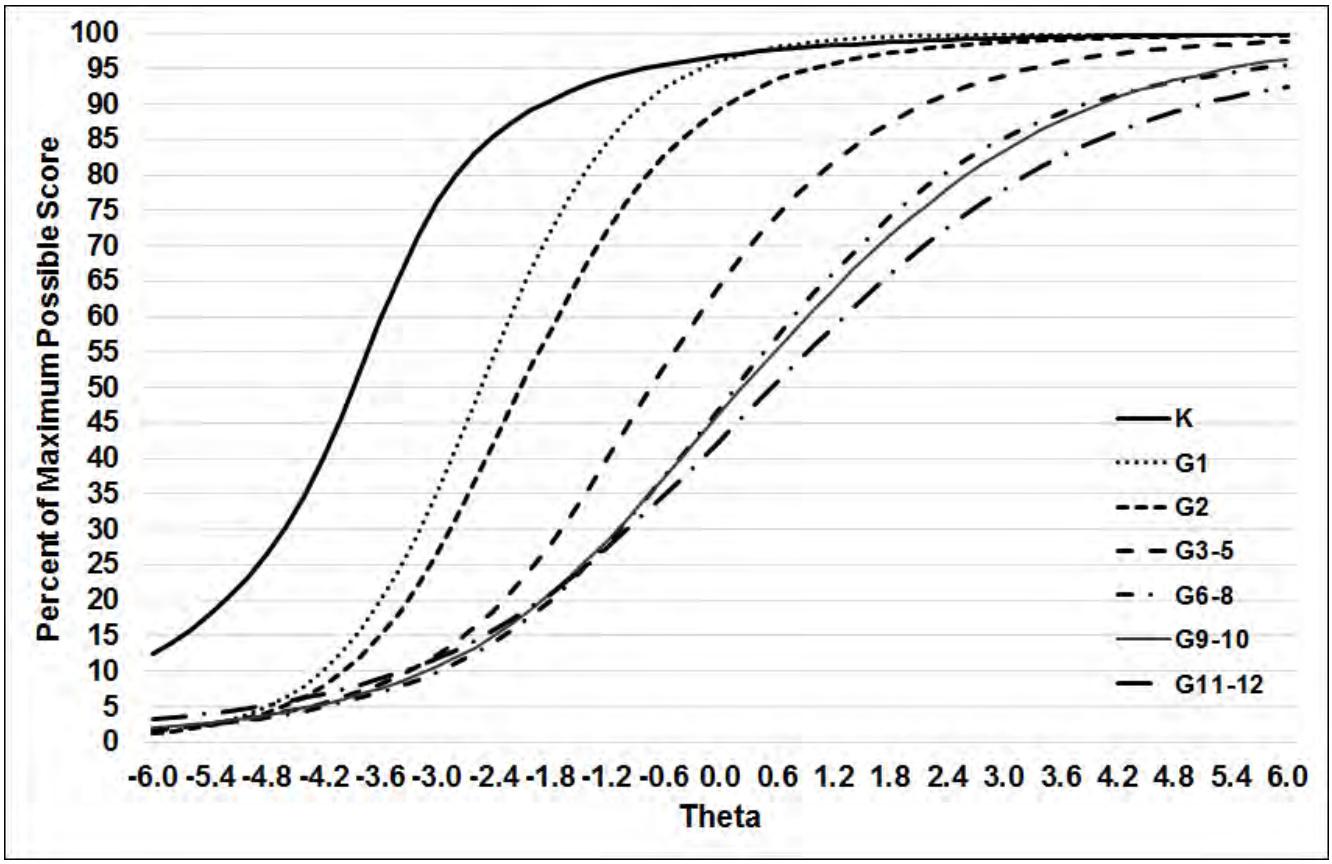

Figure 11.3 Written Language Composite Test Characteristic Curves Across Grade Levels

Plots of TCCs for adjacent grade levels and grade spans for oral scales are shown in figure 11.4. The curves in figure 11.4 are derived from the data in table 11.11.

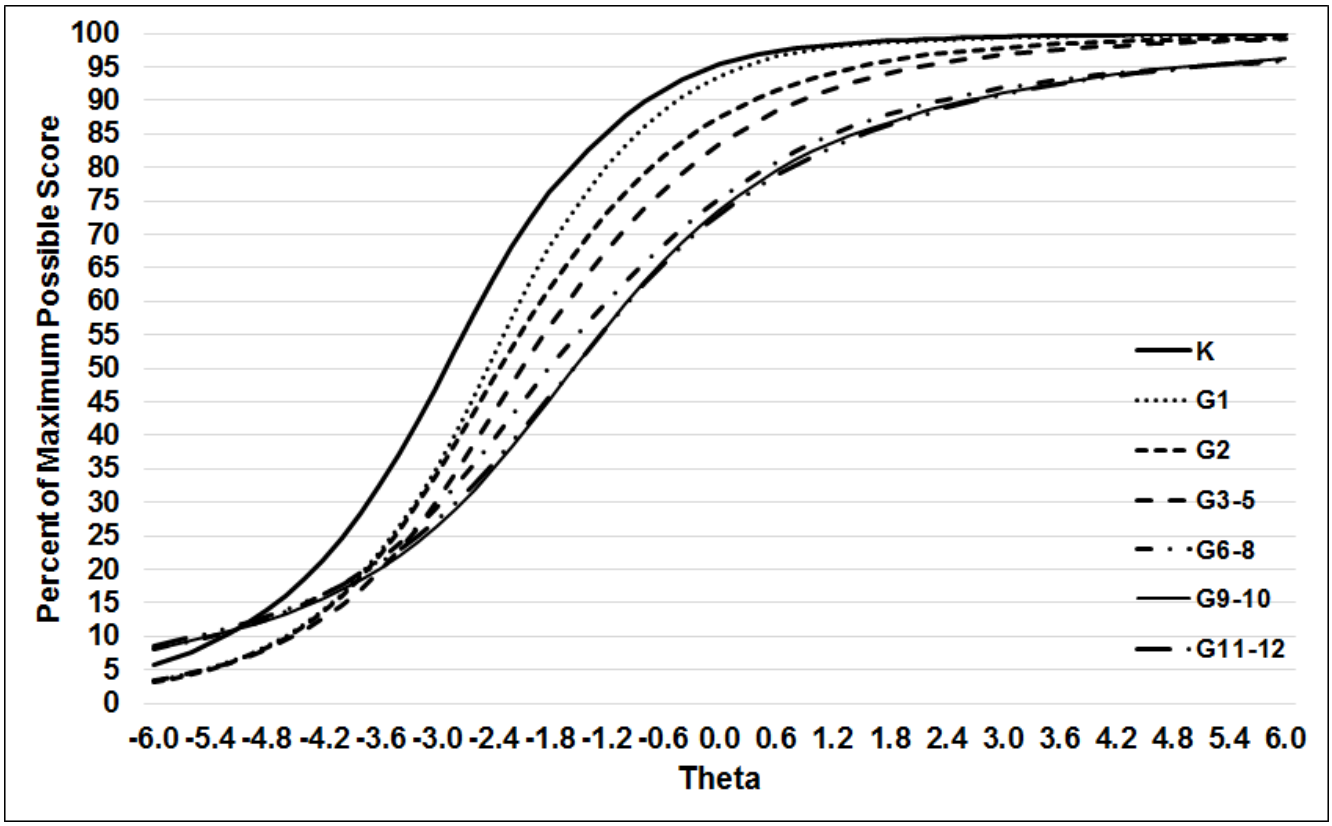

Figure 11.4 Oral Test Characteristic Curves Across Grade Levels

11.5.6. Developing Summative ELPAC Reporting Scales

This subsection summarizes procedures used for establishing reporting score scales and the highest obtainable scale score (HOSS) and lowest obtainable scale score (LOSS) for Summative ELPAC Oral and Written measures. The horizontal and vertical scales on the theta metric were established for Oral and Written measures during IRT calibration and linking. For reporting purposes, the Oral and Written student scores need to be transformed to the reporting score scale, and HOSS and LOSS need to be established.

11.5.6.1 Establishing Reporting Score Scale

Considerations in making the transformation included the following:

1. For both oral and written measures, the reporting score should be a four-digit number between 1000 and 1999. Because the thousandth place (i.e., "1") is a "prefix," the effective range was from 000 to 999.
2. The overall score was a weighted average of the oral and written reporting scores. At each grade level, the oral and written reporting scores should have had comparable distributions with the same mean and standard deviation (SD). Equal means helped simplify the interpretation of the expected overall score. Equal variances ensured that the contributions these two measures made to the variance of the overall score reflected the weights applied to calculate the overall score. Having the same mean and variance for oral and written reporting scores at each grade level, however, could not be achieved because of the vertical nature of the oral and written reporting

scores. In particular, the difference in variance composition for these two measures made it difficult, if not unrealistic, to achieve within-grade score distribution comparability while maintaining separate vertical scales for the two measures. A less ideal but more implementable approach would be to scale oral and written scores so that the reporting scores with all included grade levels had the same mean and SD across the two measures based on the same reference population.

Taking into account these considerations, it was recommended that the reporting score scales for both oral and written scores be set up in such a way that student scores, when placed on the reporting score metric, had a mean of 500 and an SD of 70 with students from all grade levels included. The mean and SD were determined as such that the reporting scores across all grade levels fell within the range of 000 to 999. The field test sample was used as the reference population. Table 11.8 presents the slope and intercept used to linearly transform vertically scaled theta scores to reporting scores for the oral and written measures respectively.

Table 11.8 Parameters for Transforming Vertically Linked Thetas to Reporting Scores

Variable	Oral	Written
Slope	41.602381083	38.871116151
Intercept	508.71944305	512.11884788

11.5.6.2 Determining HOSS and LOSS Values

Once the reporting score scales were established and the raw-to-scale score conversion tables were generated, the LOSS and HOSS values were determined by balancing the following criteria:

1. The HOSS must be high enough that it does not cause an unnecessary accumulation of scale scores at the top of the scale. Likewise, the LOSS should be low enough that it does not cause an unnecessary accumulation of scale scores at the bottom of the scale.
2. The HOSS should be greater than the scale score of the penultimate raw score (i.e., a raw score of N−1, where N represents the total raw score). The LOSS should be lower than the scale score that corresponds to a raw score of one. If doing so causes violation of other criteria, the LOSS may be adjusted upward accordingly, but it should not be set higher than the scale score that corresponds to the raw score at the guessing level.
3. The HOSS gap—which refers to the difference between the HOSS and the second-highest scale score—should be similar to the penultimate HOSS gap, which refers to the difference between the second- and the third-highest scale scores. Likewise, the LOSS gap—the difference between the LOSS and the second smallest scale score)—should be similar to the penultimate LOSS gap, which is the difference between the second- and the third-smallest scale scores.
4. The HOSS should be low enough that *CSEM*(HOSS) < 10*MIN(*CSEM*s for all scale scores), where *CSEM* is the conditional standard error of measurement. The LOSS should be high enough that *CSEM*(LOSS)<15*MIN(*CSEM*s).
5. For multiple test levels placed on the same vertical scale, the HOSS and LOSS values should increase and transition smoothly over levels.

6. For multiple test levels placed on the same vertical scale, the *CSEM*(HOSS) should be similar across levels. Likewise, *CSEM*(LOSS) should be similar across levels.
7. The HOSS and LOSS values should be consistent across oral and written measures.

Table 11.9 presents the resulting Summative ELPAC HOSS and LOSS values after balancing different criteria and considerations. It also summarizes the mean and SD of the oral and written scale scores. The distributions of oral and written scale scores after applying HOSS and LOSS values are examined.

Table 11.9 LOSS, HOSS, Mean, and Standard Deviation of Oral and Written Reporting Scale Scores Based on Field Test Data

Grade Level or Grade Span	LOSS	HOSS	Oral Mean	Oral SD	Written Mean	Written SD
K	150	700	436	45	409	70
Grade 1	150	700	468	47	462	61
Grade 2	150	700	487	49	485	54
Grade span 3–5	150	800	515	59	513	43
Grade span 6–8	150	900	528	68	535	45
Grade span 9–10	150	950	534	89	546	55
Grade span 11–12	150	950	539	85	558	50

Figure 11.5 shows the key percentiles of the score distributions, together with the HOSS and LOSS values, for oral scale scores. The curves in figure 11.5 are derived from the data in table 11.12.

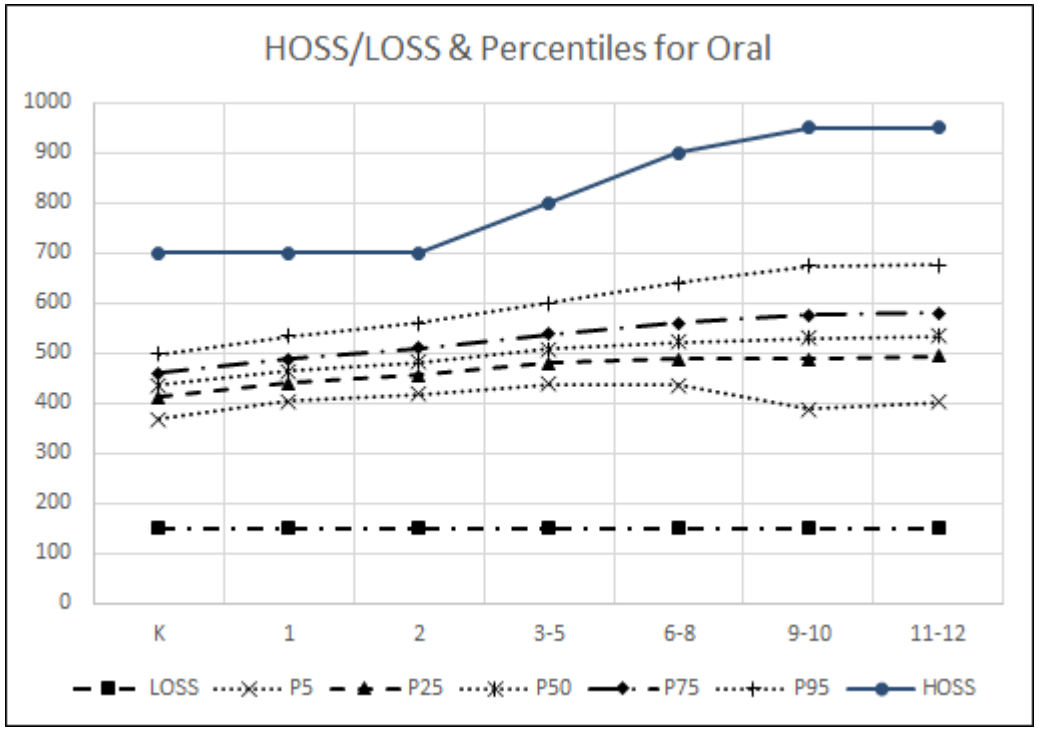

Figure 11.5 HOSS, LOSS, and Percentile for Oral Scale Scores

Figure 11.6 shows the key percentiles of the score distributions, together with the HOSS and LOSS values, for written scale scores. The curves in figure 11.6 are derived from the data in table 11.13.

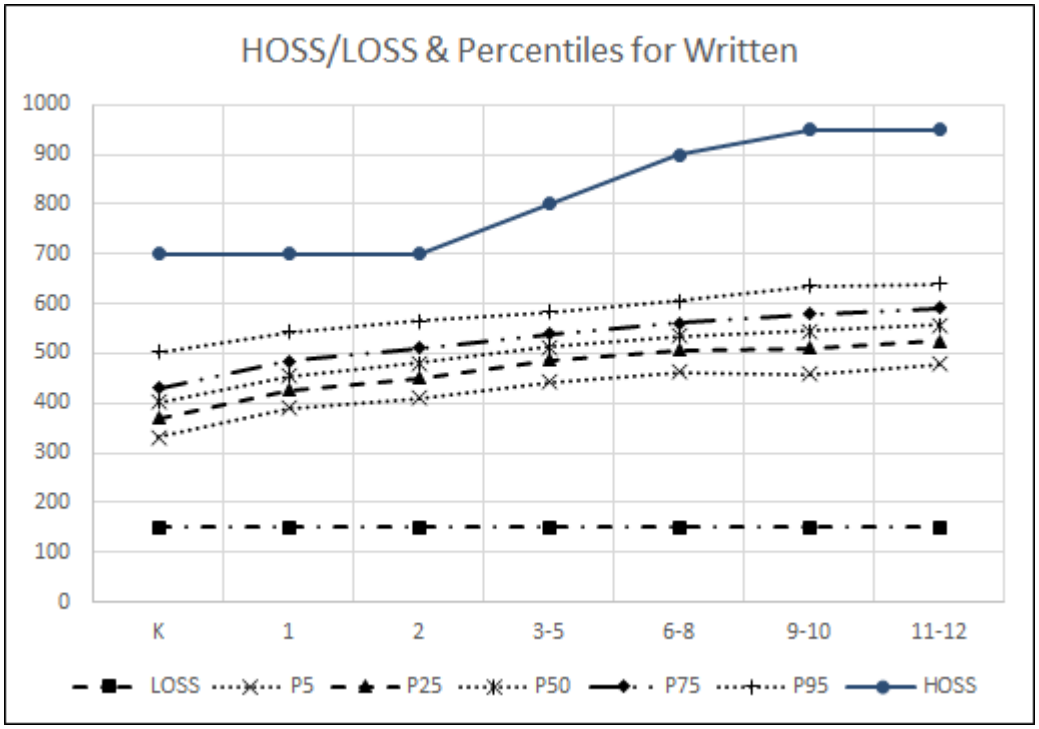

Figure 11.6 HOSS, LOSS, and Percentile for Written Scale Scores

Figure 11.7 shows the cumulative distributions of oral reporting scores at each grade level or grade span. The curves in figure 11.7 are derived from the data in table 11.14.

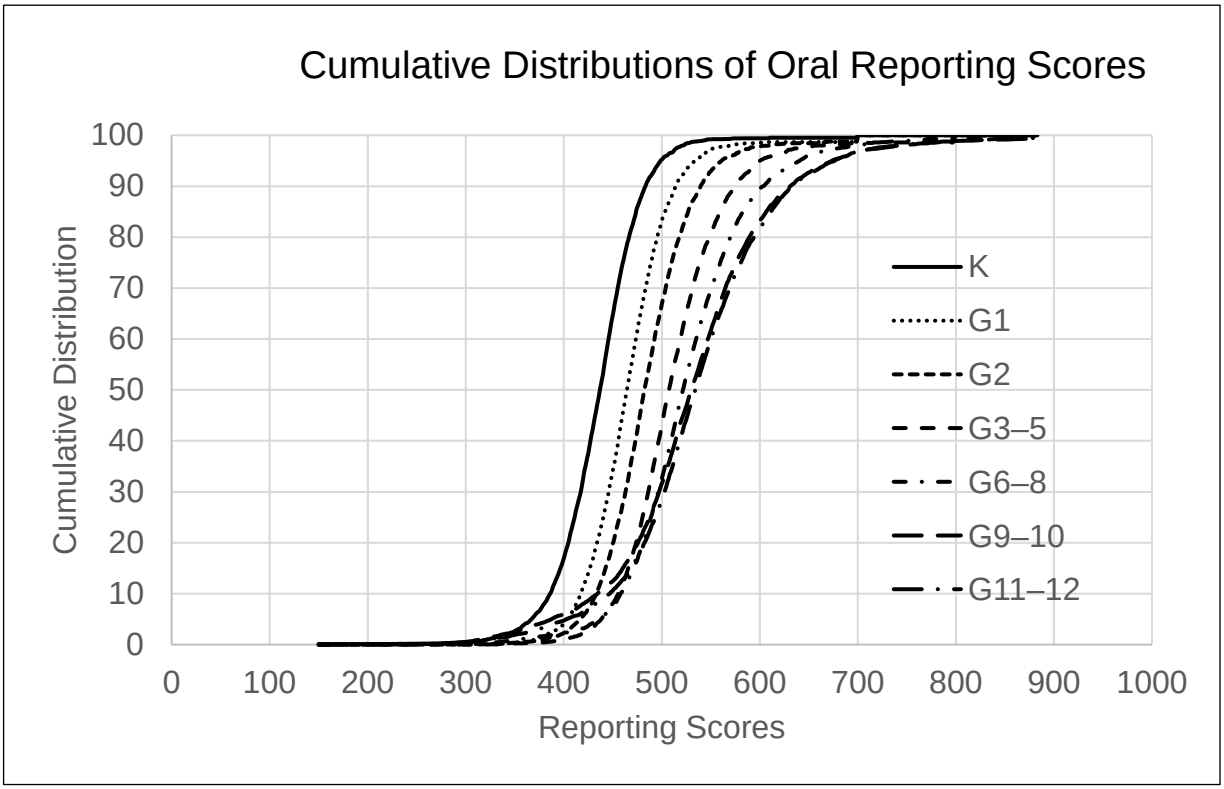

Figure 11.7 Cumulative Distributions of Oral Reporting Scores at Each Grade Level and Grade Span

Figure 11.8 shows the cumulative distributions of written reporting scores at each grade level or grade span. The curves in figure 11.8 are derived from the data in table 11.15.

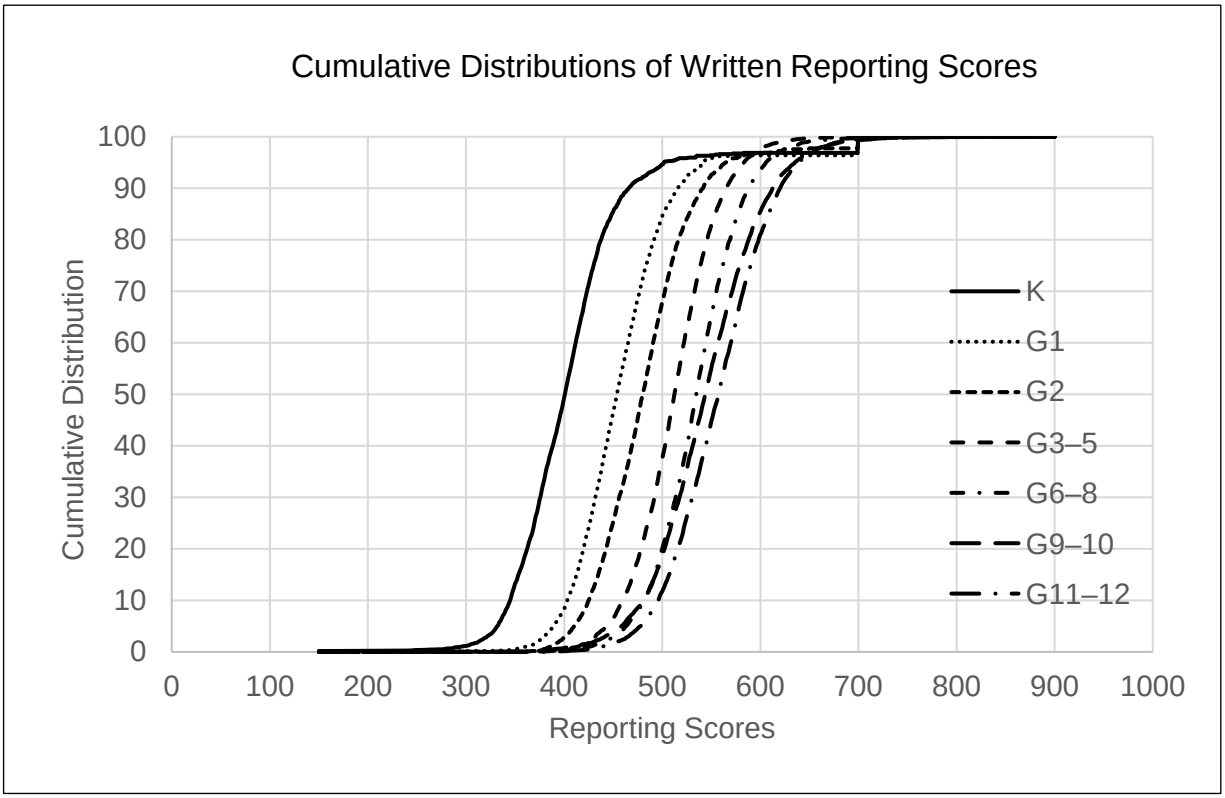

Figure 11.8 Cumulative Distributions of Written Reporting Scores at Each Grade Level and Grade Span

References

California Department of Education. (2018a). *Matrix Four: Universal tools, designated supports, and accommodations for the English Language Proficiency Assessments for California*. Retrieved from https://www.cde.ca.gov/ta/tg/ep/documents/elpacmatrix4.docx

California Department of Education. (2015). *Proposed test blueprints for the English Language Proficiency Assessments for California*. Sacramento, CA: California Department of Education. Retrieved from http://www.cde.ca.gov/ta/tg/ep/documents/elpacblueprts.pdf

Educational Testing Service. (2019). *An investigation of the factor structures for the ELAPC* (Unpublished report). Princeton, NJ: Educational Testing Service.

Stocking, M. L., & Lord, F. M. (1983). *Developing a common metric in item response theory*. *Applied Psychological Measurement, 7*, 201–10.

Accessibility References: Table Data for Field Test Graphs

Field Test Data for Test Characteristic Curves

Table 11.10 Written Language Composite Test Characteristic Curves Data for Figure 11.3

Theta	K	G1	G2	G3–5	G6–8	G9–10	G11–12
-6.0	12.44	1.24	1.31	1.79	1.75	1.99	3.21
-5.8	14.01	1.56	1.57	2.01	1.96	2.22	3.46
-5.6	15.82	1.96	1.89	2.27	2.20	2.47	3.74
-5.4	17.92	2.46	2.29	2.56	2.47	2.76	4.06
-5.2	20.36	3.11	2.78	2.90	2.78	3.09	4.40
-5.0	23.19	3.92	3.40	3.28	3.12	3.45	4.79
-4.8	26.46	4.94	4.17	3.72	3.51	3.86	5.22
-4.6	30.25	6.24	5.13	4.23	3.94	4.32	5.69
-4.4	34.68	7.89	6.34	4.80	4.42	4.84	6.22
-4.2	39.84	9.97	7.85	5.47	4.96	5.41	6.80
-4.0	45.79	12.57	9.72	6.23	5.56	6.06	7.44
-3.8	52.37	15.80	12.03	7.11	6.24	6.78	8.15
-3.6	59.17	19.69	14.83	8.12	6.99	7.59	8.94
-3.4	65.66	24.24	18.17	9.29	7.84	8.49	9.81
-3.2	71.39	29.41	22.06	10.64	8.79	9.50	10.77
-3.0	76.13	35.10	26.47	12.21	9.86	10.63	11.83
-2.8	79.92	41.22	31.33	14.01	11.08	11.89	13.01
-2.6	82.93	47.63	36.53	16.10	12.46	13.31	14.30
-2.4	85.38	54.12	41.92	18.49	14.02	14.90	15.74
-2.2	87.41	60.46	47.37	21.23	15.79	16.67	17.31
-2.0	89.12	66.44	52.78	24.33	17.78	18.64	19.02
-1.8	90.56	71.91	58.04	27.78	20.00	20.82	20.89
-1.6	91.78	76.77	63.05	31.56	22.45	23.20	22.91
-1.4	92.80	81.00	67.74	35.60	25.10	25.77	25.06
-1.2	93.67	84.60	72.06	39.84	27.94	28.49	27.33
-1.0	94.41	87.62	75.97	44.17	30.94	31.34	29.71
-0.8	95.04	90.11	79.45	48.50	34.06	34.28	32.19
-0.6	95.59	92.13	82.51	52.77	37.28	37.28	34.73
-0.4	96.06	93.76	85.16	56.89	40.57	40.32	37.33
-0.2	96.47	95.06	87.42	60.82	43.90	43.37	39.98
0.0	96.84	96.09	89.34	64.54	47.25	46.42	42.65
0.2	97.16	96.90	90.95	68.03	50.59	49.46	45.34
0.4	97.44	97.55	92.30	71.27	53.90	52.46	48.04
0.6	97.70	98.05	93.43	74.27	57.14	55.41	50.73
0.8	97.93	98.45	94.37	77.03	60.29	58.31	53.40
1.0	98.13	98.77	95.16	79.56	63.34	61.13	56.04

Table 11.10 *(continued)*

Theta	K	G1	G2	G3–5	G6–8	G9–10	G11–12
1.2	98.31	99.02	95.82	81.85	66.26	63.87	58.63
1.4	98.48	99.22	96.38	83.92	69.03	66.51	61.16
1.6	98.63	99.37	96.86	85.78	71.65	69.05	63.62
1.8	98.76	99.50	97.26	87.44	74.10	71.48	65.99
2.0	98.88	99.60	97.60	88.92	76.38	73.78	68.27
2.2	98.99	99.67	97.90	90.23	78.49	75.96	70.45
2.4	99.09	99.74	98.16	91.39	80.42	78.01	72.51
2.6	99.18	99.79	98.38	92.40	82.18	79.93	74.47
2.8	99.25	99.83	98.57	93.29	83.79	81.72	76.31
3.0	99.33	99.86	98.74	94.07	85.23	83.39	78.04
3.2	99.39	99.89	98.89	94.75	86.54	84.93	79.66
3.4	99.45	99.91	99.01	95.35	87.71	86.34	81.16
3.6	99.50	99.93	99.13	95.87	88.76	87.65	82.56
3.8	99.55	99.94	99.23	96.33	89.70	88.84	83.86
4.0	99.59	99.95	99.31	96.73	90.54	89.92	85.05
4.2	99.63	99.96	99.39	97.08	91.30	90.90	86.15
4.4	99.67	99.97	99.46	97.39	91.98	91.79	87.15
4.6	99.70	99.97	99.52	97.66	92.59	92.60	88.07
4.8	99.73	99.98	99.57	97.90	93.14	93.32	88.91
5.0	99.75	99.98	99.62	98.11	93.63	93.97	89.67
5.2	99.78	99.99	99.66	98.29	94.08	94.56	90.36
5.4	99.80	99.99	99.70	98.45	94.49	95.08	90.98
5.6	99.82	99.99	99.73	98.60	94.86	95.55	91.55
5.8	99.83	99.99	99.76	98.73	95.19	95.98	92.07
6.0	99.85	99.99	99.79	98.84	95.50	96.35	92.54

Table 11.11 Oral Language Composite Test Characteristic Curves Data for Figure 11.4

Theta	K	G1	G2	G3–5	G6–8	G9–10	G11–12
-6.0	5.76	3.37	3.18	3.45	8.15	8.21	8.69
-5.8	6.68	3.90	3.71	3.97	8.71	8.77	9.26
-5.6	7.74	4.53	4.35	4.58	9.34	9.38	9.88
-5.4	8.97	5.28	5.11	5.28	10.03	10.05	10.55
-5.2	10.39	6.17	6.01	6.09	10.79	10.77	11.29
-5.0	12.02	7.23	7.08	7.04	11.63	11.57	12.11
-4.8	13.92	8.49	8.35	8.15	12.57	12.44	13.00
-4.6	16.10	10.00	9.85	9.43	13.63	13.41	14.00
-4.4	18.62	11.79	11.62	10.93	14.82	14.49	15.10
-4.2	21.50	13.91	13.70	12.66	16.17	15.69	16.33
-4.0	24.79	16.38	16.12	14.68	17.71	17.03	17.71
-3.8	28.49	19.24	18.92	17.01	19.47	18.53	19.24
-3.6	32.60	22.51	22.12	19.68	21.48	20.21	20.95
-3.4	37.11	26.21	25.73	22.74	23.78	22.09	22.85
-3.2	41.99	30.35	29.72	26.20	26.36	24.18	24.97
-3.0	47.19	34.96	34.07	30.05	29.25	26.52	27.31
-2.8	52.62	40.06	38.68	34.22	32.39	29.10	29.89
-2.6	58.07	45.61	43.44	38.62	35.76	31.93	32.70
-2.4	63.29	51.48	48.24	43.13	39.29	35.00	35.73
-2.2	68.08	57.36	52.96	47.63	42.90	38.30	38.94
-2.0	72.38	62.94	57.52	52.03	46.52	41.77	42.31
-1.8	76.21	68.04	61.86	56.26	50.09	45.37	45.78
-1.6	79.59	72.60	65.93	60.29	53.56	49.02	49.29
-1.4	82.60	76.64	69.71	64.08	56.89	52.66	52.79
-1.2	85.28	80.20	73.17	67.62	60.05	56.22	56.21
-1.0	87.65	83.32	76.31	70.89	63.04	59.64	59.52
-0.8	89.75	86.05	79.13	73.90	65.84	62.89	62.66
-0.6	91.57	88.42	81.63	76.66	68.47	65.92	65.61
-0.4	93.10	90.47	83.85	79.16	70.93	68.73	68.35
-0.2	94.36	92.22	85.80	81.43	73.23	71.30	70.86
0.0	95.36	93.67	87.52	83.47	75.36	73.65	73.15
0.2	96.16	94.85	89.02	85.29	77.33	75.78	75.23
0.4	96.79	95.79	90.34	86.93	79.15	77.70	77.12
0.6	97.29	96.53	91.49	88.38	80.80	79.43	78.83
0.8	97.69	97.11	92.50	89.67	82.31	80.99	80.38
1.0	98.02	97.56	93.38	90.80	83.68	82.39	81.79
1.2	98.29	97.93	94.16	91.81	84.92	83.67	83.08
1.4	98.52	98.22	94.84	92.69	86.04	84.82	84.26
1.6	98.71	98.46	95.43	93.46	87.05	85.86	85.34
1.8	98.88	98.66	95.95	94.14	87.96	86.82	86.33
2.0	99.02	98.82	96.40	94.74	88.78	87.69	87.24

Table 11.11 *(continued)*

Theta	K	G1	G2	G3–5	G6–8	G9–10	G11–12
2.2	99.14	98.96	96.79	95.27	89.52	88.49	88.08
2.4	99.25	99.08	97.13	95.73	90.19	89.23	88.86
2.6	99.34	99.19	97.42	96.14	90.80	89.91	89.57
2.8	99.42	99.28	97.67	96.50	91.35	90.54	90.23
3.0	99.49	99.35	97.89	96.82	91.85	91.13	90.84
3.2	99.56	99.42	98.09	97.11	92.30	91.68	91.40
3.4	99.61	99.48	98.26	97.36	92.72	92.18	91.92
3.6	99.65	99.54	98.41	97.59	93.09	92.65	92.40
3.8	99.70	99.58	98.54	97.79	93.43	93.09	92.84
4.0	99.73	99.62	98.66	97.98	93.75	93.50	93.26
4.2	99.76	99.66	98.76	98.14	94.03	93.88	93.64
4.4	99.79	99.69	98.86	98.29	94.30	94.23	93.99
4.6	99.81	99.72	98.94	98.43	94.54	94.56	94.33
4.8	99.83	99.75	99.02	98.55	94.77	94.87	94.64
5.0	99.85	99.77	99.09	98.66	94.98	95.15	94.93
5.2	99.87	99.79	99.15	98.77	95.18	95.41	95.20
5.4	99.88	99.81	99.21	98.86	95.36	95.66	95.45
5.6	99.90	99.83	99.26	98.95	95.54	95.89	95.69
5.8	99.91	99.85	99.31	99.02	95.70	96.10	95.91
6.0	99.92	99.86	99.36	99.10	95.85	96.29	96.12

Field Test Data for HOSS, LOSS, and Percentiles

Table 11.12 HOSS, LOSS, and Percentile for Oral Scale Scores Data for Figure 11.5

Grade Level or Grade Span	LOSS	P5	P25	P50	P75	P95	HOSS
Kindergarten	150	368	412	437	461	499	700
Grade 1	150	404	441	465	489.5	534	700
Grade 2	150	418	457	483	510	561	700
Grade span 3–5	150	439	481	508	539	600.5	800
Grade span 6–8	150	436	489	522	560	641	900
Grade span 9–10	150	388	488	530	577	674	950
Grade span 11–12	150	403	494	535	581	677	950

Table 11.13 HOSS, LOSS, and Percentile for Written Scale Scores Data for Figure 11.6

Grade Level or Grade Span	LOSS	P5	P25	P50	P75	P95	HOSS
Kindergarten	150	332	370	401	430	502	700
Grade 1	150	389	427	454	484	542	700
Grade 2	150	411	450	481	510	565	700
Grade span 3–5	150	443	486	513	539	584	800
Grade span 6–8	150	463	507	535	562	606	900
Grade span 9–10	150	459	510	544	580	636	950
Grade span 11–12	150	478	524	558	591	639	950

Field Test Data for Cumulative Distributions

Table 11.14 Cumulative Distributions of Oral Reporting Scores at Each Grade Level and Grade Span Data for Figure 11.7

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
150	0.07	0.05	0.02	0.00	0.00	0.02	0.00
181	0.07	0.05	0.02	0.00	0.00	0.03	0.00
185	0.07	0.05	0.02	0.00	0.00	0.05	0.00
186	0.07	0.05	0.02	0.00	0.00	0.05	0.02
195	0.07	0.05	0.02	0.00	0.00	0.07	0.02
204	0.09	0.05	0.02	0.00	0.00	0.07	0.02
221	0.09	0.05	0.02	0.00	0.02	0.07	0.02
223	0.11	0.05	0.02	0.00	0.04	0.07	0.02
229	0.13	0.05	0.02	0.00	0.04	0.07	0.02
232	0.15	0.05	0.02	0.00	0.04	0.07	0.02
236	0.15	0.05	0.02	0.00	0.04	0.08	0.02
239	0.15	0.08	0.02	0.00	0.04	0.08	0.02
244	0.17	0.08	0.02	0.00	0.04	0.08	0.02
247	0.17	0.08	0.02	0.00	0.04	0.10	0.02
248	0.17	0.08	0.02	0.00	0.04	0.11	0.02
255	0.19	0.08	0.02	0.00	0.04	0.11	0.02
257	0.19	0.08	0.02	0.00	0.04	0.13	0.02
258	0.19	0.08	0.02	0.00	0.04	0.13	0.04
259	0.19	0.08	0.02	0.00	0.04	0.13	0.06
260	0.19	0.08	0.02	0.00	0.04	0.13	0.08
261	0.19	0.08	0.02	0.00	0.04	0.13	0.10
264	0.19	0.08	0.02	0.00	0.04	0.15	0.10
266	0.21	0.08	0.02	0.00	0.04	0.15	0.12
267	0.21	0.08	0.02	0.00	0.04	0.16	0.12
270	0.22	0.08	0.02	0.00	0.04	0.16	0.14
271	0.22	0.08	0.02	0.00	0.04	0.18	0.14
273	0.24	0.08	0.02	0.00	0.04	0.20	0.15
274	0.24	0.08	0.02	0.00	0.04	0.21	0.15
275	0.24	0.08	0.02	0.00	0.04	0.23	0.17
276	0.24	0.08	0.02	0.00	0.04	0.23	0.21
278	0.26	0.08	0.02	0.00	0.04	0.26	0.21
279	0.28	0.08	0.02	0.00	0.05	0.26	0.21
280	0.30	0.08	0.02	0.00	0.05	0.28	0.21
281	0.30	0.08	0.02	0.00	0.05	0.31	0.23
282	0.30	0.08	0.02	0.00	0.07	0.34	0.25
284	0.30	0.08	0.02	0.00	0.07	0.34	0.27
285	0.30	0.08	0.02	0.00	0.07	0.34	0.29

Table 11.14 *(continuation one)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
288	0.30	0.08	0.02	0.00	0.07	0.36	0.31
290	0.34	0.08	0.02	0.00	0.09	0.38	0.33
291	0.34	0.08	0.02	0.00	0.09	0.39	0.33
292	0.35	0.08	0.02	0.00	0.09	0.41	0.33
293	0.35	0.08	0.02	0.00	0.09	0.43	0.35
294	0.35	0.08	0.02	0.00	0.11	0.44	0.35
295	0.37	0.08	0.02	0.02	0.14	0.44	0.37
296	0.37	0.08	0.02	0.02	0.16	0.48	0.37
297	0.39	0.08	0.02	0.02	0.16	0.48	0.37
298	0.39	0.08	0.02	0.02	0.16	0.49	0.42
299	0.41	0.08	0.02	0.02	0.16	0.49	0.44
300	0.43	0.08	0.02	0.02	0.16	0.51	0.48
301	0.45	0.08	0.02	0.02	0.16	0.54	0.48
302	0.47	0.08	0.02	0.02	0.16	0.57	0.52
303	0.52	0.08	0.02	0.02	0.18	0.59	0.54
304	0.54	0.08	0.02	0.02	0.18	0.62	0.56
305	0.54	0.08	0.02	0.02	0.18	0.64	0.58
306	0.54	0.08	0.02	0.02	0.18	0.67	0.58
307	0.56	0.08	0.03	0.02	0.18	0.69	0.58
308	0.56	0.09	0.03	0.02	0.19	0.74	0.58
309	0.62	0.09	0.03	0.02	0.19	0.79	0.58
310	0.63	0.09	0.03	0.02	0.19	0.85	0.60
311	0.65	0.09	0.03	0.02	0.21	0.85	0.64
312	0.69	0.09	0.03	0.02	0.21	0.89	0.66
313	0.69	0.11	0.03	0.02	0.23	0.95	0.69
314	0.69	0.11	0.03	0.02	0.23	1.05	0.69
315	0.73	0.11	0.03	0.02	0.23	1.07	0.71
316	0.75	0.11	0.03	0.02	0.26	1.08	0.73
317	0.78	0.11	0.05	0.02	0.26	1.17	0.81
318	0.84	0.11	0.05	0.02	0.26	1.20	0.81
319	0.88	0.11	0.05	0.02	0.26	1.23	0.83
320	0.91	0.11	0.06	0.02	0.28	1.26	0.85
321	0.91	0.13	0.06	0.02	0.30	1.31	0.89
322	1.04	0.13	0.08	0.02	0.30	1.35	0.89
323	1.06	0.13	0.09	0.03	0.30	1.36	0.91
324	1.08	0.13	0.09	0.03	0.32	1.41	0.93
325	1.14	0.13	0.09	0.03	0.35	1.41	0.95
326	1.21	0.13	0.09	0.05	0.37	1.49	1.02
327	1.23	0.14	0.09	0.05	0.37	1.51	1.06

Table 11.14 *(continuation two)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
328	1.25	0.14	0.09	0.05	0.42	1.56	1.08
329	1.27	0.17	0.09	0.05	0.42	1.59	1.16
330	1.32	0.17	0.09	0.05	0.44	1.61	1.16
331	1.44	0.17	0.09	0.05	0.46	1.64	1.18
332	1.49	0.19	0.09	0.05	0.48	1.67	1.24
333	1.49	0.19	0.13	0.07	0.48	1.71	1.27
334	1.55	0.20	0.14	0.07	0.49	1.76	1.33
335	1.55	0.20	0.14	0.07	0.51	1.87	1.33
336	1.59	0.20	0.16	0.07	0.51	1.90	1.37
337	1.66	0.20	0.17	0.07	0.55	1.99	1.41
338	1.74	0.24	0.19	0.08	0.60	2.00	1.43
339	1.77	0.25	0.20	0.08	0.62	2.05	1.43
340	1.79	0.27	0.25	0.12	0.65	2.07	1.45
341	1.92	0.27	0.25	0.13	0.65	2.15	1.47
342	2.03	0.27	0.27	0.17	0.69	2.15	1.51
343	2.07	0.27	0.27	0.17	0.72	2.15	1.54
344	2.07	0.28	0.30	0.18	0.76	2.20	1.58
345	2.15	0.33	0.33	0.20	0.76	2.30	1.66
346	2.20	0.35	0.33	0.20	0.78	2.33	1.74
347	2.29	0.38	0.33	0.20	0.81	2.36	1.80
348	2.41	0.39	0.36	0.20	0.85	2.43	1.83
349	2.50	0.39	0.36	0.20	0.86	2.48	1.89
350	2.65	0.42	0.36	0.23	0.86	2.53	1.89
351	2.74	0.42	0.36	0.23	0.88	2.58	1.97
352	2.89	0.44	0.36	0.23	0.90	2.71	2.03
353	2.97	0.44	0.36	0.25	0.93	2.76	2.03
354	3.06	0.44	0.38	0.25	0.93	2.82	2.08
355	3.12	0.44	0.39	0.25	0.95	2.84	2.10
356	3.26	0.44	0.46	0.25	1.00	2.91	2.10
357	3.47	0.47	0.46	0.25	1.11	2.96	2.16
358	3.68	0.47	0.49	0.25	1.13	2.99	2.18
359	3.81	0.49	0.50	0.26	1.13	2.99	2.20
360	3.99	0.50	0.52	0.26	1.15	3.07	2.24
361	4.12	0.56	0.54	0.28	1.18	3.10	2.28
362	4.20	0.58	0.55	0.28	1.22	3.14	2.39
363	4.27	0.66	0.58	0.31	1.23	3.22	2.45
364	4.40	0.72	0.58	0.31	1.27	3.28	2.49
365	4.59	0.77	0.58	0.31	1.27	3.35	2.49
366	4.81	0.82	0.60	0.31	1.29	3.45	2.61

Table 11.14 *(continuation three)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
367	4.98	0.85	0.61	0.31	1.32	3.53	2.64
368	5.19	0.86	0.65	0.31	1.34	3.60	2.68
369	5.41	0.89	0.68	0.31	1.37	3.65	2.78
370	5.62	0.91	0.68	0.33	1.39	3.69	2.84
371	5.90	0.94	0.72	0.36	1.46	3.86	2.93
372	6.10	1.00	0.74	0.38	1.52	3.92	2.97
373	6.31	1.19	0.77	0.40	1.52	3.96	2.99
374	6.44	1.25	0.80	0.40	1.55	4.01	3.01
375	6.55	1.32	0.83	0.41	1.57	4.07	3.05
376	6.85	1.40	0.90	0.41	1.60	4.12	3.11
377	7.15	1.41	0.91	0.43	1.62	4.22	3.18
378	7.44	1.52	0.96	0.43	1.62	4.27	3.24
379	7.80	1.62	0.99	0.43	1.66	4.29	3.28
380	7.99	1.71	1.07	0.45	1.67	4.38	3.32
381	8.28	1.84	1.10	0.46	1.71	4.45	3.42
382	8.53	1.96	1.10	0.46	1.71	4.55	3.53
383	8.88	2.07	1.21	0.50	1.80	4.61	3.61
384	9.29	2.13	1.24	0.53	1.80	4.71	3.69
385	9.65	2.21	1.34	0.55	1.83	4.75	3.76
386	9.96	2.26	1.37	0.58	1.83	4.88	3.84
387	10.30	2.34	1.43	0.58	1.85	4.94	3.88
388	10.77	2.42	1.46	0.60	1.90	5.06	3.92
389	11.40	2.54	1.48	0.65	1.90	5.17	4.01
390	11.81	2.63	1.54	0.66	1.90	5.24	4.17
391	12.26	2.76	1.57	0.70	1.92	5.39	4.23
392	12.76	2.84	1.62	0.75	1.96	5.45	4.30
393	13.15	2.96	1.64	0.78	1.96	5.52	4.34
394	13.64	3.06	1.75	0.78	2.04	5.63	4.40
395	14.12	3.17	1.79	0.79	2.06	5.67	4.42
396	14.68	3.40	1.90	0.81	2.10	5.70	4.48
397	15.22	3.50	2.00	0.88	2.15	5.76	4.50
398	15.69	3.69	2.08	0.93	2.19	5.83	4.61
399	16.33	3.84	2.17	0.96	2.24	5.93	4.63
400	16.90	4.05	2.28	0.99	2.31	5.96	4.71
401	17.61	4.45	2.36	1.01	2.33	6.06	4.83
402	18.32	4.71	2.53	1.06	2.36	6.26	4.90
403	18.96	4.94	2.69	1.11	2.38	6.34	5.02
404	19.53	5.18	2.77	1.24	2.43	6.39	5.10
405	20.26	5.38	2.86	1.29	2.47	6.44	5.17

Table 11.14 *(continuation four)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
406	21.14	5.63	3.07	1.31	2.50	6.55	5.25
407	22.05	5.94	3.21	1.36	2.50	6.62	5.31
408	22.69	6.27	3.43	1.42	2.56	6.67	5.40
409	23.47	6.57	3.60	1.44	2.61	6.78	5.52
410	24.24	6.84	3.78	1.49	2.68	6.85	5.56
411	24.94	7.18	3.92	1.59	2.75	6.95	5.66
412	25.95	7.69	4.16	1.61	2.80	7.03	5.66
413	26.75	8.12	4.27	1.64	2.84	7.08	5.69
414	27.48	8.63	4.45	1.71	2.91	7.24	5.75
415	28.25	9.05	4.66	1.82	2.96	7.34	5.81
416	28.97	9.33	4.86	1.85	3.07	7.47	5.91
417	29.72	9.82	4.99	1.94	3.12	7.64	6.08
418	30.65	10.34	5.21	2.09	3.12	7.80	6.18
419	31.77	10.92	5.48	2.15	3.17	7.91	6.35
420	32.91	11.32	5.70	2.27	3.19	8.05	6.49
421	34.12	11.86	5.97	2.33	3.22	8.11	6.58
422	34.94	12.50	6.15	2.43	3.37	8.29	6.66
423	35.82	12.99	6.36	2.55	3.44	8.39	6.87
424	36.81	13.55	6.75	2.70	3.51	8.51	6.99
425	37.56	14.02	7.05	2.85	3.74	8.65	7.10
426	38.71	14.60	7.35	2.93	3.84	8.80	7.18
427	39.78	15.25	7.68	2.98	4.02	8.95	7.32
428	40.93	15.94	7.95	3.15	4.05	9.11	7.47
429	41.81	16.41	8.25	3.31	4.19	9.29	7.55
430	42.82	17.02	8.70	3.44	4.30	9.41	7.72
431	43.94	17.77	9.19	3.58	4.39	9.49	7.84
432	44.96	18.70	9.44	3.76	4.60	9.69	7.95
433	45.91	19.28	9.95	4.02	4.63	9.79	8.13
434	47.03	20.08	10.39	4.24	4.74	9.92	8.22
435	47.95	20.73	10.89	4.37	4.88	10.12	8.40
436	49.01	21.42	11.35	4.54	5.00	10.25	8.45
437	50.17	22.30	11.87	4.70	5.15	10.41	8.61
438	51.14	23.13	12.40	4.95	5.30	10.48	8.72
439	52.16	23.97	12.77	5.18	5.44	10.61	8.82
440	53.10	24.75	13.35	5.50	5.52	10.72	9.01
441	54.46	25.58	13.99	5.68	5.74	10.82	9.13
442	55.78	26.58	14.62	5.94	5.89	10.99	9.26
443	56.96	27.56	15.32	6.19	6.10	11.22	9.50
444	58.08	28.22	15.84	6.46	6.33	11.41	9.65

Table 11.14 *(continuation five)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
445	59.31	29.16	16.43	6.64	6.48	11.58	9.77
446	60.41	30.19	16.92	6.89	6.66	11.82	9.90
447	61.47	31.34	17.57	7.24	6.75	12.02	10.02
448	62.69	32.26	18.35	7.58	7.07	12.18	10.31
449	63.77	33.25	19.05	7.88	7.24	12.30	10.37
450	64.70	34.22	19.80	8.05	7.45	12.46	10.64
451	65.75	35.04	20.38	8.38	7.63	12.64	10.85
452	66.98	35.93	21.27	8.66	7.93	12.82	11.04
453	67.91	37.00	22.10	9.04	8.21	13.04	11.18
454	68.86	38.14	22.97	9.44	8.53	13.24	11.41
455	70.00	39.18	23.66	9.77	8.95	13.53	11.66
456	70.97	40.50	24.40	10.32	9.29	13.79	11.83
457	72.02	41.83	25.04	10.60	9.62	14.06	12.08
458	72.93	42.93	25.70	11.04	9.99	14.29	12.26
459	73.97	44.10	26.57	11.47	10.26	14.61	12.53
460	74.81	45.19	27.52	11.92	10.59	14.99	12.72
461	75.58	46.38	28.19	12.48	11.12	15.21	12.89
462	76.49	47.37	29.03	12.88	11.40	15.55	13.13
463	77.31	48.53	30.08	13.39	11.68	15.86	13.34
464	78.13	49.53	30.96	13.94	11.97	16.13	13.57
465	78.99	50.67	31.86	14.57	12.37	16.40	13.76
466	79.57	51.63	32.88	15.15	12.95	16.65	13.90
467	80.43	53.00	34.14	15.80	13.23	16.95	14.21
468	81.10	54.02	35.05	16.44	13.62	17.34	14.50
469	81.75	54.96	36.20	17.07	14.06	17.64	14.77
470	82.28	56.21	36.99	17.63	14.52	17.96	15.17
471	83.00	57.09	37.90	18.23	15.00	18.28	15.48
472	83.71	58.42	38.86	18.89	15.35	18.74	16.02
473	84.29	59.41	39.93	19.49	15.89	19.11	16.43
474	85.30	60.59	40.89	20.13	16.19	19.39	16.73
475	85.90	61.59	42.03	20.89	16.63	19.66	17.04
476	86.40	62.81	43.29	21.57	17.32	20.15	17.55
477	86.88	63.80	44.26	22.32	17.89	20.61	17.85
478	87.48	64.79	45.43	23.06	18.36	20.95	18.22
479	87.89	65.86	46.59	23.89	18.86	21.31	18.66
480	88.49	66.61	47.69	24.79	19.42	21.72	18.94
481	88.94	67.72	48.75	25.63	19.95	22.25	19.46
482	89.46	68.68	49.80	26.44	20.55	22.68	19.90
483	89.89	69.61	50.80	27.35	21.08	23.07	20.23

Table 11.14 *(continuation six)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
484	90.35	70.41	51.68	28.21	21.80	23.42	20.61
485	90.88	71.55	52.53	28.94	22.54	23.84	21.06
486	91.16	72.44	53.55	29.83	23.15	24.27	21.52
487	91.59	73.24	54.49	30.76	23.70	24.78	22.06
488	91.87	74.25	55.49	31.76	24.41	25.27	22.53
489	92.03	75.00	56.43	32.77	25.09	25.76	23.05
490	92.39	75.78	57.41	33.59	25.62	26.27	23.59
491	92.65	76.74	58.51	34.64	26.47	26.75	23.90
492	92.97	77.68	59.36	35.56	27.12	27.29	24.26
493	93.36	78.31	60.60	36.47	27.82	27.75	24.76
494	93.68	78.95	61.66	37.17	28.51	28.23	25.21
495	93.97	79.82	62.46	38.23	29.15	28.72	25.57
496	94.20	80.57	63.21	39.29	29.74	29.31	26.31
497	94.55	81.26	64.05	40.23	30.43	29.90	26.87
498	94.80	81.82	64.99	41.06	31.12	30.53	27.39
499	95.08	82.51	65.80	41.97	31.89	31.08	27.95
500	95.24	83.33	66.85	42.96	32.56	31.63	28.57
501	95.39	83.83	67.67	43.87	33.36	32.10	29.09
502	95.69	84.57	68.42	45.15	34.04	32.68	29.51
503	95.82	84.91	69.12	46.19	34.87	33.15	30.13
504	95.93	85.32	70.09	47.17	35.56	33.84	30.71
505	95.99	85.74	71.01	48.15	36.26	34.48	31.37
506	96.14	86.34	71.87	48.92	37.16	35.06	32.08
507	96.33	86.76	72.74	49.79	38.13	35.58	32.60
508	96.49	87.14	73.40	50.65	38.80	36.13	33.45
509	96.53	87.59	74.26	51.49	39.54	36.86	33.84
510	96.60	87.92	75.08	52.48	40.34	37.44	34.38
511	96.64	88.57	75.73	53.61	41.29	38.10	35.03
512	96.88	88.96	76.56	54.55	42.13	38.77	35.67
513	97.16	89.51	77.19	55.35	42.98	39.44	36.34
514	97.31	89.85	77.85	56.08	43.84	40.16	37.02
515	97.44	90.18	78.42	56.81	44.78	40.85	37.68
516	97.59	90.65	79.05	57.48	45.71	41.46	38.51
517	97.67	91.25	79.54	58.38	46.45	42.04	39.20
518	97.71	91.69	80.26	59.21	47.35	43.01	39.68
519	97.82	91.88	80.80	60.10	48.00	43.56	40.44
520	97.93	92.05	81.19	60.86	48.74	44.11	41.17
521	97.97	92.25	81.90	61.85	49.60	44.61	41.75
522	98.02	92.38	82.43	62.75	50.36	45.14	42.46

Table 11.14 *(continuation seven)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
523	98.10	92.64	82.97	63.59	51.12	45.68	43.22
524	98.32	93.08	83.58	64.49	52.07	46.26	43.68
525	98.40	93.26	84.10	65.23	52.88	46.78	44.24
526	98.45	93.57	84.46	65.93	53.71	47.44	44.90
527	98.45	93.70	85.14	66.72	54.40	48.10	45.57
528	98.49	93.81	85.69	67.37	54.96	48.80	46.32
529	98.56	93.96	86.12	68.16	55.86	49.38	46.81
530	98.66	94.20	86.50	68.81	56.58	50.03	47.50
531	98.68	94.32	87.03	69.65	57.15	50.59	48.08
532	98.73	94.62	87.22	70.43	57.90	51.03	48.47
533	98.75	94.92	87.60	71.24	58.63	51.71	49.24
534	98.75	95.03	87.83	72.04	59.35	52.20	49.93
535	98.77	95.15	88.20	72.67	60.00	52.78	50.53
536	98.81	95.22	88.70	73.31	60.63	53.38	51.40
537	98.81	95.28	89.12	73.96	61.27	53.97	52.00
538	98.84	95.34	89.49	74.50	62.19	54.65	52.52
539	98.86	95.81	89.75	75.27	62.80	55.32	53.25
540	98.88	96.27	90.18	75.70	63.42	55.96	53.87
541	98.90	96.38	90.48	76.23	64.26	56.47	54.41
542	98.96	96.42	90.75	76.77	64.78	56.93	55.14
543	98.97	96.49	91.19	77.38	65.32	57.49	55.63
544	99.05	96.55	91.45	78.01	65.99	58.29	56.38
545	99.05	96.60	91.67	78.49	66.68	58.82	56.90
546	99.16	96.88	91.82	78.84	67.40	59.34	57.54
547	99.18	96.93	91.99	79.31	67.93	60.02	58.16
548	99.20	96.96	92.56	79.95	68.49	60.51	58.85
549	99.20	97.15	92.78	80.41	69.15	61.12	59.51
550	99.20	97.21	92.95	80.86	69.60	61.64	60.14
551	99.22	97.30	93.20	81.42	70.22	62.33	60.53
552	99.24	97.40	93.42	81.72	70.77	62.91	61.09
553	99.24	97.44	93.58	82.09	71.37	63.35	61.57
554	99.24	97.44	93.77	82.60	71.89	63.92	62.05
555	99.24	97.55	94.03	83.11	72.56	64.47	62.65
556	99.24	97.57	94.16	83.48	73.00	65.22	63.27
557	99.24	97.60	94.32	83.91	73.53	65.81	63.79
558	99.24	97.66	94.60	84.32	74.13	66.44	64.41
559	99.24	97.71	94.66	84.55	74.61	67.09	64.93
560	99.24	97.73	94.85	85.03	75.28	67.62	65.32
561	99.24	97.73	95.01	85.38	75.79	68.01	65.95

Table 11.14 *(continuation eight)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
562	99.24	97.74	95.22	85.99	76.21	68.47	66.28
563	99.25	97.79	95.33	86.31	76.79	69.03	66.80
564	99.25	97.80	95.47	86.71	77.27	69.57	67.44
565	99.25	97.85	95.51	87.17	77.83	70.10	67.92
566	99.27	97.87	95.55	87.45	78.24	70.72	68.40
567	99.27	97.90	95.56	87.95	78.64	71.18	68.91
568	99.27	97.90	95.77	88.15	79.01	71.49	69.33
569	99.27	97.98	95.84	88.46	79.56	72.12	69.95
570	99.29	97.99	95.84	88.68	80.05	72.46	70.51
571	99.33	97.99	95.86	88.97	80.53	72.82	70.86
572	99.35	97.99	95.96	89.35	81.00	73.28	71.42
573	99.38	98.13	96.13	89.65	81.41	73.68	71.86
574	99.38	98.18	96.16	89.87	81.97	74.07	72.30
575	99.38	98.18	96.19	90.08	82.33	74.34	72.75
576	99.38	98.23	96.51	90.32	82.63	74.76	73.11
577	99.38	98.24	96.51	90.70	83.01	75.45	73.46
578	99.38	98.34	96.52	91.03	83.49	75.93	73.87
579	99.38	98.34	96.87	91.28	83.77	76.32	74.31
580	99.38	98.34	96.87	91.46	84.07	76.68	74.70
581	99.40	98.34	96.87	91.59	84.42	77.00	75.18
582	99.40	98.34	96.90	91.77	84.78	77.42	75.58
583	99.40	98.37	96.98	91.97	85.04	77.72	76.03
584	99.40	98.37	96.98	92.20	85.43	78.08	76.47
585	99.40	98.43	97.18	92.37	85.76	78.44	76.99
586	99.40	98.43	97.29	92.53	86.01	78.82	77.36
587	99.40	98.43	97.29	92.65	86.29	79.23	77.78
588	99.40	98.43	97.31	92.73	86.61	79.59	78.11
589	99.40	98.45	97.34	92.96	86.87	79.90	78.44
590	99.40	98.45	97.59	93.06	87.03	80.23	78.87
591	99.40	98.45	97.59	93.13	87.33	80.56	79.14
592	99.40	98.45	97.64	93.21	87.58	80.94	79.37
593	99.40	98.45	97.66	93.81	87.88	81.30	79.64
594	99.40	98.45	97.72	94.34	88.05	81.54	79.97
595	99.40	98.45	97.72	94.37	88.37	81.82	80.24
596	99.40	98.45	97.73	94.47	88.56	82.20	80.62
597	99.40	98.54	97.75	94.64	88.83	82.40	80.97
598	99.40	98.54	97.81	94.82	88.92	82.64	81.16
599	99.40	98.54	97.81	94.95	89.13	82.89	81.49
600	99.40	98.54	97.83	95.00	89.43	83.20	81.84

Table 11.14 *(continuation nine)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
601	99.40	98.54	97.84	95.12	89.69	83.48	82.09
602	99.40	98.54	97.91	95.22	89.89	83.73	82.34
603	99.40	98.54	97.92	95.32	89.96	83.96	82.61
604	99.40	98.67	97.92	95.41	90.06	84.25	82.80
605	99.40	98.67	97.94	95.53	90.22	84.52	83.19
606	99.40	98.67	97.94	95.56	90.49	84.81	83.58
607	99.40	98.67	97.94	95.61	90.63	85.14	83.83
608	99.40	98.67	97.95	95.73	90.86	85.39	84.19
609	99.40	98.67	97.95	95.75	91.10	85.70	84.48
610	99.52	98.67	97.97	95.83	91.37	85.96	84.75
611	99.52	98.67	98.17	95.84	91.61	86.26	84.98
612	99.52	98.67	98.17	95.89	91.79	86.37	85.22
613	99.52	98.67	98.19	95.93	91.98	86.50	85.54
614	99.52	98.67	98.19	95.94	92.09	86.68	85.79
615	99.52	98.67	98.24	96.08	92.18	86.95	86.01
616	99.52	98.67	98.24	96.16	92.35	87.19	86.30
617	99.52	98.67	98.25	96.16	92.37	87.47	86.55
618	99.52	98.67	98.25	96.26	92.51	87.77	86.80
619	99.52	98.67	98.25	96.33	92.58	87.95	87.09
620	99.52	98.67	98.25	96.36	92.72	88.03	87.34
621	99.52	98.67	98.25	96.46	92.81	88.24	87.65
622	99.52	98.67	98.32	96.62	92.97	88.46	87.84
623	99.52	98.67	98.32	96.62	93.08	88.60	87.94
624	99.52	98.67	98.32	96.62	93.22	88.82	88.25
625	99.52	98.67	98.32	96.66	93.36	88.88	88.42
626	99.52	98.67	98.32	96.72	93.52	89.13	88.69
627	99.52	98.67	98.43	96.76	93.74	89.34	88.79
628	99.52	98.67	98.43	96.97	93.92	89.46	89.02
629	99.52	98.67	98.43	97.14	94.03	89.62	89.27
630	99.52	98.67	98.43	97.15	94.22	89.79	89.46
631	99.52	98.67	98.43	97.19	94.33	90.03	89.66
632	99.52	98.67	98.43	97.19	94.40	90.13	89.75
633	99.52	98.67	98.43	97.25	94.41	90.35	89.96
634	99.52	98.67	98.44	97.30	94.45	90.44	90.27
635	99.52	98.67	98.44	97.40	94.49	90.59	90.41
636	99.52	98.67	98.44	97.42	94.57	90.82	90.60
637	99.52	98.67	98.44	97.42	94.63	91.03	90.76
638	99.52	98.67	98.44	97.43	94.73	91.23	91.03
639	99.52	98.67	98.44	97.43	94.86	91.36	91.12

Table 11.14 *(continuation 10)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
640	99.52	98.67	98.44	97.43	94.91	91.49	91.22
641	99.52	98.67	98.44	97.62	95.07	91.59	91.30
642	99.52	98.67	98.44	97.63	95.21	91.71	91.43
643	99.52	98.67	98.44	97.67	95.31	91.74	91.53
644	99.52	98.67	98.44	97.73	95.40	91.92	91.59
645	99.52	98.67	98.44	97.75	95.54	92.10	91.78
646	99.52	98.67	98.44	97.75	95.67	92.20	91.87
647	99.52	98.67	98.44	97.75	95.81	92.32	91.89
648	99.52	98.67	98.44	97.77	96.00	92.48	91.99
649	99.52	98.67	98.44	97.77	96.05	92.61	92.09
650	99.52	98.67	98.44	97.82	96.09	92.66	92.20
651	99.52	98.67	98.44	97.82	96.21	92.78	92.32
652	99.52	98.67	98.44	97.83	96.30	92.92	92.42
653	99.52	98.67	98.44	97.85	96.35	93.05	92.65
654	99.52	98.67	98.44	97.86	96.39	93.15	92.72
655	99.52	98.67	98.44	97.96	96.42	93.28	92.80
656	99.52	98.67	98.44	97.96	96.42	93.35	92.97
657	99.52	98.67	98.44	97.96	96.48	93.51	93.09
658	99.52	98.67	98.44	97.96	96.55	93.65	93.19
659	99.52	98.67	98.44	98.00	96.62	93.76	93.30
660	99.52	98.67	98.60	98.00	96.65	93.79	93.44
661	99.52	98.67	98.60	98.00	96.69	93.86	93.50
662	99.52	98.67	98.60	98.00	96.74	93.96	93.71
663	99.52	98.67	98.60	98.00	96.81	94.06	93.77
664	99.52	98.67	98.60	98.00	96.88	94.12	93.84
665	99.52	98.67	98.60	98.00	96.95	94.22	93.98
666	99.52	98.67	98.60	98.00	97.00	94.34	94.13
667	99.52	98.67	98.77	98.05	97.06	94.40	94.27
668	99.52	98.67	98.77	98.05	97.09	94.48	94.35
669	99.52	98.67	98.77	98.05	97.09	94.55	94.48
670	99.52	98.67	98.77	98.05	97.09	94.65	94.56
671	99.52	98.67	98.77	98.05	97.11	94.76	94.63
672	99.52	98.67	98.77	98.05	97.23	94.81	94.71
673	99.52	98.67	98.77	98.05	97.27	94.96	94.77
674	99.52	98.67	98.77	98.05	97.27	95.01	94.83
675	99.52	98.67	98.77	98.08	97.29	95.14	94.89
676	99.52	98.67	98.77	98.08	97.34	95.21	94.94
677	99.52	98.67	98.77	98.10	97.34	95.30	95.12
678	99.52	98.67	98.77	98.10	97.37	95.34	95.23

Table 11.14 *(continuation 11)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
679	99.52	98.67	98.77	98.10	97.43	95.42	95.29
680	99.52	98.67	98.77	98.10	97.46	95.42	95.39
681	99.52	98.67	98.77	98.21	97.46	95.50	95.46
682	99.52	98.67	98.77	98.21	97.50	95.53	95.52
683	99.52	98.67	98.77	98.21	97.55	95.65	95.58
684	99.52	98.67	98.77	98.21	97.57	95.70	95.66
685	99.52	98.67	98.77	98.21	97.59	95.80	95.75
686	99.52	98.67	98.77	98.21	97.60	95.81	95.87
687	99.52	98.67	98.77	98.28	97.60	95.86	95.93
688	99.52	98.67	98.77	98.28	97.66	95.88	95.97
689	99.52	98.67	98.77	98.28	97.67	95.94	96.08
690	99.52	98.67	98.77	98.34	97.74	96.06	96.10
691	99.52	98.67	98.77	98.34	97.76	96.19	96.20
692	99.52	98.67	98.77	98.34	97.78	96.29	96.31
693	99.52	98.67	98.77	98.41	97.80	96.44	96.43
694	99.52	98.67	98.77	98.41	97.82	96.49	96.47
695	99.52	98.67	98.77	98.41	97.85	96.50	96.55
696	99.52	98.67	98.77	98.41	97.89	96.55	96.60
697	99.52	98.67	98.77	98.41	97.94	96.60	96.74
698	99.52	98.67	98.77	98.41	97.94	96.70	96.76
699	99.52	98.67	98.77	98.41	97.96	96.72	96.76
700	100.00	100.00	100.00	98.41	97.99	96.73	96.84
701	100.00	100.00	100.00	98.43	98.03	96.78	96.91
702	100.00	100.00	100.00	98.43	98.04	96.85	96.97
703	100.00	100.00	100.00	98.43	98.12	96.91	96.99
704	100.00	100.00	100.00	98.43	98.19	96.96	97.05
705	100.00	100.00	100.00	98.43	98.20	97.01	97.11
706	100.00	100.00	100.00	98.43	98.20	97.06	97.16
707	100.00	100.00	100.00	98.43	98.22	97.09	97.16
708	100.00	100.00	100.00	98.49	98.24	97.13	97.18
709	100.00	100.00	100.00	98.49	98.29	97.16	97.18
710	100.00	100.00	100.00	98.49	98.29	97.19	97.18
711	100.00	100.00	100.00	98.49	98.29	97.26	97.24
712	100.00	100.00	100.00	98.49	98.34	97.31	97.24
713	100.00	100.00	100.00	98.49	98.40	97.31	97.26
714	100.00	100.00	100.00	98.49	98.41	97.32	97.28
715	100.00	100.00	100.00	98.49	98.41	97.37	97.30
716	100.00	100.00	100.00	98.59	98.47	97.42	97.32
717	100.00	100.00	100.00	98.59	98.50	97.47	97.34

Table 11.14 *(continuation 12)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
718	100.00	100.00	100.00	98.59	98.54	97.49	97.34
719	100.00	100.00	100.00	98.59	98.56	97.52	97.38
720	100.00	100.00	100.00	98.59	98.59	97.54	97.38
721	100.00	100.00	100.00	98.59	98.61	97.60	97.41
722	100.00	100.00	100.00	98.59	98.63	97.62	97.43
723	100.00	100.00	100.00	98.59	98.63	97.64	97.47
724	100.00	100.00	100.00	98.59	98.64	97.65	97.51
725	100.00	100.00	100.00	98.59	98.66	97.67	97.53
726	100.00	100.00	100.00	98.59	98.66	97.69	97.53
727	100.00	100.00	100.00	98.59	98.66	97.72	97.57
728	100.00	100.00	100.00	98.59	98.66	97.73	97.61
729	100.00	100.00	100.00	98.59	98.68	97.73	97.65
730	100.00	100.00	100.00	98.59	98.70	97.78	97.65
731	100.00	100.00	100.00	98.59	98.71	97.82	97.67
732	100.00	100.00	100.00	98.59	98.73	97.85	97.70
733	100.00	100.00	100.00	98.59	98.73	97.85	97.72
734	100.00	100.00	100.00	98.59	98.77	97.85	97.74
735	100.00	100.00	100.00	98.59	98.77	97.85	97.78
736	100.00	100.00	100.00	98.59	98.78	97.85	97.78
737	100.00	100.00	100.00	98.59	98.78	97.88	97.86
738	100.00	100.00	100.00	98.59	98.78	97.93	97.90
739	100.00	100.00	100.00	98.59	98.78	97.98	97.92
740	100.00	100.00	100.00	98.59	98.78	98.01	97.94
742	100.00	100.00	100.00	98.59	98.82	98.03	97.95
743	100.00	100.00	100.00	98.59	98.82	98.05	97.95
744	100.00	100.00	100.00	98.59	98.82	98.10	97.97
745	100.00	100.00	100.00	98.59	98.86	98.11	97.99
746	100.00	100.00	100.00	98.59	98.87	98.11	98.07
747	100.00	100.00	100.00	98.59	98.87	98.11	98.13
748	100.00	100.00	100.00	98.59	98.91	98.14	98.15
749	100.00	100.00	100.00	98.59	98.91	98.14	98.17
750	100.00	100.00	100.00	98.59	98.91	98.14	98.19
752	100.00	100.00	100.00	98.59	98.91	98.14	98.21
753	100.00	100.00	100.00	98.59	98.91	98.18	98.21
754	100.00	100.00	100.00	98.59	98.91	98.18	98.22
755	100.00	100.00	100.00	98.59	98.94	98.19	98.26
756	100.00	100.00	100.00	98.59	98.94	98.21	98.28
757	100.00	100.00	100.00	98.59	98.94	98.21	98.34
758	100.00	100.00	100.00	98.59	98.94	98.23	98.36

Table 11.14 *(continuation 13)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
759	100.00	100.00	100.00	98.59	98.96	98.24	98.38
760	100.00	100.00	100.00	98.59	99.03	98.29	98.40
761	100.00	100.00	100.00	98.59	99.05	98.34	98.40
763	100.00	100.00	100.00	98.59	99.07	98.37	98.40
764	100.00	100.00	100.00	98.59	99.08	98.39	98.44
765	100.00	100.00	100.00	98.59	99.08	98.41	98.53
767	100.00	100.00	100.00	98.59	99.12	98.41	98.53
768	100.00	100.00	100.00	98.59	99.14	98.41	98.53
770	100.00	100.00	100.00	98.59	99.24	98.41	98.55
771	100.00	100.00	100.00	98.59	99.26	98.44	98.55
772	100.00	100.00	100.00	98.59	99.26	98.46	98.59
773	100.00	100.00	100.00	98.59	99.26	98.57	98.61
776	100.00	100.00	100.00	98.59	99.26	98.57	98.63
777	100.00	100.00	100.00	98.59	99.28	98.57	98.63
778	100.00	100.00	100.00	98.59	99.28	98.60	98.63
779	100.00	100.00	100.00	98.59	99.28	98.62	98.63
781	100.00	100.00	100.00	98.59	99.28	98.64	98.63
782	100.00	100.00	100.00	98.59	99.30	98.67	98.63
783	100.00	100.00	100.00	98.59	99.30	98.69	98.65
784	100.00	100.00	100.00	98.59	99.31	98.74	98.71
787	100.00	100.00	100.00	98.59	99.31	98.74	98.77
788	100.00	100.00	100.00	98.59	99.31	98.77	98.77
789	100.00	100.00	100.00	98.59	99.33	98.82	98.77
790	100.00	100.00	100.00	98.59	99.35	98.82	98.77
791	100.00	100.00	100.00	98.59	99.35	98.83	98.77
793	100.00	100.00	100.00	98.59	99.37	98.83	98.77
794	100.00	100.00	100.00	98.59	99.37	98.85	98.78
795	100.00	100.00	100.00	98.59	99.37	98.85	98.80
796	100.00	100.00	100.00	98.59	99.37	98.87	98.86
797	100.00	100.00	100.00	98.59	99.37	98.88	98.88
800	100.00	100.00	100.00	100.00	99.37	98.88	98.88
802	100.00	100.00	100.00	100.00	99.37	98.90	98.88
804	100.00	100.00	100.00	100.00	99.47	98.90	98.90
805	100.00	100.00	100.00	100.00	99.47	98.90	98.92
806	100.00	100.00	100.00	100.00	99.47	98.92	98.92
807	100.00	100.00	100.00	100.00	99.49	98.92	98.92
809	100.00	100.00	100.00	100.00	99.51	98.92	98.92
810	100.00	100.00	100.00	100.00	99.51	98.93	98.92
812	100.00	100.00	100.00	100.00	99.51	98.93	98.94

Table 11.14 *(continuation 14)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
813	100.00	100.00	100.00	100.00	99.51	98.93	98.98
816	100.00	100.00	100.00	100.00	99.51	98.95	99.00
817	100.00	100.00	100.00	100.00	99.51	98.97	99.00
818	100.00	100.00	100.00	100.00	99.52	98.97	99.02
820	100.00	100.00	100.00	100.00	99.52	98.98	99.02
821	100.00	100.00	100.00	100.00	99.52	99.00	99.02
822	100.00	100.00	100.00	100.00	99.52	99.00	99.04
823	100.00	100.00	100.00	100.00	99.52	99.02	99.04
825	100.00	100.00	100.00	100.00	99.52	99.05	99.05
826	100.00	100.00	100.00	100.00	99.52	99.06	99.05
827	100.00	100.00	100.00	100.00	99.52	99.06	99.07
829	100.00	100.00	100.00	100.00	99.52	99.06	99.11
830	100.00	100.00	100.00	100.00	99.52	99.06	99.13
831	100.00	100.00	100.00	100.00	99.52	99.06	99.15
832	100.00	100.00	100.00	100.00	99.52	99.08	99.17
834	100.00	100.00	100.00	100.00	99.52	99.10	99.17
835	100.00	100.00	100.00	100.00	99.52	99.10	99.21
837	100.00	100.00	100.00	100.00	99.52	99.11	99.23
839	100.00	100.00	100.00	100.00	99.52	99.11	99.25
842	100.00	100.00	100.00	100.00	99.52	99.13	99.25
843	100.00	100.00	100.00	100.00	99.52	99.15	99.27
844	100.00	100.00	100.00	100.00	99.52	99.16	99.29
852	100.00	100.00	100.00	100.00	99.52	99.16	99.32
855	100.00	100.00	100.00	100.00	99.52	99.18	99.34
857	100.00	100.00	100.00	100.00	99.52	99.21	99.34
858	100.00	100.00	100.00	100.00	99.52	99.23	99.34
860	100.00	100.00	100.00	100.00	99.52	99.25	99.34
862	100.00	100.00	100.00	100.00	99.52	99.25	99.40
868	100.00	100.00	100.00	100.00	99.52	99.28	99.40
869	100.00	100.00	100.00	100.00	99.52	99.29	99.44
876	100.00	100.00	100.00	100.00	99.52	99.29	99.46
878	100.00	100.00	100.00	100.00	99.54	99.29	99.46
880	100.00	100.00	100.00	100.00	99.54	99.33	99.46
881	100.00	100.00	100.00	100.00	99.54	99.33	99.48
883	100.00	100.00	100.00	100.00	99.54	99.34	99.48

Table 11.15 Cumulative Distributions of Written Reporting Scores at Each Grade Level and Grade Span Data for Figure 11.8

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
150	0.17	0.11	0.00	0.00	0.00	0.00	0.00
175	0.19	0.11	0.00	0.00	0.00	0.00	0.00
210	0.22	0.11	0.00	0.00	0.00	0.00	0.00
222	0.24	0.11	0.00	0.00	0.00	0.00	0.00
239	0.26	0.11	0.00	0.00	0.00	0.00	0.00
245	0.30	0.11	0.00	0.00	0.00	0.00	0.00
248	0.36	0.11	0.00	0.00	0.00	0.00	0.00
252	0.37	0.11	0.00	0.00	0.00	0.00	0.00
256	0.39	0.11	0.00	0.00	0.00	0.00	0.00
259	0.41	0.11	0.00	0.00	0.00	0.00	0.00
261	0.43	0.11	0.00	0.00	0.00	0.00	0.00
262	0.45	0.11	0.00	0.00	0.00	0.00	0.00
268	0.47	0.11	0.00	0.00	0.00	0.00	0.00
271	0.51	0.11	0.00	0.00	0.00	0.00	0.00
275	0.51	0.13	0.00	0.00	0.00	0.00	0.00
277	0.52	0.13	0.00	0.00	0.00	0.00	0.00
278	0.56	0.13	0.00	0.00	0.00	0.00	0.00
279	0.62	0.13	0.00	0.00	0.00	0.02	0.00
280	0.64	0.13	0.00	0.00	0.00	0.02	0.00
281	0.66	0.13	0.00	0.00	0.00	0.02	0.00
283	0.67	0.13	0.00	0.00	0.00	0.02	0.00
285	0.69	0.13	0.00	0.00	0.00	0.02	0.00
286	0.71	0.13	0.00	0.00	0.00	0.02	0.00
287	0.77	0.13	0.00	0.00	0.00	0.03	0.00
288	0.84	0.13	0.00	0.00	0.00	0.03	0.00
289	0.88	0.13	0.00	0.00	0.00	0.03	0.00
290	0.92	0.13	0.00	0.00	0.00	0.03	0.00
292	0.94	0.13	0.00	0.00	0.00	0.03	0.00
293	0.96	0.13	0.00	0.00	0.00	0.03	0.00
294	0.99	0.13	0.00	0.00	0.00	0.03	0.00
295	1.01	0.13	0.00	0.00	0.00	0.03	0.00
296	1.05	0.13	0.00	0.00	0.00	0.03	0.00
297	1.09	0.13	0.00	0.00	0.00	0.03	0.00
298	1.14	0.13	0.00	0.00	0.00	0.03	0.00
300	1.16	0.13	0.00	0.00	0.00	0.03	0.00
301	1.20	0.14	0.00	0.00	0.00	0.03	0.00
302	1.24	0.14	0.02	0.00	0.00	0.03	0.00
303	1.31	0.14	0.02	0.00	0.00	0.03	0.00

Table 11.15 *(continuation one)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
304	1.37	0.16	0.02	0.00	0.00	0.03	0.00
305	1.48	0.16	0.02	0.00	0.00	0.03	0.00
306	1.55	0.16	0.02	0.00	0.00	0.03	0.00
307	1.59	0.16	0.02	0.00	0.00	0.03	0.00
308	1.63	0.17	0.02	0.00	0.00	0.03	0.00
309	1.72	0.17	0.02	0.00	0.00	0.03	0.00
311	1.76	0.19	0.02	0.00	0.00	0.03	0.00
312	1.91	0.19	0.02	0.00	0.00	0.03	0.00
313	2.02	0.19	0.02	0.00	0.00	0.05	0.00
314	2.14	0.19	0.02	0.00	0.00	0.05	0.00
315	2.25	0.19	0.02	0.00	0.00	0.05	0.00
316	2.34	0.19	0.02	0.00	0.00	0.05	0.00
317	2.43	0.19	0.02	0.00	0.00	0.05	0.00
318	2.58	0.19	0.02	0.00	0.00	0.05	0.00
319	2.70	0.19	0.02	0.00	0.00	0.05	0.00
320	2.75	0.19	0.02	0.00	0.00	0.05	0.00
321	2.90	0.20	0.02	0.00	0.00	0.05	0.00
322	3.11	0.22	0.02	0.00	0.00	0.05	0.00
323	3.20	0.22	0.02	0.00	0.00	0.05	0.00
324	3.37	0.22	0.02	0.00	0.00	0.05	0.00
325	3.45	0.22	0.02	0.00	0.00	0.05	0.00
326	3.58	0.22	0.02	0.00	0.00	0.05	0.00
327	3.78	0.22	0.02	0.00	0.00	0.05	0.00
328	3.97	0.22	0.02	0.00	0.00	0.05	0.00
329	4.23	0.22	0.02	0.00	0.00	0.05	0.00
330	4.46	0.22	0.02	0.00	0.00	0.05	0.00
331	4.81	0.22	0.02	0.00	0.00	0.05	0.02
332	5.04	0.24	0.02	0.00	0.00	0.05	0.02
333	5.30	0.24	0.02	0.00	0.00	0.05	0.02
334	5.81	0.24	0.02	0.00	0.00	0.05	0.02
335	5.96	0.25	0.02	0.00	0.00	0.05	0.02
336	6.39	0.27	0.02	0.00	0.00	0.05	0.02
337	6.82	0.30	0.02	0.00	0.00	0.05	0.02
338	7.19	0.30	0.02	0.00	0.00	0.05	0.02
339	7.42	0.31	0.02	0.00	0.00	0.05	0.02
340	7.90	0.31	0.02	0.00	0.00	0.05	0.02
341	8.35	0.31	0.02	0.00	0.00	0.05	0.02
342	8.62	0.33	0.02	0.00	0.00	0.05	0.02
343	9.12	0.33	0.03	0.00	0.00	0.05	0.02

Table 11.15 *(continuation two)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
344	9.55	0.35	0.03	0.00	0.00	0.06	0.02
345	10.06	0.39	0.03	0.00	0.00	0.06	0.02
346	10.70	0.41	0.03	0.00	0.00	0.06	0.02
347	11.44	0.42	0.03	0.00	0.00	0.06	0.02
348	12.03	0.44	0.03	0.00	0.00	0.06	0.02
349	12.61	0.52	0.03	0.00	0.00	0.06	0.02
350	13.26	0.56	0.03	0.00	0.00	0.06	0.02
351	13.64	0.58	0.03	0.00	0.00	0.06	0.02
352	14.09	0.61	0.03	0.00	0.00	0.06	0.02
353	14.80	0.71	0.03	0.02	0.02	0.06	0.02
354	15.28	0.77	0.03	0.02	0.02	0.08	0.02
355	15.81	0.80	0.05	0.02	0.02	0.08	0.02
356	16.24	0.88	0.06	0.02	0.02	0.08	0.02
357	16.76	0.91	0.08	0.02	0.02	0.08	0.02
358	17.38	0.93	0.09	0.02	0.02	0.08	0.02
359	18.08	0.94	0.09	0.02	0.02	0.08	0.02
360	18.69	0.99	0.09	0.02	0.02	0.08	0.04
361	19.24	1.00	0.11	0.03	0.02	0.08	0.04
362	19.84	1.08	0.13	0.03	0.02	0.10	0.04
363	20.49	1.18	0.13	0.03	0.02	0.10	0.04
364	21.24	1.18	0.14	0.03	0.02	0.11	0.04
365	21.78	1.26	0.14	0.03	0.03	0.11	0.04
366	22.33	1.30	0.14	0.03	0.03	0.13	0.04
367	22.81	1.36	0.17	0.05	0.05	0.16	0.04
368	23.53	1.41	0.17	0.05	0.05	0.16	0.04
369	24.39	1.44	0.17	0.05	0.05	0.16	0.04
370	25.38	1.58	0.17	0.05	0.07	0.18	0.04
371	26.35	1.71	0.20	0.05	0.07	0.18	0.04
372	26.86	1.77	0.22	0.07	0.07	0.19	0.04
373	27.68	1.95	0.25	0.10	0.07	0.19	0.04
374	28.62	2.07	0.30	0.10	0.07	0.21	0.04
375	29.58	2.24	0.38	0.10	0.07	0.24	0.04
376	30.25	2.46	0.42	0.10	0.07	0.28	0.04
377	31.22	2.60	0.49	0.10	0.09	0.29	0.04
378	32.14	2.78	0.49	0.12	0.09	0.31	0.06
379	32.89	2.86	0.50	0.13	0.09	0.34	0.06
380	33.70	3.11	0.58	0.15	0.09	0.34	0.06
381	34.69	3.31	0.63	0.15	0.10	0.34	0.06
382	35.61	3.39	0.68	0.15	0.14	0.34	0.06

Table 11.15 *(continuation three)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
383	36.24	3.69	0.82	0.15	0.14	0.34	0.06
384	36.94	3.89	0.88	0.18	0.14	0.36	0.06
385	37.84	4.11	0.97	0.18	0.14	0.36	0.06
386	38.47	4.31	1.04	0.18	0.14	0.41	0.06
387	39.17	4.47	1.16	0.21	0.17	0.45	0.06
388	39.76	4.68	1.22	0.21	0.21	0.52	0.06
389	40.40	5.00	1.33	0.23	0.21	0.52	0.06
390	41.21	5.19	1.43	0.25	0.21	0.55	0.08
391	41.92	5.54	1.54	0.26	0.23	0.57	0.08
392	42.56	5.82	1.70	0.31	0.23	0.62	0.08
393	43.51	6.12	1.77	0.33	0.24	0.62	0.10
394	44.30	6.39	1.88	0.39	0.26	0.63	0.10
395	45.08	6.70	2.06	0.43	0.28	0.65	0.10
396	45.98	6.98	2.17	0.44	0.30	0.65	0.12
397	46.71	7.39	2.28	0.46	0.30	0.65	0.12
398	47.50	7.66	2.40	0.48	0.30	0.70	0.13
399	48.46	8.02	2.65	0.53	0.30	0.71	0.13
400	49.37	8.32	2.76	0.56	0.31	0.78	0.13
401	50.29	8.82	2.95	0.58	0.33	0.81	0.13
402	51.26	9.33	3.12	0.63	0.33	0.84	0.15
403	52.15	9.92	3.28	0.64	0.35	0.88	0.15
404	53.14	10.48	3.49	0.64	0.35	0.92	0.17
405	53.91	10.81	3.67	0.66	0.38	0.96	0.19
406	54.79	11.37	3.97	0.67	0.38	1.01	0.25
407	55.52	11.83	4.11	0.71	0.38	1.02	0.25
408	56.57	12.44	4.47	0.81	0.40	1.05	0.25
409	57.46	12.82	4.69	0.86	0.42	1.07	0.27
410	58.42	13.40	4.95	0.87	0.45	1.09	0.29
411	59.39	13.96	5.31	0.87	0.49	1.15	0.29
412	60.24	14.59	5.54	0.94	0.54	1.23	0.29
413	61.23	15.09	5.84	0.99	0.59	1.27	0.33
414	61.96	15.77	5.95	1.04	0.59	1.31	0.33
415	62.93	16.54	6.31	1.07	0.62	1.33	0.33
416	63.74	17.29	6.55	1.18	0.64	1.38	0.35
417	64.62	18.00	6.91	1.32	0.66	1.40	0.35
418	65.39	18.76	7.21	1.40	0.66	1.43	0.35
419	66.12	19.50	7.57	1.48	0.75	1.49	0.37
420	67.17	20.29	7.99	1.50	0.76	1.51	0.37
421	68.18	21.10	8.27	1.56	0.80	1.51	0.38

Table 11.15 *(continuation four)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
422	68.93	22.01	8.62	1.73	0.87	1.51	0.42
423	69.81	22.73	9.01	1.81	0.94	1.57	0.44
424	70.63	23.42	9.59	1.94	1.01	1.69	0.48
425	71.31	24.10	10.00	2.06	1.02	1.70	0.52
426	72.19	24.93	10.44	2.24	1.04	1.70	0.52
427	72.90	25.70	10.82	2.35	1.08	1.75	0.52
428	73.74	26.58	11.49	2.45	1.11	1.75	0.52
429	74.32	27.49	12.09	2.53	1.20	1.87	0.62
430	75.03	28.37	12.36	2.72	1.25	1.96	0.63
431	75.78	29.28	12.99	2.90	1.37	2.01	0.65
432	76.31	30.23	13.44	2.99	1.41	2.14	0.73
433	77.02	31.10	13.99	3.32	1.49	2.19	0.77
434	77.73	31.83	14.54	3.51	1.56	2.30	0.83
435	78.67	32.74	15.18	3.62	1.58	2.34	0.85
436	79.27	33.50	15.86	3.77	1.70	2.40	0.87
437	79.75	34.42	16.36	3.95	1.77	2.45	0.94
438	80.22	35.35	16.91	4.11	1.87	2.63	1.00
439	80.88	36.43	17.52	4.25	1.93	2.69	1.06
440	81.35	37.29	18.26	4.31	2.00	2.77	1.10
441	81.85	38.52	18.84	4.59	2.10	2.89	1.15
442	82.30	39.25	19.53	4.81	2.19	3.03	1.15
443	82.58	40.05	20.25	5.00	2.29	3.13	1.19
444	83.18	40.85	20.91	5.15	2.34	3.26	1.25
445	83.57	42.05	21.78	5.30	2.40	3.31	1.27
446	83.80	43.05	22.48	5.50	2.48	3.41	1.38
447	84.23	43.87	23.25	5.71	2.57	3.55	1.42
448	84.77	44.70	23.91	5.92	2.71	3.67	1.44
449	85.11	45.64	24.56	6.15	2.88	3.83	1.60
450	85.48	46.58	25.25	6.52	2.99	3.99	1.62
451	85.97	47.41	26.00	6.81	3.07	4.09	1.71
452	86.18	48.38	26.72	7.09	3.18	4.19	1.87
453	86.33	49.22	27.59	7.50	3.28	4.30	1.92
454	86.81	50.25	28.47	7.93	3.49	4.45	1.98
455	87.04	51.11	29.17	8.28	3.64	4.53	2.08
456	87.60	52.04	29.94	8.66	3.71	4.64	2.08
457	87.84	53.06	30.60	9.08	3.85	4.84	2.19
458	88.33	53.94	31.10	9.45	3.99	4.92	2.21
459	88.48	54.90	31.72	9.77	4.22	5.09	2.31
460	88.63	55.55	32.50	10.19	4.32	5.27	2.42

Table 11.15 *(continuation five)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
461	89.06	56.32	33.24	10.65	4.55	5.47	2.46
462	89.21	57.48	34.09	11.06	4.79	5.68	2.54
463	89.34	58.36	34.79	11.44	5.02	5.95	2.77
464	89.64	59.21	35.64	11.72	5.15	6.23	2.83
465	89.79	59.85	36.38	12.14	5.40	6.47	2.94
466	90.05	60.75	37.24	12.67	5.66	6.69	3.10
467	90.30	61.77	38.22	13.07	5.95	6.86	3.23
468	90.43	62.68	39.03	13.68	6.21	7.07	3.44
469	90.75	63.41	39.98	14.46	6.37	7.27	3.58
470	90.94	64.23	40.71	15.07	6.63	7.40	3.69
471	91.14	65.03	41.55	15.65	6.80	7.64	3.87
472	91.16	65.97	42.35	16.18	7.05	7.95	4.06
473	91.42	66.96	43.32	16.69	7.39	8.15	4.14
474	91.48	67.79	44.26	17.25	7.65	8.34	4.39
475	91.61	68.64	45.33	17.74	7.95	8.58	4.54
476	91.72	69.23	46.29	18.38	8.30	8.76	4.77
477	91.80	70.14	47.07	19.12	8.57	8.94	4.90
478	91.83	70.88	48.31	19.75	8.82	9.14	5.02
479	91.91	71.57	49.18	20.26	9.32	9.30	5.29
480	92.23	72.40	49.98	21.10	9.56	9.65	5.58
481	92.23	73.16	50.81	21.80	9.93	9.95	5.83
482	92.32	74.05	51.52	22.51	10.27	10.24	6.04
483	92.49	74.68	52.47	23.35	10.59	10.66	6.14
484	92.55	75.29	53.60	24.12	11.02	10.97	6.29
485	92.83	75.97	54.55	24.73	11.40	11.31	6.54
486	92.96	76.66	55.41	25.49	11.80	11.76	7.00
487	93.09	77.11	56.57	26.08	12.31	12.17	7.19
488	93.11	77.68	57.33	26.71	12.83	12.49	7.50
489	93.26	78.55	58.17	27.55	13.12	12.90	7.83
490	93.33	79.06	59.30	28.24	13.62	13.40	8.16
491	93.39	79.51	60.36	29.04	14.14	14.00	8.48
492	93.54	80.28	61.12	29.95	14.80	14.44	8.77
493	93.67	80.80	62.00	30.84	15.31	14.88	9.02
494	93.71	81.39	62.91	31.87	15.93	15.38	9.39
495	93.95	81.77	63.73	32.86	16.59	15.95	9.66
496	93.97	82.19	64.63	33.88	17.27	16.37	10.00
497	94.19	82.88	65.65	34.94	17.89	16.91	10.58
498	94.25	83.56	66.23	35.64	18.73	17.54	11.10
499	94.53	84.08	67.04	36.40	19.39	18.11	11.46

Table 11.15 *(continuation six)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
500	94.64	84.61	67.95	37.52	20.08	18.77	11.75
501	94.72	85.16	68.66	38.42	20.74	19.29	12.21
502	95.13	85.55	69.45	39.30	21.43	19.84	12.58
503	95.21	85.86	70.34	40.22	22.13	20.46	13.12
504	95.22	86.18	71.28	41.32	22.93	21.05	13.58
505	95.22	86.38	72.02	42.31	23.69	21.84	13.98
506	95.28	86.87	72.71	43.20	24.23	22.46	14.54
507	95.28	87.12	73.53	44.45	25.03	23.06	15.00
508	95.28	87.37	74.20	45.53	25.63	23.71	15.45
509	95.28	87.92	74.94	46.39	26.40	24.49	15.95
510	95.28	88.17	75.82	47.39	27.21	25.12	16.56
511	95.39	88.42	76.45	48.41	27.89	25.77	17.18
512	95.39	88.80	77.30	49.60	28.90	26.38	17.87
513	95.41	89.10	77.81	50.63	29.42	27.18	18.33
514	95.58	89.44	78.62	51.67	30.34	27.78	18.98
515	95.64	89.90	79.18	52.81	31.01	28.40	19.56
516	95.66	90.07	79.65	53.73	31.74	28.88	20.00
517	95.82	90.40	80.17	54.62	32.65	29.60	20.64
518	95.84	90.74	80.80	55.74	33.50	30.31	21.24
519	95.84	90.85	81.30	56.59	34.50	30.93	21.79
520	95.84	90.98	81.65	57.74	35.34	31.49	22.39
521	95.84	91.18	82.27	59.08	36.26	32.45	23.08
522	95.84	91.39	82.70	59.90	37.16	33.13	24.06
523	95.84	91.58	83.11	60.90	38.04	34.06	24.85
524	95.84	92.11	83.61	62.07	38.88	34.72	25.51
525	95.86	92.31	84.11	62.89	40.02	35.42	26.22
526	95.92	92.63	84.43	63.54	40.98	36.25	27.01
527	95.92	92.63	84.88	64.44	41.74	37.22	27.60
528	95.95	92.80	85.30	65.51	42.71	38.03	28.16
529	95.95	92.83	85.90	66.33	43.77	38.49	28.93
530	95.95	92.96	86.20	67.39	45.16	39.28	29.54
531	95.95	93.30	86.48	68.37	46.29	40.00	30.31
532	95.95	93.30	86.91	69.15	47.21	40.68	30.91
533	95.95	93.43	87.38	70.05	48.02	41.44	31.62
534	95.95	93.60	87.64	71.01	49.29	42.20	32.28
535	96.31	93.60	87.97	71.98	50.42	43.16	33.08
536	96.31	93.83	88.37	72.70	51.55	44.01	33.87
537	96.31	94.05	88.65	73.52	52.59	44.70	34.41
538	96.31	94.05	89.01	74.40	53.75	45.34	35.28

Table 11.15 *(continuation seven)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
539	96.31	94.21	89.39	75.10	54.58	46.02	36.12
540	96.31	94.85	89.73	75.84	55.52	46.93	36.95
541	96.31	94.85	89.94	76.70	56.58	47.82	37.68
542	96.31	95.12	90.19	77.34	57.46	48.52	38.32
543	96.31	95.12	90.71	78.13	58.56	49.33	39.05
544	96.31	95.61	90.86	78.67	59.39	50.14	39.64
545	96.31	95.78	91.10	79.46	60.59	50.87	40.51
546	96.31	95.78	91.33	80.19	61.65	51.81	41.35
547	96.31	95.78	91.66	81.01	62.51	52.54	42.01
548	96.31	95.78	92.07	81.59	63.45	53.40	43.09
549	96.40	95.78	92.28	82.18	64.41	54.24	43.86
550	96.40	95.78	92.64	82.85	65.22	55.02	44.76
551	96.40	95.91	92.73	83.38	66.14	55.98	45.45
552	96.40	95.97	92.83	83.96	67.01	56.58	46.22
553	96.40	95.97	93.01	84.42	68.17	57.36	47.07
554	96.54	95.97	93.19	84.99	69.11	58.09	47.64
555	96.54	95.97	93.23	85.52	69.79	58.84	48.32
556	96.54	96.24	93.48	86.14	70.83	59.39	48.86
557	96.54	96.24	93.59	86.52	71.56	60.15	49.72
558	96.63	96.24	93.72	86.98	72.25	60.88	50.68
559	96.63	96.24	93.96	87.43	73.00	61.46	51.45
560	96.63	96.25	94.21	87.94	73.72	62.18	52.34
561	96.63	96.25	94.54	88.30	74.51	62.78	52.99
562	96.63	96.25	94.62	88.73	75.22	63.65	53.90
563	96.63	96.25	94.80	89.24	75.95	64.42	54.78
564	96.63	96.25	94.90	89.63	76.81	65.07	55.67
565	96.63	96.25	95.05	90.09	77.60	65.97	56.65
566	96.63	96.25	95.09	90.36	78.19	66.62	57.20
567	96.63	96.25	95.46	90.69	79.02	67.35	57.82
568	96.63	96.25	95.57	91.05	79.63	67.79	58.51
569	96.63	96.36	95.65	91.36	79.99	68.49	59.36
570	96.63	96.36	95.68	91.67	80.67	69.22	60.07
571	96.63	96.36	95.73	92.05	81.29	69.80	60.80
572	96.76	96.36	95.82	92.32	81.93	70.47	61.70
573	96.76	96.36	95.84	92.55	82.59	71.22	62.26
574	96.76	96.39	95.84	92.79	83.25	71.96	63.17
575	96.76	96.39	95.86	93.11	83.96	72.64	64.03
576	96.76	96.39	95.87	93.42	84.57	73.39	64.74
577	96.76	96.39	95.93	93.71	85.08	73.89	65.55

Table 11.15 *(continuation eight)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
578	96.76	96.39	96.15	93.96	85.54	74.43	66.76
579	96.76	96.39	96.17	94.22	86.05	75.00	67.49
580	96.76	96.39	96.23	94.34	86.50	75.56	68.30
581	96.76	96.39	96.36	94.50	87.05	76.18	69.07
582	96.76	96.39	96.36	94.78	87.57	76.78	69.69
583	96.87	96.39	96.45	94.95	88.13	77.32	70.49
584	96.87	96.39	96.45	95.21	88.42	77.72	71.23
585	96.87	96.39	96.45	95.43	88.89	78.24	71.88
586	96.87	96.39	96.45	95.62	89.28	78.89	72.61
587	96.87	96.39	96.45	95.85	89.69	79.47	73.38
588	96.87	96.39	96.48	95.99	90.02	80.01	73.90
589	96.87	96.39	96.51	96.08	90.40	80.56	74.42
590	96.87	96.39	96.51	96.23	90.75	81.00	74.94
591	96.87	96.39	96.69	96.48	91.03	81.54	75.80
592	96.87	96.39	96.72	96.64	91.36	82.25	76.42
593	96.87	96.39	96.72	96.68	91.76	82.67	77.15
594	96.87	96.39	96.73	96.87	92.05	83.11	77.61
595	96.87	96.39	96.73	97.05	92.40	83.50	78.23
596	96.87	96.39	96.73	97.17	92.68	83.86	78.82
597	96.87	96.39	96.73	97.24	92.85	84.23	79.28
598	96.87	96.39	96.73	97.33	93.11	84.62	79.82
599	96.87	96.39	96.73	97.55	93.32	85.12	80.55
600	96.87	96.39	96.73	97.61	93.54	85.46	81.04
601	96.87	96.39	96.78	97.84	93.79	85.90	81.38
602	96.87	96.39	96.78	97.94	94.05	86.34	81.94
603	96.87	96.39	96.78	97.98	94.31	86.70	82.46
604	96.87	96.39	96.89	98.09	94.48	87.10	82.80
605	96.87	96.39	96.89	98.14	94.79	87.46	83.29
606	96.87	96.39	96.89	98.22	95.07	87.81	83.77
607	96.87	96.39	96.89	98.26	95.26	88.07	84.27
608	96.87	96.39	96.89	98.32	95.49	88.40	84.90
609	96.87	96.39	96.91	98.37	95.61	88.67	85.40
610	96.87	96.39	96.91	98.49	95.77	89.03	85.75
611	96.87	96.39	97.02	98.54	95.84	89.36	86.07
612	96.87	96.39	97.24	98.62	95.97	89.79	86.57
613	96.87	96.39	97.24	98.68	96.10	90.23	86.98
614	96.87	96.39	97.24	98.72	96.29	90.46	87.50
615	96.87	96.39	97.24	98.75	96.34	90.90	87.96
616	96.87	96.39	97.24	98.80	96.51	91.09	88.36

Table 11.15 *(continuation nine)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
617	96.87	96.39	97.24	98.83	96.65	91.37	88.79
618	96.87	96.39	97.24	98.96	96.72	91.68	89.27
619	96.87	96.39	97.24	98.98	96.95	92.07	89.61
620	96.87	96.39	97.24	99.03	97.03	92.23	89.96
621	96.87	96.39	97.24	99.05	97.15	92.49	90.40
622	96.87	96.39	97.24	99.08	97.38	92.68	90.77
623	96.87	96.39	97.24	99.14	97.45	92.89	91.15
624	96.87	96.39	97.58	99.16	97.50	93.06	91.36
625	96.87	96.39	97.58	99.16	97.62	93.27	91.73
626	96.87	96.39	97.58	99.23	97.76	93.43	92.02
627	96.87	96.39	97.58	99.26	97.83	93.62	92.23
628	96.87	96.39	97.58	99.26	97.88	93.75	92.48
629	96.87	96.39	97.58	99.28	97.92	93.90	92.73
630	96.87	96.39	97.58	99.31	97.97	94.11	93.02
631	96.87	96.39	97.58	99.34	98.00	94.32	93.35
632	96.87	96.39	97.58	99.36	98.09	94.50	93.67
633	96.87	96.39	97.58	99.41	98.16	94.63	93.88
634	96.87	96.39	97.58	99.41	98.21	94.78	94.13
635	96.87	96.39	97.58	99.44	98.28	94.95	94.36
636	96.87	96.39	97.58	99.49	98.32	95.17	94.61
637	96.87	96.39	97.58	99.52	98.40	95.31	94.77
638	96.87	96.39	97.58	99.56	98.42	95.44	94.94
639	96.87	96.39	97.58	99.56	98.54	95.60	95.08
640	96.87	96.39	97.58	99.59	98.65	95.72	95.29
641	96.87	96.39	97.58	99.59	98.73	95.80	95.46
642	96.87	96.39	97.58	99.61	98.82	95.98	95.52
643	96.87	96.39	97.58	99.64	98.87	96.14	95.81
644	96.87	96.39	97.58	99.64	98.89	96.20	95.98
645	96.87	96.39	97.58	99.65	98.92	96.25	96.12
646	96.87	96.39	97.58	99.65	98.94	96.32	96.23
647	96.87	96.39	97.58	99.65	98.96	96.48	96.38
648	96.87	96.39	97.58	99.67	98.99	96.56	96.56
649	96.87	96.39	97.58	99.69	99.06	96.72	96.81
650	96.87	96.39	97.76	99.69	99.12	96.87	96.90
651	96.87	96.39	97.76	99.72	99.13	96.93	96.94
652	96.87	96.39	97.76	99.75	99.20	97.01	97.06
653	96.87	96.39	97.76	99.75	99.20	97.11	97.13
654	96.87	96.39	97.76	99.75	99.22	97.18	97.23
655	96.87	96.39	97.76	99.75	99.25	97.34	97.37

Table 11.15 *(continuation 10)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
656	96.87	96.39	97.76	99.77	99.25	97.47	97.46
657	96.87	96.39	97.76	99.77	99.27	97.52	97.63
658	96.87	96.39	97.76	99.77	99.27	97.62	97.69
659	96.87	96.39	97.76	99.77	99.27	97.68	97.81
660	96.87	96.39	97.76	99.77	99.29	97.75	97.87
661	96.87	96.39	97.76	99.80	99.29	97.81	97.92
662	96.87	96.39	97.76	99.80	99.32	97.89	97.96
663	96.87	96.39	97.76	99.80	99.34	97.94	98.02
664	96.87	96.39	97.76	99.80	99.38	97.97	98.06
665	96.87	96.39	97.76	99.84	99.39	97.99	98.10
666	96.87	96.39	97.76	99.84	99.39	98.07	98.12
667	96.87	96.39	97.76	99.85	99.39	98.15	98.21
668	96.87	96.39	97.76	99.87	99.41	98.18	98.23
669	96.87	96.39	97.76	99.87	99.45	98.25	98.35
670	96.87	96.39	97.76	99.87	99.48	98.35	98.44
671	96.87	96.39	97.76	99.87	99.50	98.43	98.52
672	96.87	96.39	97.76	99.87	99.53	98.51	98.56
673	96.87	96.39	97.76	99.87	99.53	98.54	98.63
674	96.87	96.39	97.76	99.87	99.55	98.57	98.71
675	96.87	96.39	97.76	99.87	99.57	98.62	98.79
676	96.87	96.39	97.76	99.89	99.57	98.64	98.81
677	96.87	96.39	97.76	99.89	99.58	98.69	98.87
678	96.87	96.39	97.76	99.90	99.58	98.72	98.90
679	96.87	96.39	97.76	99.90	99.60	98.80	98.96
680	96.87	96.39	97.76	99.90	99.60	98.86	98.96
681	96.87	96.39	97.76	99.90	99.62	98.88	99.00
682	96.87	96.39	97.76	99.90	99.62	98.93	99.04
683	96.87	96.39	97.76	99.90	99.62	99.01	99.10
684	96.87	96.39	97.76	99.90	99.62	99.03	99.12
685	96.87	96.39	97.76	99.90	99.64	99.03	99.17
686	96.87	96.39	97.76	99.92	99.64	99.09	99.19
687	96.87	96.39	97.76	99.92	99.65	99.11	99.21
688	96.87	96.39	97.76	99.92	99.65	99.12	99.21
689	96.87	96.39	97.76	99.92	99.65	99.14	99.27
690	96.87	96.39	97.76	99.92	99.67	99.17	99.29
691	96.87	96.39	97.76	99.92	99.69	99.17	99.31
692	96.87	96.39	97.76	99.92	99.69	99.24	99.33
693	96.87	96.39	97.76	99.92	99.69	99.29	99.37
694	96.87	96.39	97.76	99.92	99.69	99.30	99.38

Table 11.15 *(continuation 11)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
695	96.87	96.39	97.76	99.92	99.72	99.30	99.38
696	96.87	96.39	97.76	99.92	99.74	99.30	99.42
697	96.87	96.39	97.76	99.92	99.77	99.30	99.44
698	96.87	96.39	97.76	99.92	99.77	99.30	99.46
699	96.87	96.39	97.76	99.92	99.77	99.34	99.50
700	100.00	100.00	100.00	99.92	99.77	99.34	99.52
701	100.00	100.00	100.00	99.95	99.77	99.35	99.60
702	100.00	100.00	100.00	99.95	99.77	99.38	99.60
703	100.00	100.00	100.00	99.97	99.77	99.38	99.60
705	100.00	100.00	100.00	99.97	99.77	99.40	99.60
706	100.00	100.00	100.00	99.97	99.77	99.42	99.62
707	100.00	100.00	100.00	99.97	99.77	99.43	99.62
708	100.00	100.00	100.00	99.97	99.77	99.43	99.64
709	100.00	100.00	100.00	99.97	99.77	99.45	99.67
710	100.00	100.00	100.00	99.97	99.77	99.47	99.67
713	100.00	100.00	100.00	99.97	99.79	99.50	99.67
714	100.00	100.00	100.00	99.97	99.79	99.53	99.67
715	100.00	100.00	100.00	99.97	99.79	99.58	99.67
716	100.00	100.00	100.00	99.97	99.79	99.61	99.69
717	100.00	100.00	100.00	99.97	99.79	99.61	99.71
718	100.00	100.00	100.00	99.97	99.79	99.61	99.73
719	100.00	100.00	100.00	99.97	99.79	99.66	99.73
721	100.00	100.00	100.00	99.97	99.81	99.66	99.73
722	100.00	100.00	100.00	99.97	99.81	99.69	99.73
724	100.00	100.00	100.00	99.97	99.83	99.69	99.75
725	100.00	100.00	100.00	99.97	99.83	99.71	99.75
726	100.00	100.00	100.00	99.97	99.83	99.71	99.77
727	100.00	100.00	100.00	99.97	99.84	99.71	99.77
730	100.00	100.00	100.00	99.97	99.84	99.71	99.79
733	100.00	100.00	100.00	99.97	99.84	99.71	99.81
734	100.00	100.00	100.00	99.97	99.84	99.74	99.81
735	100.00	100.00	100.00	99.97	99.84	99.74	99.83
736	100.00	100.00	100.00	99.97	99.84	99.76	99.83
737	100.00	100.00	100.00	99.97	99.84	99.77	99.83
739	100.00	100.00	100.00	99.97	99.84	99.79	99.83
740	100.00	100.00	100.00	99.97	99.86	99.79	99.83
741	100.00	100.00	100.00	99.97	99.86	99.79	99.85
742	100.00	100.00	100.00	99.97	99.86	99.79	99.87
746	100.00	100.00	100.00	99.97	99.86	99.81	99.89

Table 11.15 *(continuation 12)*

Scale Score	K	G1	G2	G3–5	G6–8	G9–10	G11–12
747	100.00	100.00	100.00	99.97	99.86	99.82	99.89
748	100.00	100.00	100.00	99.97	99.88	99.82	99.89
749	100.00	100.00	100.00	99.97	99.88	99.82	99.90
751	100.00	100.00	100.00	99.97	99.88	99.84	99.90
754	100.00	100.00	100.00	99.97	99.88	99.85	99.90
755	100.00	100.00	100.00	99.97	99.90	99.85	99.90
764	100.00	100.00	100.00	99.97	99.90	99.87	99.90
767	100.00	100.00	100.00	99.97	99.91	99.87	99.90
773	100.00	100.00	100.00	99.97	99.93	99.89	99.92
774	100.00	100.00	100.00	99.97	99.93	99.90	99.92
782	100.00	100.00	100.00	99.97	99.93	99.92	99.92
783	100.00	100.00	100.00	99.97	99.93	99.94	99.92
788	100.00	100.00	100.00	99.98	99.93	99.94	99.92
791	100.00	100.00	100.00	99.98	99.93	99.95	99.92
793	100.00	100.00	100.00	99.98	99.93	99.95	99.94
799	100.00	100.00	100.00	99.98	99.93	99.97	99.94
800	100.00	100.00	100.00	100.00	99.93	99.97	99.94
802	100.00	100.00	100.00	100.00	99.93	99.98	99.94
808	100.00	100.00	100.00	100.00	99.95	99.98	99.94
809	100.00	100.00	100.00	100.00	99.95	99.98	99.96
813	100.00	100.00	100.00	100.00	99.97	99.98	99.96
832	100.00	100.00	100.00	100.00	99.97	99.98	99.98
841	100.00	100.00	100.00	100.00	99.97	100.00	99.98
861	100.00	100.00	100.00	100.00	99.97	100.00	100.00
900	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Chapter 12: Continuous Program Improvement

This chapter summarizes the completed and ongoing improvements for the Summative ELPAC in the areas of threshold validation, test development, and test delivery.

12.1. Summative Threshold Score Validation Study

12.1.1. Overview of the Threshold Score Validation Study

At the request of the California Department of Education (CDE), Educational Testing Service (ETS) conducted a threshold score validation study to provide additional validity evidence for the Summative English Language Proficiency Assessments for California (ELPAC) preliminary threshold scores (CDE, 2018). The process that was implemented is known as the Contrasting Groups Method (Zieky & Livingston, 2008). In this method, teachers familiar with the 2012 *California English Language Development Standards, Kindergarten Through Grade 12* and with the students in their classroom who are classified as English learners (ELs) were asked to make judgments about students' ELPAC performance levels based on the approved ELPAC performance level descriptors (PLDs). Students' classifications into performance levels based on preliminary threshold scores were compared to the performance levels based on the teacher judgments from the contrasting groups method. Results from the validation study allowed the CDE to consider information across standard setting methods, as described here.

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 ELPAC consistently distinguished between levels of students' English language proficiency, based on teacher ratings from a multistep process (standard setting in October 2017 and the validation study). Implementing a multistep process offers increased confidence in decisions utilizing threshold scores based on ELPAC results. Because the proficiency classification of ELs entails relatively high-stakes decisions for an individual student's academic path, school program funding, and resource plans, it is crucial to cross-validate the threshold scores for each proficiency level to the extent possible.

12.1.2. Results

Results from the rater-agreement analysis of the contrasting groups study indicated a trend for most grades for most performance levels (CDE, 2018). Teacher ratings of expected levels were compared to the students' ELPAC levels based on the preliminary threshold scores. The comparison indicated an exact or adjacent agreement rate between 85 percent and 90 percent in kindergarten and grades one and two, and a rate at 90 percent or above for grades three through twelve. This means, for example, where teacher ratings of students indicated Level 3, exact or adjacent agreement was met if the students' ELPAC level was Level 2, 3, or 4, based on the ELPAC preliminary threshold scores.

For most grades, the difference in performance level classification between the teachers' ratings of students and the preliminary threshold scores shows that teachers who rated the students who are ELs in their classrooms placed their students at a lower performance level than what the ELPAC preliminary threshold score would have indicated. If the results from this study were used to set threshold scores, at least some threshold scores would be increased.

12.1.3. Impact of Threshold Validation and Additional Analyses

In addition to the results from the contrasting groups study indicating a trend toward an increase in threshold scores, additional analyses were conducted by the California Comprehensive Center at WestEd. These analyses evaluated the relationships between classifications of student proficiency based on the Smarter Balanced English language arts/literacy performance and the classifications of students based on ELPAC scores and performance levels. Results from the two studies were reviewed by the State Board of Education (SBE) and revised threshold scores were approved (SBE, 2017).

12.2. Test Development

One goal in the development of the 2017–18 Summative ELPAC was to appropriately consider and implement feedback received from the 2016–17 stand-alone field test. Feedback was collected from three sources:

1. "Blue cards" that test examiners administering the field test filled out during training to provide feedback
2. Input gathered in the Summative ELPAC Field Test Feedback Focus Group Sessions for Test Examiners and ELPAC Coordinators
3. Results of the Summative ELPAC online survey that was made available to some ELPAC coordinators and test examiners

The CDE reviewed the feedback from the three sources and provided ETS with proposed revisions to the 2017–18 Summative ELPAC. Based on the feedback, more than 100 changes to the test materials were implemented in the 2017–18 Summative ELPAC, including the following:

- **Speaking Audio Recordings:** Audio recordings were provided for the *Speaking—Summarize an Academic Presentation* section to standardize the delivery of the presentation.
- **Speaking Section in the Test Books:** Graphics for *Speaking—Present and Discuss Information* and *Speaking—Summarize an Academic Presentation* section were entered in the Test Books so students could write notes directly on the graphics to prepare responses.
- **One- or Two-Session Administrations:** At grade spans three through five, six through eight, nine and ten, and eleven and twelve, the option of having one-session or two-session administrations was provided for Reading and Writing.
- **Write-on Lines:** In the Writing domain, additional write-on lines were provided in the Answer Books for certain task types.

ETS assessment specialists worked with ETS production staff to implement the revisions in the 2017–18 Summative ELPAC.

12.3. Test Delivery

12.3.1. Post Administration Survey

During the 2017–18 Summative ELPAC administration, ETS administered a post-test survey to local educational agencies (LEAs). The survey focused on gathering information on the test materials delivery; clarity of the test administration, *Examiner's Manuals*, and return instructions; and overall administration experience.

In response to the LEA feedback ETS is implementing the following improvements for the 2018–19 administration:

- Reorganization of the test administration manual so the flow of information is easier for LEAs to follow
- Implementation of a new return process allowing LEAs to produce return labels, increasing materials tracking accuracy
- Simplification of return instructions
- Earlier delivery of test materials.

12.3.2. Training and Communication

As ETS continues work on the Summative ELPAC, recruitment, training, and communication will be a focal point moving forward. ETS will continue to provide timely communications for each critical component of the ELPAC administration, including material order dates and deadlines and training schedules. ETS will continue to work with the Sacramento County Office of Education to emphasize the importance and necessity of training, along with providing statewide training to LEA staff so they are prepared to administer the test. Training will continue to focus on local scoring of the Speaking and Listening domains.

ETS will continue to support familiarizing students with the ELPAC items using practice tests and informational videos. Parent engagement continues to be an important factor for student engagement and familiarization. To that end, ETS will work with the CDE to increase communication and information targeted at parents. Communications will also encourage LEAs to use the practice tests to prepare students to become more familiar with the ELPAC.

References

California Department of Education. (2018). *English Language Proficiency Assessments for California threshold score validation study final report* (Unpublished report). Sacramento, CA: California Department of Education.

California State Board of Education. (November 8, 2017). *Item 08 Addendum – English Language Proficiency Assessments for California: Approve the operational summative assessment threshold scores and composite weights for the English Language Proficiency Assessments for California and approve the local educational agency apportionment rates.* Presented at the California State Board of Education meeting, Sacramento, CA. Retrieved from https://www.cde.ca.gov/be/ag/ag/yr17/documents/nov17item08addendum.doc

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.