

California Department of Education

Report to the Legislature:

Kindergarten and Grade One Early Literacy Assessment: Results and Administrative Process



Prepared by:
Educational Data Systems
Assessment Development and Administration Division
District, School and Innovation Branch

MARCH 2013

Description: Report outlines the California English Language Development Test, Kindergarten and Grade One Literacy Assessment results and the administrative process.

Authority: California Education Code Section 60810(b)(1)

Recipient: The Legislature

Due Date: January 1, 2013

California Department of Education
Report to the Legislature:
Kindergarten and Grade One Early Literacy Assessment: Results and
Administrative Process

Table of Contents

| | |
|---|----|
| Executive Summary | 1 |
| 1. Background and Purpose | 3 |
| 2. Development and Description of the Kindergarten and Grade One Early Literacy Assessment..... | 5 |
| 3. Test Administration Process..... | 13 |
| 4. Test Reliability and Validity | 15 |
| 5. Results for Three Administrations of the Early Literacy Assessment | 21 |
| 6. Growth Over Time | 27 |
| 7. Summary and Recommendations for Future Development..... | 33 |
| Appendix | 35 |

California Department of Education
Report to the Legislature:
Kindergarten and Grade One Early Literacy Assessment: Results and
Administrative Process

List of Tables

| | |
|--|----|
| Table 2.1. CELDT K–1 Reading and Writing Test Performance Descriptors | 8 |
| Table 2.2. CELDT Cut Scores..... | 10 |
| Table 4.1. Average Scale Score Comparisons of EL Students and EF Students..... | 17 |
| Table 5.1. Number of K–1 Students Tested | 21 |
| Table 5.2 Program Participation of Students Tested..... | 22 |
| Table 5.3. Annual Assessment CELDT Average Scale Scores..... | 24 |
| Table 5.4. Percentage of Students in Performance Levels | 24 |
| Table 5.5. Student Performance Level..... | 25 |
| Table 6.1. Average Student Scale Score Increase by Grade | 29 |
| Table A.1. English Learner Expert Panel Members | 35 |
| Table A.2. CELDT Blueprint for K – 1 Early Literacy Assessment | 37 |

List of Figures

| | |
|--|----|
| Fig. 2.1. Early Literacy Assessment Development Path..... | 7 |
| Fig. 2.2. Sample Questions from the Early Literacy Assessment..... | 11 |
| Fig. 4.1. Percentage of Survey Respondents Rating Each Domain of the CELDT as Useful for Initial Classification | 18 |
| Fig. 4.2. Percentage of Survey Respondents Rating Each Domain of the CELDT as Useful for Instructional Decisions..... | 19 |
| Fig. 5.1. Reading Scale Score Distribution..... | 23 |
| Fig. 5.2. Writing Scale Score Distribution | 23 |
| Fig. 6.1. Average Reading Scale Scores for Students Matched from 2010–11 to 2011–12..... | 27 |
| Fig. 6.2. Average Writing Scale Scores for Students Matched from 2010–11 to 2011–12..... | 28 |
| Fig. 6.3. Average Overall Scale Scores for Students Matched from 2010–11 to 2011–12..... | 28 |
| Fig. 6.4. Average Individual Gain in Overall CELDT Scale Score from K–1..... | 30 |
| Fig. 6.5. Average Individual Gain in Reading Scale Score from K–1 | 31 |
| Fig. 6.6. Average Individual Gain in Writing Scale Score from K–1..... | 31 |

California Department of Education

Report to the Legislature:

Kindergarten and Grade One Early Literacy Assessment: Results and Administrative Process

Executive Summary

This report has been prepared by Educational Data Systems (EDS) under contract number CN088346 for the California Department of Education (CDE). The California State Legislature established a requirement that the CDE administer an early literacy assessment, consisting of reading and writing, to kindergarten and grade one (K–1) students as part of the California English Language Development Test (CELDT) in accordance with California *Education Code* Section 60810(b)(1). The legislation included a requirement that the CDE report to the Legislature on the K–1 early literacy assessment results as well as on the administrative process. This report has been prepared in response to that legislative requirement.

The report addresses the following topics: (1) development of the assessment, including the process used and the subject-matter experts contributing to its development; (2) description of the early literacy assessment, including sample questions used in the assessment; (3) administration process of the early literacy assessment; (4) reliability and validity of the assessment, including (a) its statistical properties; (b) the results of an empirical study comparing the performance of students receiving English language instruction and those not receiving such instruction; and (c) the results of a survey of teachers and administrators involved with the early literacy assessment; (5) statistical results for the first three years of test administration, including demographic characteristics of the test population, score averages, and performance-level results; and (6) improvement in individual student performance over time, including results of studies comparing the growth of students' English skills from one year to the next.

The key findings are as follows: (1) more than one million K–1 students took the early literacy assessment over the first three years of its use; (2) the assessment is reliable and valid for its intended purpose; (3) score differences between English-fluent students and English learners are highly significant, both statistically and practically; and (4) students retested after a year of school show great increases in test scores.

The report concludes with recommendations for further improvements to the early literacy assessment and its use.

A copy of this report is located on the CDE [CELDT Technical Documentation](#) Web page. A copy of the report can be obtained from Educational Data Systems, CELDT contractor, by phone at 408-776-7646.

1. Background and Purpose

The California English Language Development Test (CELDT) was developed by the California Department of Education (CDE), with the approval of the State Board of Education (SBE), in response to legislation requiring school districts to assess annually the English language proficiency of all students whose primary language is other than English. The CELDT originally consisted of listening, speaking, reading, and writing domains for students in grades two through twelve; and listening and speaking domains for kindergarten through grade one students. In 2009, reading and writing domains (the kindergarten and grade one [K–1] early literacy assessment) were administered to K–1 students for the first time in response to changes in the California *Education Code (EC)*. That legislation, *EC* Section 60810(b)(1), specified that:

The early literacy assessment shall be administered for a period of three years beginning after the initial administration of the assessment or until July 1, 2012, whichever occurs first. Six months after the results of the last administered assessment are collected, but no later than January 1, 2013, the department shall report to the Legislature on the administration of the kindergarten and grade 1 early literacy assessment results, as well as on the administrative process, in order to determine whether reauthorization of the early literacy assessment is appropriate.

This K–1 legislative report has been prepared in response to that legislative requirement. The report describes the process by which the early literacy assessment and test items were developed and validated; provides a description of the test administration process; addresses test reliability and validity; presents results for three groups of students who have taken the CELDT (the 2009–10, 2010–11, and 2011–12 student populations); and describes the great improvements in English literacy made by students enrolled in English-language development (ELD) programs, particularly in the first few years of instruction. This report concludes with a set of recommendations for the CDE and SBE to consider in planning future CELDT development.

EC Section 60810(d) states the purpose of the CELDT.

The test shall be used for the following purposes:

- (1) To identify pupils who are limited English proficient.
- (2) To determine the level of English language proficiency of pupils who are limited English proficient.
- (3) To assess the progress of limited English proficient pupils in acquiring the skills of listening, reading, speaking, and writing in English.

Responding to those requirements, the CDE, with the approval of the State Superintendent of Public Instruction and the SBE, developed the CELDT. Its first operational administration occurred between May 14 and October 31, 2001. The CELDT

assesses English learners (ELs) in the domains of listening, speaking, reading, and writing. It consists of five separate tests: K–1, grade two, grades three through five, grades six through eight, and grades nine through twelve.

Approximately 1.6 million students are tested each year, of which about 400,000 are K–1 students. Each year that the CELDT is administered, the CDE prepares an extensive technical report covering all domains and grades at which the CELDT is administered. This K–1 legislative report contains a high-level summary of a small portion of the data presented in these technical reports. The technical reports provide detailed information on the following topics:

- The item development and review process, including details regarding subject-matter experts involved in this process
- The rules for item selection and test assembly
- Procedures for assessing students with disabilities
- Scoring procedures
- The process of equating successive CELDT editions to the base form
- Results of analyses conducted to document the statistical properties of the CELDT
- Quality control procedures
- Historical comparative data

The development and administration of the K–1 early literacy assessment followed all the standard procedures as the other CELDT grades and domains. The annual technical reports available on the CDE Web page at <http://www.cde.ca.gov/ta/tg/el/techreport.asp> contain additional details about the CELDT and the results of each administration, including detailed results of the 2009–10, 2010–11, and 2011–12 K–1 early literacy assessment administrations. References will be made to the *CELDT 2010–11 Edition Technical Report* throughout this document.

2. Development and Description of the Kindergarten and Grade One Early Literacy Assessment

Although the focus of this report is on the K–1 reading and writing domains, these two domains are always administered along with the listening and speaking domains to students taking the CELDT. For each student, a scale score and performance level are produced for each domain. In addition to those scores, the CELDT also provides a comprehension score (based on the listening and reading domains) and an overall score (based on the scores of all four domains). References to these derived scores are made later in this report.

2.1 Content of the K–1 early literacy assessment

The CDE convened a meeting of the CELDT English Learner Expert Panel on May 28, 2008, in Sacramento, California, to guide the process of developing the early literacy assessment. Members were selected for their expertise in areas that included applied linguistics, English-language development and acquisition, early childhood development, testing, and psychometrics. The names and affiliations of the panel members are shown in Table A.1 of the appendix.

The panel began by developing a number of principles to guide test development, including the following:

- Clear rationale for aligning an item to a particular domain is important. Because an overlap among English language skills often occurs and listening and speaking likely are needed to engage students in reading and writing tasks, reading and writing items should require students to recognize and/or produce written English letters or words.
- Performance expectations of ELs generally should not be greater than those of English proficient or “English only” students of comparable age. A key reference for expectations of English proficient or “English only” K–1 students is the state’s grade-level English-language arts standards.
- For writing, particularly in kindergarten, there likely will be more overlap between the proficiency expectations of ELs and English proficient students of the same age. Therefore, skills can be tested even though the skill is introduced in the kindergarten curriculum (e.g., writing the letters of the alphabet). Some skills may be so fundamental or foundational that it is important to assess them on the CELDT.
- ELD standards assessed in grade two may not be appropriate for the K–1 reading and writing assessments or may be appropriate only for grade one.

- The English language proficiency level at which a skill is assessed should be consistent with the state ELD standards and reflect appropriate levels of complexity and rigor.

Panel members then assisted in the development of reading and writing item specifications. The final set of specifications for the tests is shown in Table A.2 of the appendix. Twelve reading and writing ELD standards were identified as important and assessable for K–1.

The CDE then contracted with WestEd for the development of an initial pool of reading and writing items that aligned with the early literacy blueprint. A total of 142 new items were delivered to the CDE in early 2009. English-language development experts, content and bias panels, editors, and CDE staff reviewed these items to ensure that they met the specifications and were appropriate for inclusion on the test. (Appendix B of each CELDT annual technical report provides documentation of experts who contributed to the development of that edition.)

Prior to operational administration of the early literacy assessment, the items were field-tested to check that their statistical properties were appropriate. This field-test took place in mid-March 2009. Four test forms, each consisting of 10 reading and 10 writing items, were field tested at 57 different California schools representing 35 districts throughout the state.

The total sample size for the field-test study was 2,548, and the numbers of students taking each of the four forms were approximately equal. Each school administered one form to approximately 25 kindergarten students and 25 grade one students.

The process used to create the early literacy assessment follows the same path used with the CELDT in general and is shown in Figure 2.1. This process is further documented in the *CELDT 2010–11 Edition Technical Report*, chapter 3 (pp. 25–30). The approach carefully follows professional testing standards to ensure the validity of the CELDT.

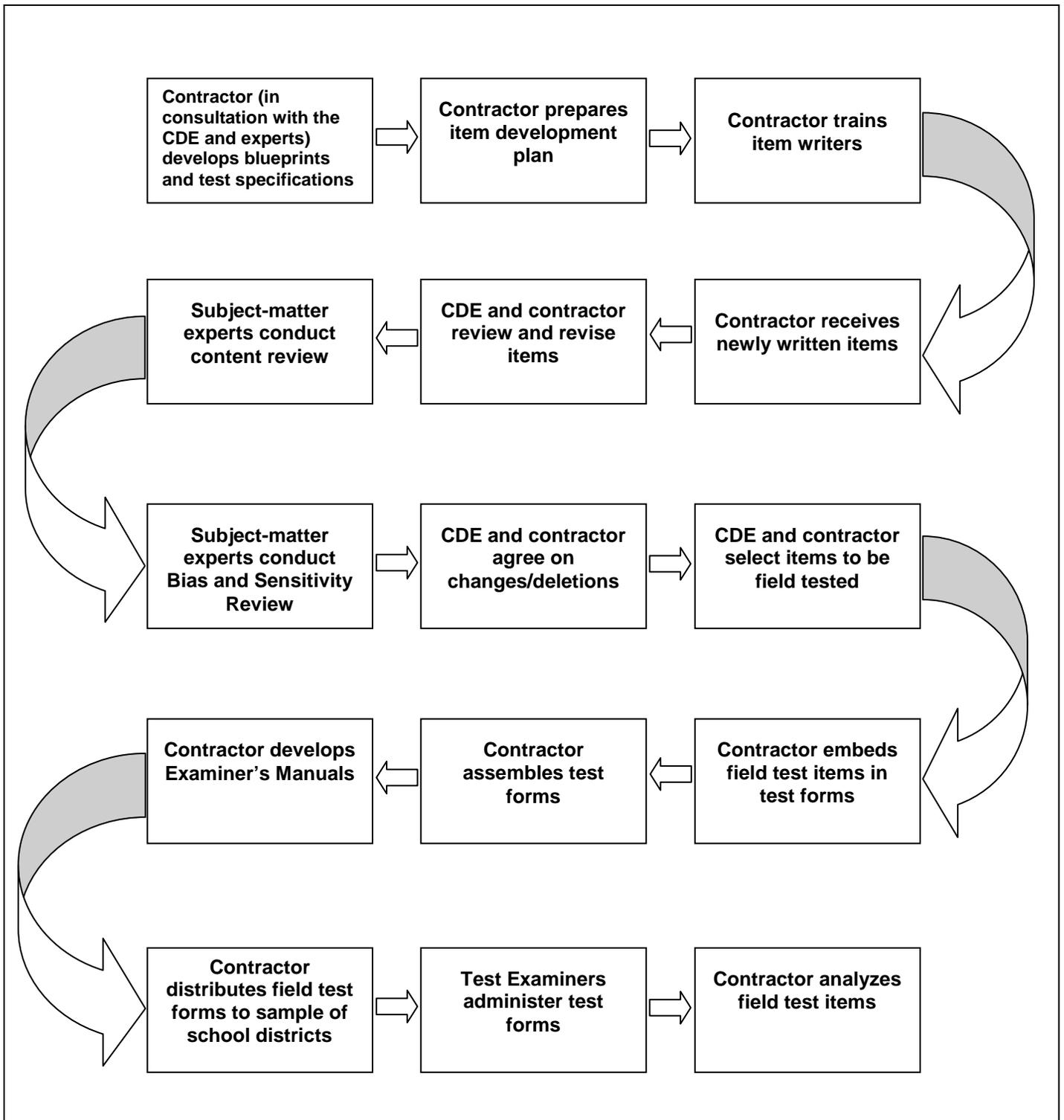


Fig. 2.1. Early Literacy Assessment Development Path

The first operational early literacy assessment was administered to K–1 students beginning July 1, 2009. Data from that administration were used to link the new K–1 reading and writing scales to the common scales that are used to report all other CELDT results. Scores on these scales range from approximately 200 to 600 and permit comparisons of student performance across grades within each domain.

The CDE also initiated a process to identify five levels of performance on the test that correspond to the proficiency levels in the 1999 California ELD standards: Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced. Each successive level represents increasing mastery of English. The K–1 reading and writing test performance descriptors are shown in Table 2.1.

Table 2.1. CELDT K–1 Reading and Writing Test Performance Descriptors

| Performance Level | Descriptor |
|---------------------------|---|
| Reading | |
| Advanced | Students at this level of English language performance typically name all uppercase and lowercase letters of the alphabet; recognize all phonemes that are specific to the English language; identify regular letter-sound correspondences and use them to read one- and two-syllable words; recognize most high-frequency words; including some irregular words; use decoding skills to achieve grade-level appropriate, independent reading; and demonstrate comprehension of most grade-level-appropriate text on both familiar and unfamiliar topics. |
| Early Advanced | Students at this level of English language performance typically name all uppercase and lowercase letters of the alphabet; recognize most phonemes that are specific to the English language; identify regular letter-sound correspondences and use them to read one- and some two-syllable words; recognize many one-syllable and high-frequency words; apply decoding skills to read most one- and two-syllable words; and demonstrate reading comprehension of some grade-level-appropriate text on familiar topics. |
| Intermediate | Students at this level of English language performance typically name most uppercase and lowercase letters of the alphabet, recognize some phonemes that are specific to the English language; identify letter-sound correspondence for initial and some final consonants; recognize some one-syllable high-frequency words; apply basic knowledge of English morphemes, phonics, and syntax to decode one-syllable words; and demonstrate reading comprehension limited to short, common words. |
| Early Intermediate | Students at this level of English language performance typically name some uppercase and lowercase letters of the alphabet; recognize a few English phonemes that do not correspond to phonemes in primary language; identify letter-sound correspondence for some initial consonants; recognize a few simple, one-syllable high-frequency words; and apply basic knowledge of English morphemes, phonics, and syntax to accomplish decoding tasks such as identifying initial sounds. |

| Performance Level | Descriptor |
|---------------------------|---|
| Reading | |
| Beginning | Students at this level of English language performance have little or no receptive skills and may recognize basic concepts of print (e.g., following words left to right, top to bottom; title), name a few uppercase and lowercase letters of the alphabet, recognize English phonemes that correspond to phonemes in primary language, and rely on graphic support to read high-frequency words. |
| Writing | |
| Advanced | Students at this level of English language performance typically copy words, including lowercase and uppercase letters; capitalize proper nouns and the first word of a sentence; place periods and question marks appropriately at the end of simple sentences; write two syllable-words; and write a word based on a story read out loud. |
| Early Advanced | Students at this level of English language performance typically copy most words with lowercase and uppercase letters, although letter reversals may occur; capitalize some proper nouns; use some ending punctuation; write high-frequency one-syllable words; and write a letter or sound based on a story read out loud. |
| Intermediate | Students at this level of English language performance typically copy most letters of the alphabet and some words legibly and write some one-syllable words correctly or use phonetic spelling. |
| Early Intermediate | Students at this level of English language performance typically copy some letters of the alphabet legibly and write first or last letter when attempting to write a word. |
| Beginning | Students at this level of English language performance may demonstrate no productive skills, may scribble when attempting to copy letters of the alphabet, or may write incomprehensible symbols when producing written language. |

These test performance descriptors are provided on the back of each student's performance level report to aid in interpreting the scale score and performance levels attained by the student.

A standard setting panel of experts met to identify cut scores that would differentiate the performance levels for the K–1 reading and writing domains. Panelists were recruited from across the State of California and were selected on the basis of their expertise in English-language development, their experience in the field of education, and their knowledge of the CELDT. The panel of 15 California educators met in Sacramento on January 13, 2010. Results of the standard setting are summarized in Table 2.2 as a set of cut scores on the reading and writing common scales.

Table 2.2. CELDT Cut Scores

| Grade | Performance Level | Scale Scores | | | |
|-------|--------------------|--------------|---------|--------------------|---------|
| | | Reading | Writing | Compre- hension | Overall |
| K | Early Intermediate | 282 | 341 | 322 | 352 |
| | Intermediate | 319 | 371 | 364 | 400 |
| | Early Advanced | 377 | 398 | 416 | 449 |
| | Advanced | 446 | 427 | 474 | 498 |
| 1 | Early Intermediate | 360 | 393 | 361 | 359 |
| | Intermediate | 398 | 409 | 403 | 406 |
| | Early Advanced | 446 | 435 | 450 | 454 |
| | Advanced | 570 | 475 | 536 | 507 |

The cut scores for comprehension were derived by calculating the average of the cut scores for listening and reading. The overall cut scores were derived by calculating the weighted average of the cut scores of the four domains (.45 Listening + .45 Speaking + .05 Reading + .05 Writing). As Table 2.2 shows, the minimum scores required for entry in each performance level are lower for kindergarten students because the expectations for performance in reading and writing increase at each grade. Chapter 6 of the *CELDT 2010–11 Edition Technical Report* (pp. 43–50) contains additional details regarding performance standards.

Subsequent operational administrations of the early literacy assessments occurred in 2010–11 and 2011–12. New K–1 reading and writing items were developed and field-tested in each of the first three operational years of the early literacy assessment.

2.2 Types of questions included in the assessment

The K–1 reading domain assesses students’ skills required to process information presented in written materials in English. Test items require students to recognize English phonemes; name upper- and lowercase letters of the alphabet; recognize sound/symbol relationships; read simple words and phrases; and identify basic text features, such as book titles. Many of the items use graphics to ensure that the reading load is not beyond what K–1 students can reasonably handle.

The K–1 writing domain assesses students’ skills in written language. Test items require students to copy upper- and lowercase letters and commonly used words; write words in response to examiner directions or stories read aloud; identify correct use of capital letters; and identify correct sentence-ending punctuation.

A few sample questions from the test may help illustrate these tasks. Figure 2.2 presents four examples, two from reading and two from writing.

Reading Sample 1

Point to the picture that begins with the sound /g/.



A



B



C

Reading Sample 2

The boy kicks.



A



B



C

Writing Sample 1

Write the Word



Writing Sample 2

When will we eat

•

A

,

B

?

C

Fig. 2.2. Sample Questions from the Early Literacy Assessment

The first sample reading question tests students' abilities to recognize English phonemes by asking students to choose the picture of an object that begins with

a “g” sound. The direction is read aloud, and the test examiner names the pictured objects (hat, bed, goat). Students must correctly identify goat as starting with the “g” sound.

In the second sample reading question, students are tested on their ability to read a simple sentence and match it with the picture that correctly conveys the meaning of the sentence.

The first sample writing question shows students a picture of an object (in this case, a can). Students are told what the object is and asked to write the word in the space provided. Unlike the other examples shown here, which are multiple-choice items, this is a constructed-response task—students must write the word. Those who write the entire word legibly receive two points for the question. Students who write only part of the word correctly receive one point, and students who write or draw illegibly receive no points.

To correctly answer the second sample writing question, students must select the correct punctuation mark from the three provided (period, comma, question mark). The test examiner reads the sentence aloud and asks the students to point to the mark that goes at the end of the sentence. Items such as this test students’ knowledge of standard English conventions.

The technical blueprint for the early literacy assessment (Table A.2 in the appendix) shows the reading domain consists of 20 items, which represent seven ELD standards. The writing domain consists of 20 items, which represent five ELD standards. A total of 194 reading and 188 writing items have been developed over a three-year period for use on the K–1 CELDT.

Additional details regarding the content and structure of the early literacy assessment can be found in chapter 2 of the *CELDT 2010–11 Edition Technical Report* (pp. 13–23) and in the *Released Test Questions*, both of which are located on the CDE Resources Web page at <http://www.cde.ca.gov/ta/tg/el/resources.asp>.

3. Test Administration Process

3.1 Selection of students to take the K–1 early literacy assessment

Districts administer a home language survey when students first enroll in the California school system. The survey asks (1) what language the child first used when learning to speak; (2) what language the child most frequently uses at home; (3) what language the parents or guardians use when speaking to the child; and (4) what language is most frequently spoken by adults in the home. If the answer to any of the first three questions is a language other than English, students are required to take the CELDT. Test results for these students are described as initial assessment (IA) data. Based on the test results and other information available to the district, a student may be classified as an EL or as initially fluent English proficient.

Students who previously have been identified as an EL, based on a prior CELDT administration, must take the CELDT once each year between July 1 and October 31 during the annual assessment window until they are reclassified as fluent English proficient. Test results for these students are described as annual assessment (AA) data.

3.2 Test administration

The CELDT is a written and oral test. Trained test examiners administer the early literacy assessment individually, although some sections may be given in small groups to grade one students. While the test is untimed, administering the four domains requires about 65 minutes.

Written CELDT procedures exist for all phases of the testing process to ensure that tests are administered in a fair and standardized manner throughout the state. The procedures are incorporated into manuals designed for specific roles.

Test examiners must complete training in the current administration of the CELDT before administering the test and must follow the directions prescribed in the *Examiner's Manuals*. Training is provided both live at workshops presented annually throughout the state and via online presentations that are available at any time. Administrative adjustments are allowed to accommodate individual needs if specified in the student's individualized educational program (IEP) or Section 504 plan.

Chapter 5 of the *CELDT 2010–11 Edition Technical Report* (pp. 33–42) provides additional details about test administration and information about procedures used to ensure security and standardization throughout all phases of the test administration process.

3.3 Scoring and reporting procedures

The students' answer documents are scanned, and the multiple-choice items are machine scored. Constructed-response items that require a written response are scored by individuals who are carefully selected, trained, and then continuously monitored throughout the scoring process. A number of quality control procedures are in place to ensure the accuracy of the resulting scores.

CELDT reports communicate results to teachers, parents, and administrators, thereby providing information needed to guide student learning and evaluate instructional programs. Four types of student score reports present test results by grade, school, district, and state.

In July 2010, the SBE modified the definition of English proficient for K–1 on the CELDT to be Early Advanced or Advanced overall and Intermediate or higher on listening and speaking only. Further, the overall score is based on a weighted design that emphasizes listening and speaking at 45 percent each and minimizes reading and writing at 5 percent each, based on the advice of technical, linguistic, and early childhood experts.

For additional details of the scoring and reporting procedures, refer to Chapter 7 of the *CELDT 2010–11 Edition Technical Report* (pp. 51–61), which is located on the CDE Resources Web page at <http://www.cde.ca.gov/ta/tg/el/resources.asp>.

4. Test Reliability and Validity

Chapter 8 of the *CELDT 2010–11 Edition Technical Report* (pp. 63–77) provides significant details regarding psychometric properties of the assessment. The following sections present an overview of the more detailed information available in that report.

4.1 Reliability of the early literacy assessment

Reliability is a measure of the extent to which the test score remains stable or consistent under some change of conditions, such as over time or when a different set of items (i.e., test questions) is used. The latter is particularly important in large-scale testing because test items change each year to ensure that the items do not become well known.

Test reliability is usually expressed in terms of a statistical coefficient that ranges from 0.00, indicating no consistency, to 1.00, indicating perfect consistency. The reliability coefficients for the early literacy assessment in the three years the early literacy assessment has been administered fell between 0.71 and 0.81, which are typical and reasonable coefficients for assessments of these lengths in these grades. Because the difficulties of the items generally target the middle of the proficiency range, scores at the upper and lower ends of the scale are somewhat less reliable than those from the center of the score distribution.

Test reliability is a characteristic of the test score. An important characteristic related to, but distinct from, test reliability is the reliability of **classifications** made on the basis of test scores. In this sense, reliability is the extent to which the test's classification of students into performance levels agrees with the students' **true** classifications, which are theoretical values that cannot be known for certain but can be estimated. It is important to evaluate classification reliability because decisions—for instance, whether to classify a student initially as EL or to reclassify an EL student as English proficient—are based on the performance levels of the test scores, not just on the scores. At the Early Intermediate–Intermediate cut point, which is the critical one for CELDT classification decisions, the reliability coefficients range from 0.67 to 0.94, which again are reasonable values for tests of this length.

4.2 Validity of the assessment

In terms of tests such as the early literacy assessment, validity refers to the extent to which a test's content is representative of the actual skills learned and whether the test can, therefore, allow accurate conclusions to be made concerning achievement.

Test validation is an ongoing process 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.

In addition to multi-level reviews of the items and test forms by internal and external panels of content and measurement experts, further checks ensure that the early literacy assessment remains valid. The assessment must conform to precise psychometric criteria: item difficulties must represent the full range of ability so that all students tested are able to demonstrate what they know and can do. Additionally, because the test items change each year the test is given, each new edition must be equated to ensure that scores retain their same meaning. For example, a score at the Advanced level must represent the same challenge for students in one year as in the next. The equating process is carefully executed and verified to provide this assurance.

Test scores' validity can be compromised if the test is administered in a non-standardized way to students. Therefore, CELDT training is provided to test administrators in multiple ways. *Examiner's Manuals* carefully outline the testing process and contain scripts for examiners to follow during test administration.

Additionally, test administration and scoring training is provided at in-person workshops. At least one person from each school district or charter school that administers the CELDT is required to attend a workshop in which participants receive training in how to (a) standardize administration; (b) reliably score the speaking and writing items that require individual judgment; and (c) qualify other persons in the district to administer and score the CELDT. These training sessions are offered each year at a variety of locations throughout the state to provide convenient venues for all districts.

To further ensure standardization, general test administration training is delivered online through presentations that cover a range of topics relevant to CELDT administration, including test material ordering, "what's new" for a given edition, general testing information, and information related to maintaining test security throughout the process. These presentations are recorded and available online for viewing throughout the administration year.

4.3 Empirical evidence for the validity of the assessment

In the fall of 2010, a study was conducted to compare the performance of K–1 English fluent (EF) students and EL students on the CELDT. EF students are those who, based on the results of a home language survey, either were not administered the CELDT in 2009 or 2010 or were designated Initial Fluent English Proficient in kindergarten, based on a previous CELDT. A total of 1,386 kindergarten and 495 grade one EF students from 100 schools were administered the 2010–11 edition of the CELDT. Their performance on the CELDT was compared to that of EL students in the same schools who took the test at the same time. For details of the "California English Language Development Test—A Comparison Study of Kindergarten and Grade One English-Fluent Students and English Learners on the 2010–11 Edition of the CELDT," consult the report available on the CDE CELDT Technical Documentation Web page at <http://www.cde.ca.gov/ta/tg/el/techreport.asp>.

The data in Table 4.1 show that the CELDT validly differentiates EF and EL students in practically significant ways.

Table 4.1. Average Scale Score Comparisons of EL Students and EF Students

| Grade | Group | Group Size | Listening | Speaking | Reading | Writing | Comprehension | Overall |
|-------|-------|------------|-----------|----------|---------|---------|---------------|---------|
| K | EL | 4,350 | 330 | 317 | 285 | 332 | 307 | 322 |
| | EF | 1,386 | 417 | 452 | 318 | 357 | 367 | 424 |
| 1 | EL | 3,985 | 394 | 393 | 365 | 393 | 380 | 392 |
| | EF | 495 | 449 | 488 | 405 | 410 | 427 | 462 |

The difference in the overall scale score (based on the weights of 45 percent each listening and speaking and 5 percent each reading and writing) shows that the average kindergarten EL student scores in the Beginning performance level while the average kindergarten EF student scores in the Intermediate level. The difference in grade one shows that the average EL student scores at the Early Intermediate level while the average EF student scores at the Early Advanced level. (See Table 2.2.)

The listening and speaking domains differentiate the two groups more sharply than do reading and writing (the early literacy assessment). The differences are roughly twice as large for kindergarten students as for grade one students. The largest differences occur in speaking, where kindergarten EF students score 135 scale score points higher than kindergarten EL students. The smallest differences occur in writing, where grade one EF students score 17 points higher than grade one EL students.

Although the differences between the two groups are larger for listening and speaking than for reading and writing, reading and writing differences between the two groups are nonetheless significant and support the validity of the assessment. Such a finding is not unexpected because kindergarten students, in particular, are unlikely to have had very much instruction in reading and writing before coming to school, regardless of their language background. This supports the notion of differentially weighting the four domains at the K–1 level.

4.4 Teachers' and English learner administrators' views regarding the usefulness of the early literacy assessment

Many teachers, administrators, and others involved in EL instruction have worked with the CELDT for more than a decade. Because of their familiarity with the CELDT, their input regarding the utility of the CELDT was sought to inform the content of this report.

CELDT District Coordinators received an e-mail asking them to request the participation of district staff, teachers, and EL administrators in a brief online survey. The survey link was to be forwarded to those who had experience using CELDT scores for making decisions for K–1 students. More than 1,000 people responded during the two weeks that the survey was operational. Half of the respondents identified themselves as EL

teachers, 30 percent as EL coordinators, 27 percent as administrators, and 13 percent as “Other.” Because respondents were able to select more than one option, the sum is more than 100 percent.

Respondents were asked how helpful they found the test results for making local educational decisions. Specifically, they were asked (a) how helpful the results are for determining initial English proficiency; (b) how accurate they are for identifying English performance levels; and (c) how helpful they are for making instructional decisions.

Respondents were asked to rate these three questions for both reading and writing on a four point scale: “Very helpful,” “Somewhat helpful,” “Somewhat unhelpful,” and “Unhelpful.” The “Very helpful” and “Somewhat helpful” ratings were added together to identify a “Helpful” rating.

For both grade levels, survey respondents found the listening and speaking results to be more helpful than the reading and writing results. However, for kindergarten, where reading and writing skills are usually minimal at the time of entry, roughly 40 percent of respondents found results of the early literacy assessment to be helpful for initial classification and instructional decisions.

Perhaps of greater interest is that helpfulness increased to about 60 percent when the decision was to consider the test’s contribution to instructional decisions for grade one.

Figures 4.1 and 4.2 present these results graphically. The helpfulness of the oral domains (listening and speaking) was about the same for both grades.

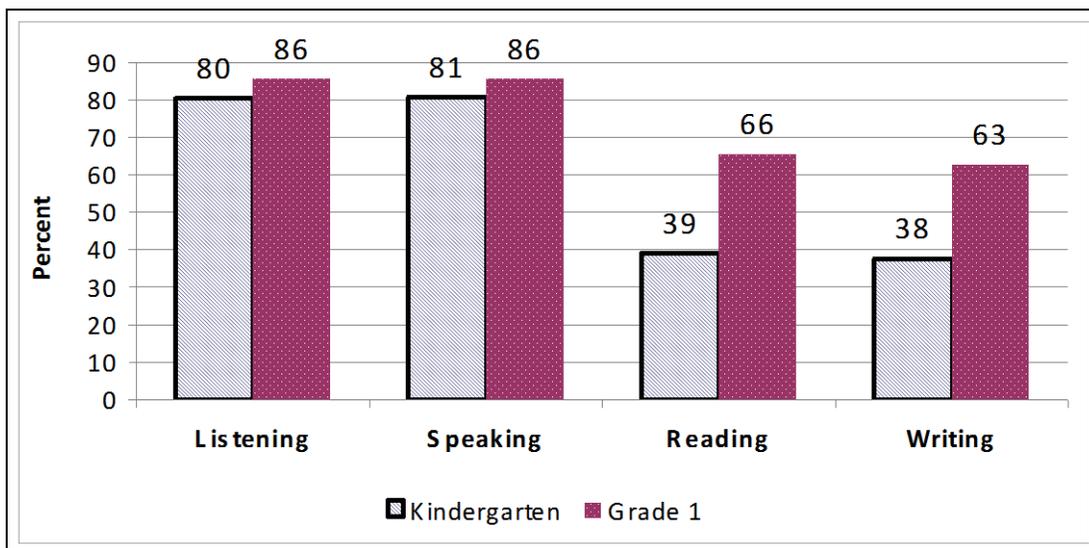


Fig. 4.1. Percentage of Survey Respondents Rating Each Domain of the CELDT as Useful for Initial Classification

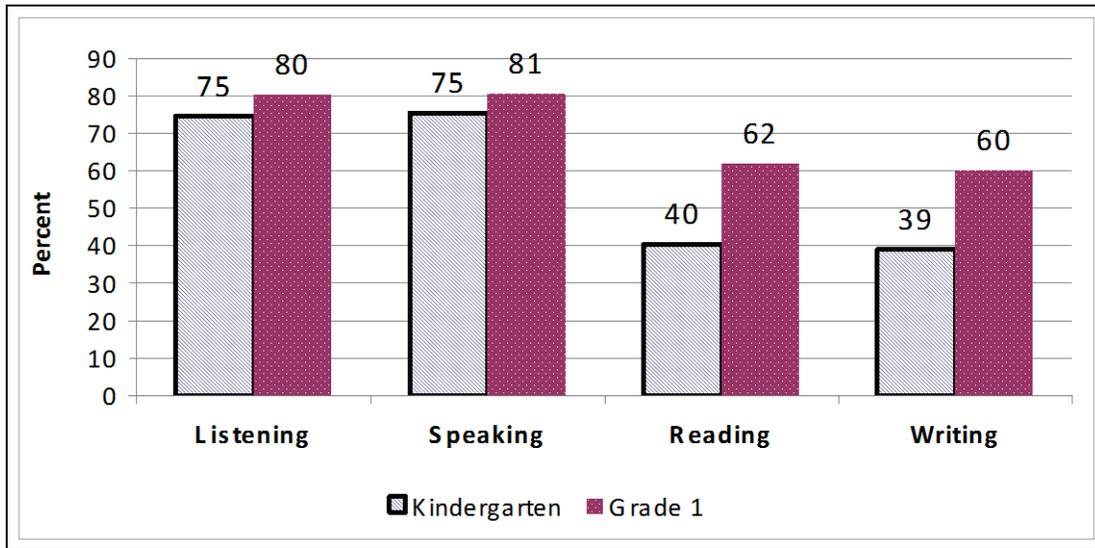


Fig. 4.2. Percentage of Survey Respondents Rating Each Domain of the CELDT as Useful for Instructional Decisions

Respondents also had an opportunity to write their observations or suggestions. The most frequent responses dealt with recommendations to change the CELDT administration time from the fall to the spring. Respondents pointed out that students, particularly K–1 students, were still adjusting to the classroom at the time of the regular fall assessment. Respondents also felt the test was too long and the reading and writing content too advanced for K–1 students and that the test measured literacy, not English-language development. Nearly one-fifth of the open-ended responses to the survey indicated that the current reading and writing items were not appropriate for kindergarten students in particular, and they suggested a shorter, simpler test for kindergartners.

5. Results for Three Administrations of the Early Literacy Assessment

This section presents an overview of the first three years of the administration of the early literacy assessment. The number of students tested, average scale scores, and performance levels are shown for each group of students.

5.1 Demographic characteristics of the K–1 students assessed

Table 5.1 shows the total number of kindergarten and grade one students tested each year. The “Initial Assessment” column shows the number of tests administered for initial evaluation. The “Annual Assessment” column shows the number of tests administered during the July 1–October 31 test window to students already identified as ELs. The “Other” category includes students tested for purposes of annual assessment, but testing took place outside the designated window; it also includes students whose testing purpose was not indicated.

As the table shows, the number of K–1 students tested has been slightly more than 400,000 each year, with a slight decline from year to year in the number tested. The table also shows that a small number of kindergarten students, about 5,000 each year, have been retained and, thus, appear in the Annual Assessment (AA) column. However, the majority of kindergarten students are in the Initial Assessment group.

Table 5.1. Number of K–1 Students Tested

| Year | Grade | Initial Assessment | Annual Assessment | Other | Total |
|---------|-------|--------------------|-------------------|-------|---------|
| 2009–10 | K | 206,888 | 5,374 | 216 | 212,478 |
| | 1 | 19,311 | 176,848 | 1,833 | 197,992 |
| 2010–11 | K | 204,359 | 4,992 | 340 | 209,691 |
| | 1 | 18,831 | 176,263 | 1,589 | 196,683 |
| 2011–12 | K | 205,738 | 5,293 | 238 | 211,269 |
| | 1 | 16,140 | 178,350 | 1,453 | 195,943 |

Table 5.2 presents information regarding the number of K–1 students tested with the CELDT who were in migrant education and gifted and talented programs. It also provides information on the types of instructional programs in which these students were enrolled. This table uses data for the 2011–12 AA population only, but the numbers are not dramatically different across the years for the same population.

Table 5.2. Program Participation of Students Tested

| Program | Number of Students | Percent |
|--|---------------------------|----------------|
| Migrant Education | 5,693 | 3.1 |
| Gifted and Talented Program Participation | 59 | 0.0 |
| English-Language Development and Specially Designed Academic Instruction in English | 98,468 | 53.6 |
| English-Language Development and Specially Designed Academic Instruction in English with Primary Language Support | 47,089 | 25.6 |
| English-Language Development | 15,381 | 8.4 |
| English-Language Development and Academic Subjects through Primary Language | 13,578 | 7.4 |

There are 58 different languages reported on home language surveys, with Spanish reported by just over 80 percent; Vietnamese reported by 3 percent; Cantonese and Mandarin together reported by 2.8 percent; Filipino reported by 1.4 percent; and Korean reported by 1 percent of the students tested. Over 2,600 students, or 1.5 percent, reported languages falling into the category “All Other Non-English Languages.”

5.2 Student performance

The score distributions for kindergarten and grade one students are very different. Figures 5.1 and 5.2 show the reading and writing scale score distributions for the 2011–12 AA test populations separately, by grade.

Instead of a single, bell-shaped distribution that would be expected of a relatively homogeneous population, the distributions essentially are bimodal. Scores for the kindergarten students tend to bunch at the lower ends of the scales while scores for grade one students tend to bunch nearer the upper ends of the scales.

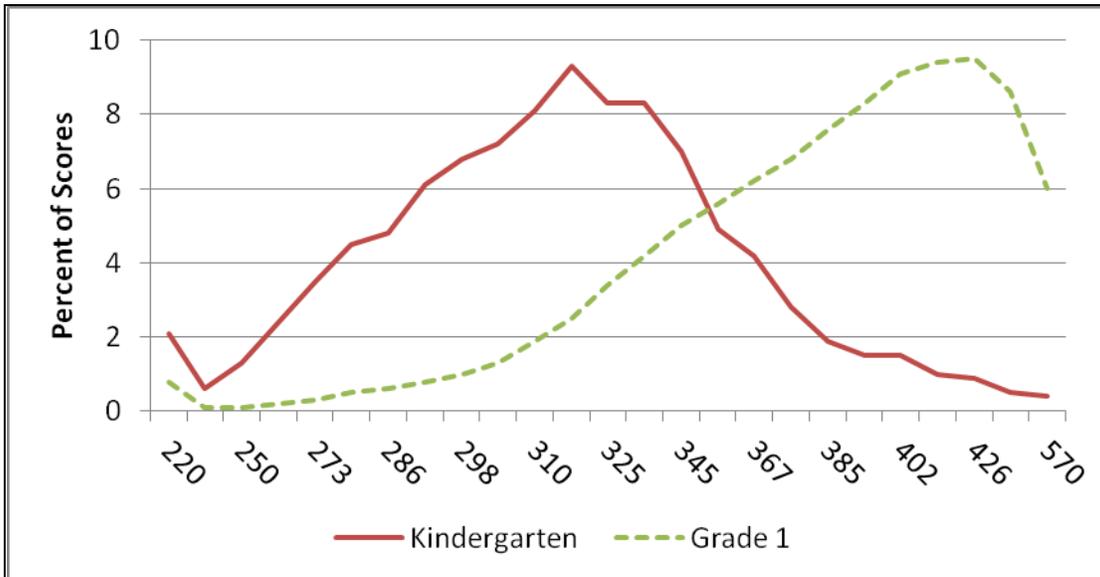


Fig. 5.1. Reading Scale Score Distribution

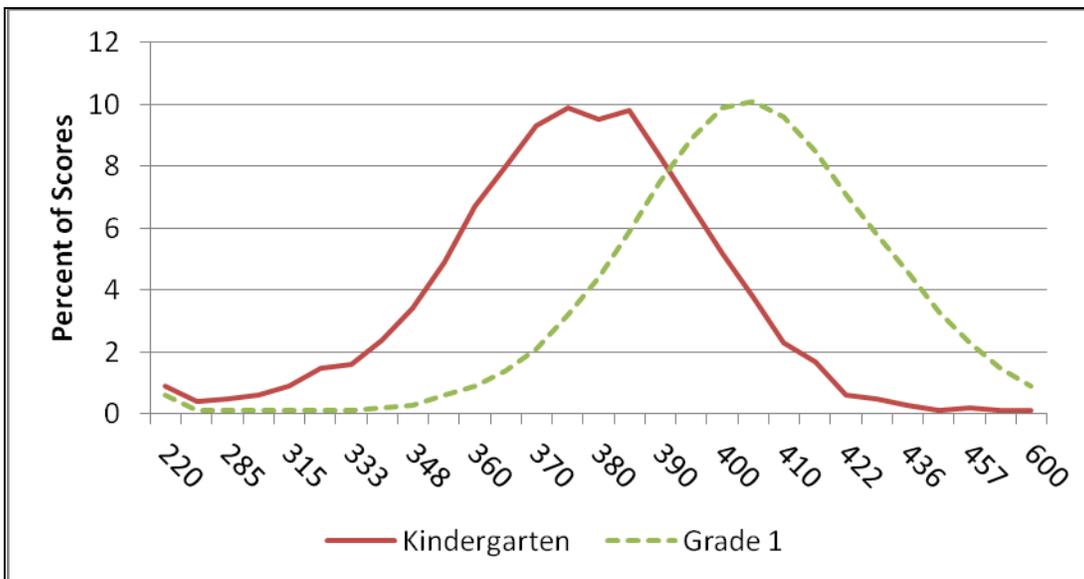


Fig. 5.2. Writing Scale Score Distribution

Table 5.3 presents the average scores of students tested in each year of administration. These results are based only on annual assessment scores or scores of students who have taken the CELDT previously. Each row of the table is based on a different group of students, which needs to be considered in comparing results across years. The differences from one year to the next are generally small and indicate only that the

groups are roughly equal in proficiency. Chapter 6 presents a different perspective by showing results when the same students are followed from one year to the next.

Table 5.3. Annual Assessment CELDT Average Scale Scores

| Grade | Year | N-tested | Scale Scores | | | | | Overall |
|-------|---------|----------|--------------|----------|---------|---------|--------------------|---------|
| | | | Listening | Speaking | Reading | Writing | Compre- hension | |
| K | 2009–10 | 206,888 | 378 | 367 | 316 | 366 | 347 | 372 |
| | 2010–11 | 204,359 | 375 | 371 | 315 | 364 | 345 | 369 |
| | 2011–12 | 205,738 | 380 | 385 | 320 | 375 | 350 | 378 |
| 1 | 2009–10 | 19,311 | 430 | 435 | 393 | 406 | 411 | 432 |
| | 2010–11 | 18,831 | 426 | 436 | 391 | 403 | 408 | 427 |
| | 2011–12 | 16,140 | 431 | 438 | 392 | 406 | 411 | 430 |

Table 5.4 presents the percentages of EL students in each performance level. These results are based only on AA scores (i.e., students who have taken the CELDT previously).

Table 5.4. Percentage of Students in Performance Levels

| Grade | Year | N-tested | Percent of Students: Reading | | | | |
|------------------------------|---------|----------|------------------------------|-----------------------|--------------|-------------------|----------|
| | | | Beginning | Early Intermediate | Intermediate | Early Advanced | Advanced |
| K | 2009–10 | 206,888 | 17.0 | 33.6 | 41.6 | 6.4 | 1.3 |
| | 2010–11 | 204,359 | 20.1 | 40.7 | 32.7 | 5.2 | 1.4 |
| | 2011–12 | 205,738 | 14.1 | 42.7 | 35.9 | 6.5 | 0.8 |
| 1 | 2009–10 | 19,311 | 23.8 | 34.2 | 23.1 | 11.0 | 8.0 |
| | 2010–11 | 18,831 | 28.7 | 31.5 | 20.7 | 10.8 | 8.3 |
| | 2011–12 | 16,140 | 28.3 | 29.0 | 28.0 | 8.6 | 6.0 |
| Percent of Students: Writing | | | | | | | |
| K | 2009–10 | 206,888 | 11.9 | 30.7 | 44.0 | 12.4 | 1.0 |
| | 2010–11 | 204,359 | 13.5 | 41.2 | 35.7 | 8.9 | .8 |
| | 2011–12 | 205,738 | 5.9 | 35.0 | 44.5 | 13.7 | .9 |
| 1 | 2009–10 | 19,311 | 22.2 | 29.4 | 34.5 | 11.3 | 2.6 |
| | 2010–11 | 18,831 | 34.0 | 28.1 | 27.7 | 8.3 | 1.9 |
| | 2011–12 | 16,140 | 27.7 | 28.8 | 31.0 | 10.1 | 2.4 |

The level of skill (i.e., cut-off score) required to qualify at each level increases from grade to grade. For example, a grade one student must score better than a kindergarten student in the reading domain to be considered Advanced.

The bulk of the annual CELDT K–1 scores is concentrated in the lowest three performance levels. This is reasonable considering that the kindergarten students in the annual assessment group, a very small fraction of the total number of kindergarten students taking the CELDT, have been retained for some reason. The grade one students in general have had only one year of instruction by the time the test is administered, but they do much better, particularly in reading.

The results are presented in Table 5.5. The values represent the averages across three CELDT editions.

Table 5.5. Student Performance Level

| Student Performance Level: Reading | | | | | |
|---|--------------------|-----------------------------|-----------------------|-------------------------|-------------------|
| Grade | % Beginning | % Early Intermediate | % Intermediate | % Early Advanced | % Advanced |
| K | 17 | 39 | 37 | 6 | 1 |
| 1 | 27 | 32 | 24 | 10 | 7 |
| Student Performance Level: Writing | | | | | |
| K | 10 | 36 | 41 | 12 | 1 |
| 1 | 28 | 29 | 31 | 10 | 2 |

6. Growth Over Time

The results presented in section 5 compare the performance of groups of students (e.g., comparing the performance of K–1 students tested in 2009–10 with K–1 students tested in 2010–11). Because these data compare different groups of students, they cannot provide any indication of how individual students improve as a result of instruction.

A second set of results may be instructive in considering this question: How much does the typical EL student gain each year he or she receives instruction? The results presented in this section answer that question.

Using the Statewide Student Identifier (SSID), student test scores were matched from one year to the next (i.e., 2009–10 to 2010–11 and 2010–11 to 2011–12). For kindergarten, the match most frequently paired a student’s IA record with the student’s AA record in grade one. At other grade levels, the match would most often pair one AA record with another.

The results of this process are shown in Figures 6.1 through 6.3 for the 2010–11 to 2011–12 match, which are quite similar to those for the 2009–10 to 2010–11 match. The solid, higher line represents the average scores for students in the 2011–12 school year, and the lower, dotted line represents the average scores for students in the 2010–11 school year. Results are shown for reading (Figure 6.1) and writing (Figure 6.2) as well as for the overall scale score (Figure 6.3).

Although this report focuses on the K–1 assessments, results are shown for all grades so the gains in K–1 can be viewed in context.

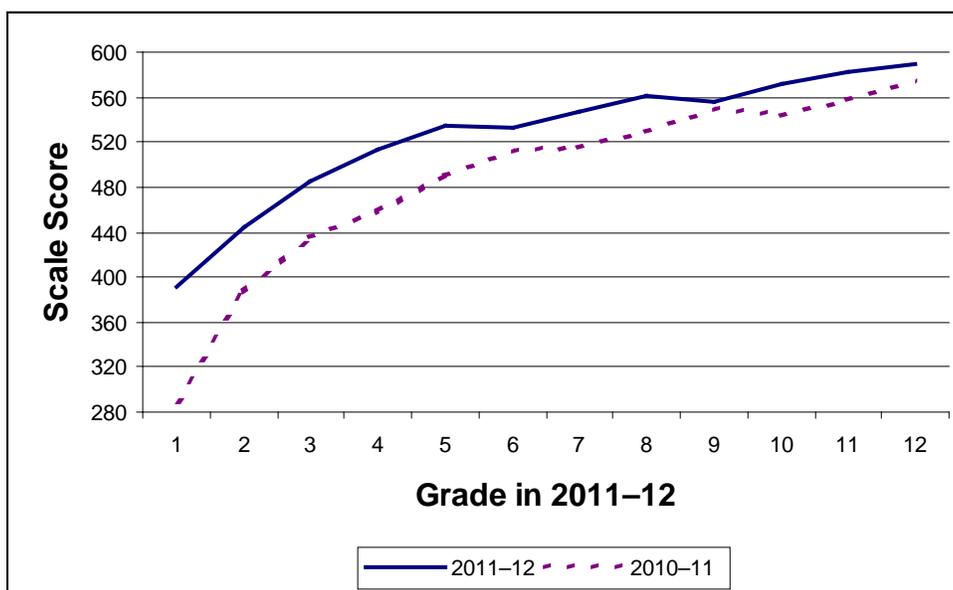


Fig. 6.1. Average Reading Scale Scores for Students Matched from 2010–11 to 2011–12

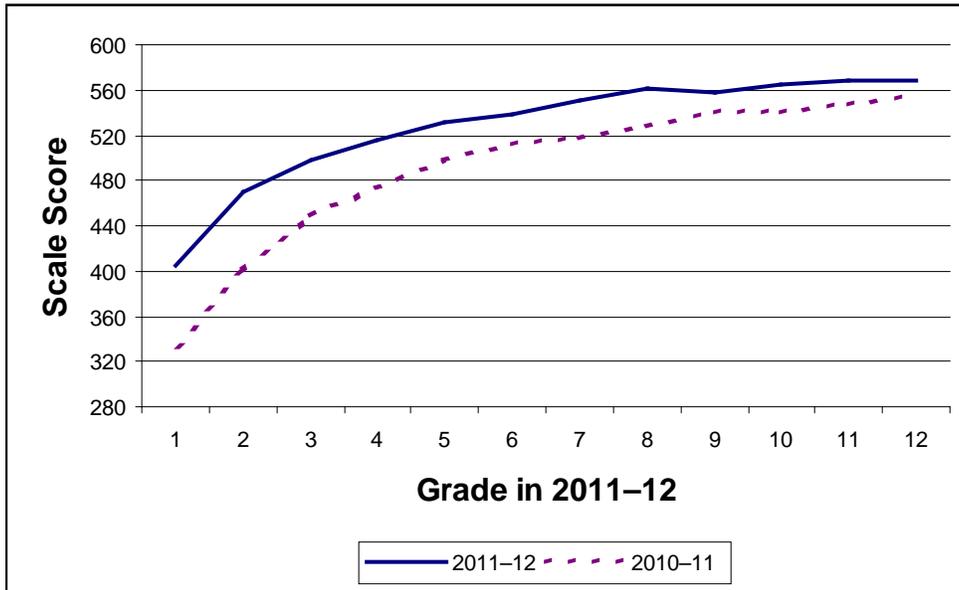


Fig. 6.2. Average Writing Scale Scores for Students Matched from 2010-11 to 2011-12

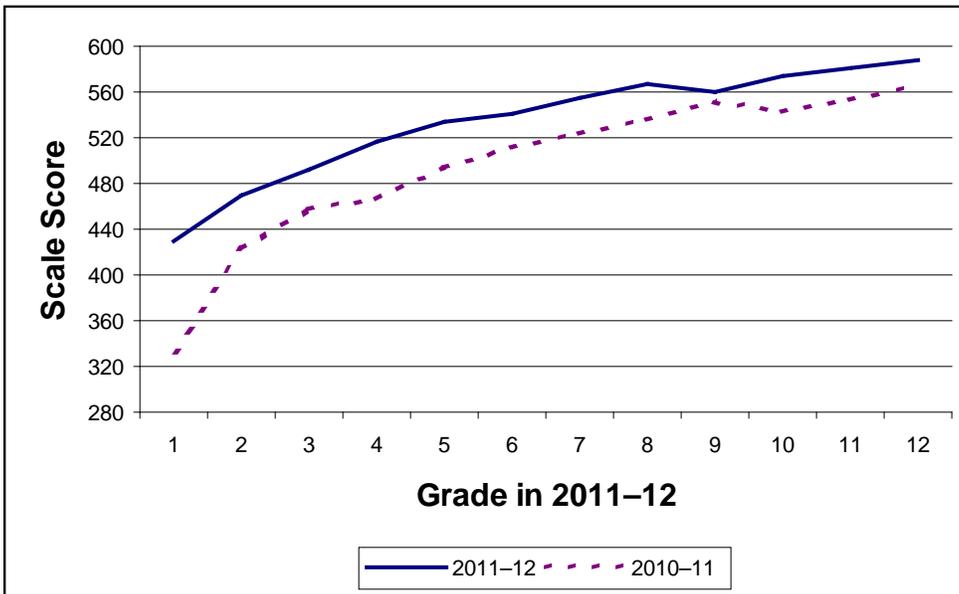


Fig. 6.3. Average Overall Scale Scores for Students Matched from 2010-11 to 2011-12

Table 6.1 summarizes the average scale score gain made by students in all grades. The first column shows the grades matched. The table shows, for example, that students gained an average of approximately 100 scale score points in reading from kindergarten to grade one (98, 101) and approximately 65 scale score points in writing (60, 71) during that same period.

Table 6.1. Average Student Scale Score Increase by Grade

| Grades Matched | 2009–10 to 2010–11 | | | | 2010–11 to 2011–12 | | | |
|----------------|--------------------|---------|--------------------|---------|--------------------|---------|--------------------|---------|
| | Reading | Writing | Compre- hension | Overall | Reading | Writing | Compre- hension | Overall |
| K–1 | 98 | 60 | 88 | 85 | 101 | 71 | 96 | 95 |
| 1–2 | 52 | 51 | 48 | 36 | 57 | 68 | 58 | 46 |
| 2–3 | 39 | 44 | 20 | 24 | 50 | 49 | 42 | 36 |
| 3–4 | 41 | 38 | 42 | 38 | 56 | 42 | 63 | 50 |
| 4–5 | 35 | 32 | 34 | 30 | 44 | 33 | 50 | 40 |
| 5–6 | 18 | 19 | 24 | 23 | 22 | 26 | 31 | 29 |
| 6–7 | 21 | 21 | 25 | 27 | 31 | 32 | 35 | 32 |
| 7–8 | 24 | 19 | 22 | 25 | 32 | 34 | 35 | 31 |
| 8–9 | 1 | 8 | -10 | 1 | 8 | 17 | 4 | 9 |
| 9–10 | 19 | 24 | 11 | 19 | 28 | 23 | 37 | 31 |
| 10–11 | 18 | 22 | 10 | 18 | 25 | 21 | 34 | 29 |
| 11–12 | 13 | 18 | 3 | 13 | 16 | 15 | 26 | 23 |

The scale score gains are largest in the early grades and diminish as the grade levels progress. This finding is consistent with a wide body of research on student achievement which shows that increases in test performance are most dramatic in the early elementary grades.

These same results perhaps can be seen more easily in graphic form. Figure 6.4 shows results for the overall score. Figures 6.5 and 6.6 show results for the reading and writing domains, respectively. In each case, the graphs show mathematically “smoothed” representations of the gains that capture the results very compactly.

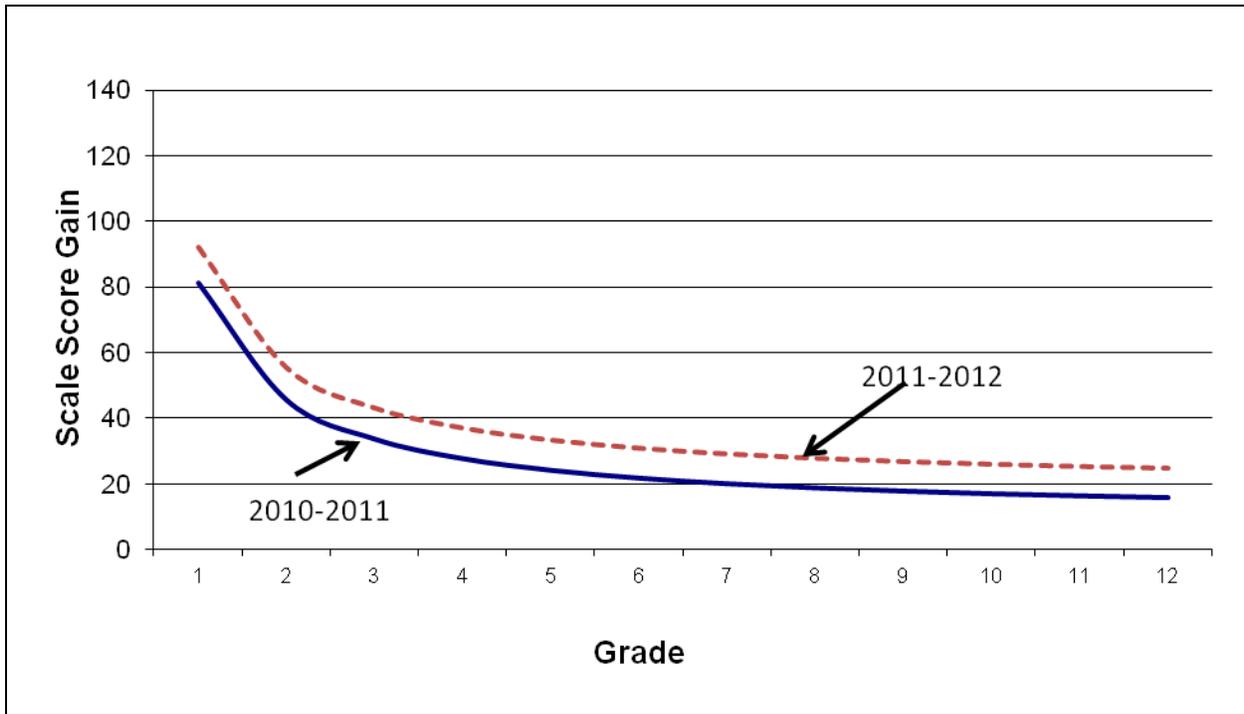


Fig. 6.4. Average Individual Gain in Overall CELDT Scale Score from K to Grade One

Figure 6.4 shows a gain from kindergarten to grade one of approximately 85 points in the overall CELDT scale score. Students gained about 45 points from grade one to grade two. The larger gains to the left of the chart point to the importance and impact of the early elementary grades. Students continue to gain throughout their school years, but gains in these first few grades are typically the most dramatic and, thus, the most critical.

Figures 6.5 and 6.6 show results for reading and writing specifically. In each case, the shape of the curve essentially is the same as that seen in Figure 6.4. The gains in the first two years of instruction are very large. Because the majority of students arrive in kindergarten with very limited reading and writing skills, regardless of their EL status, the gains in these domains are even larger than those seen for the overall assessment.

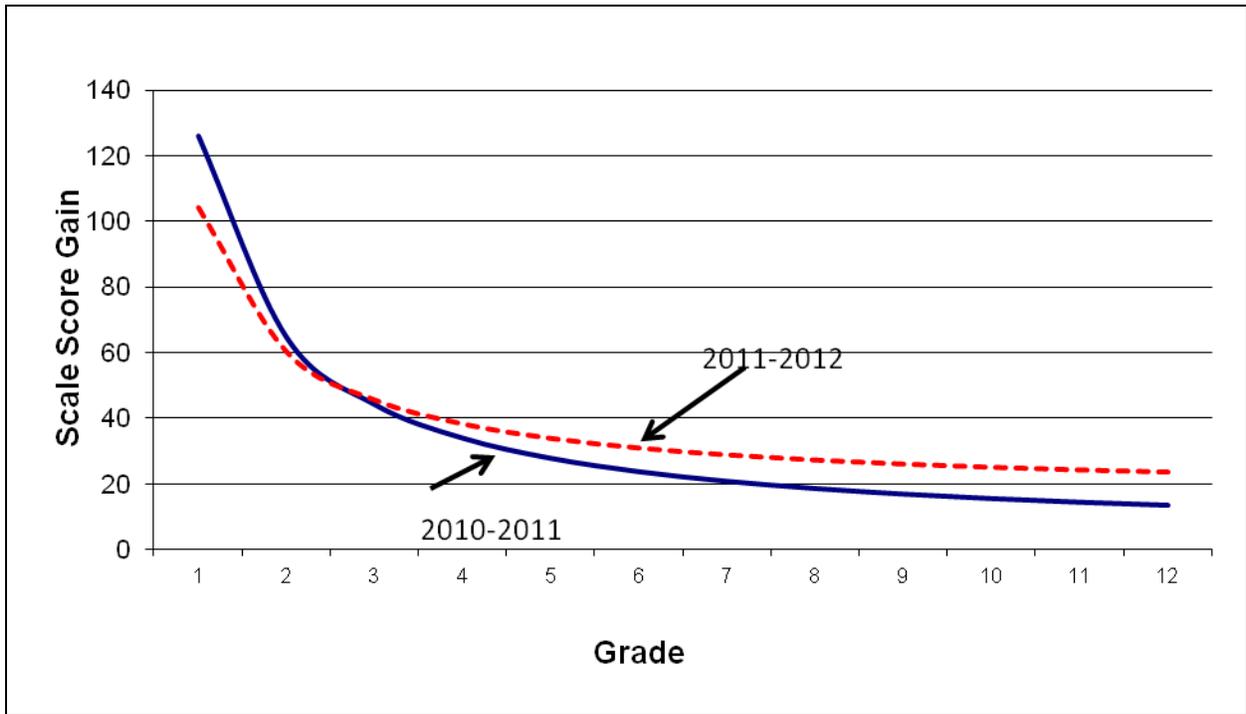


Fig. 6.5. Average Individual Gain in Reading Scale Score from K to Grade One

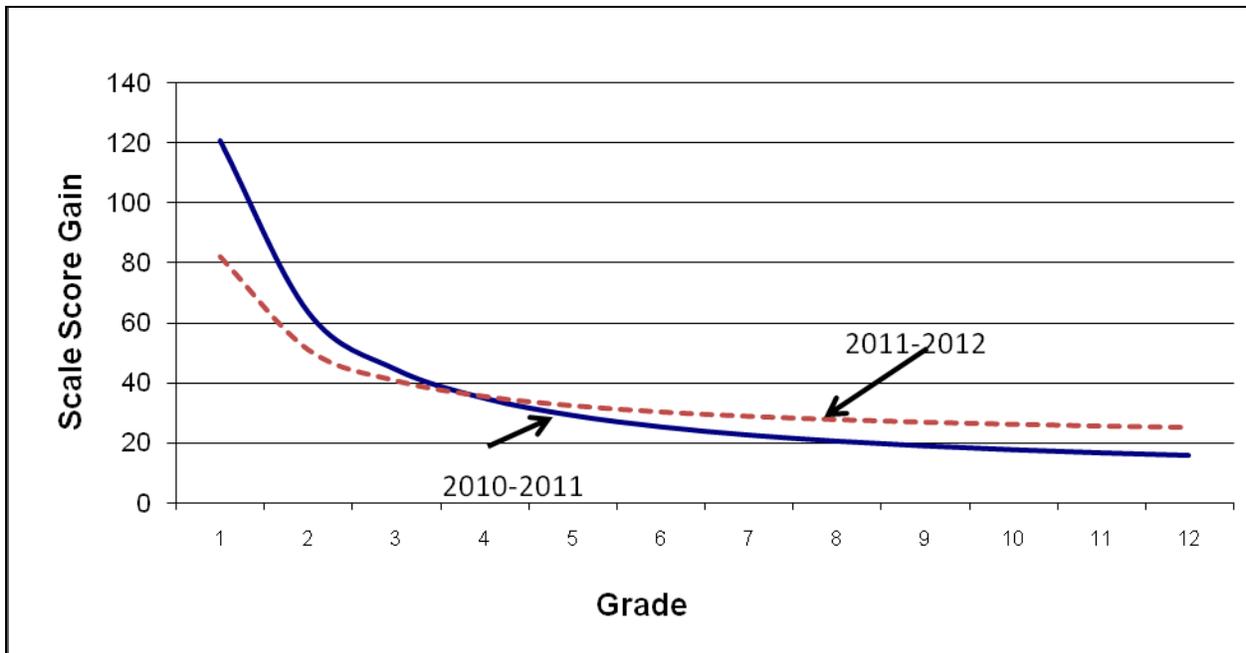


Fig. 6.6. Average Individual Gain in Writing Scale Score from K to Grade One

In addition to differences across domains, there are slight differences across years. However, with only two years' worth of matching data, it is difficult to conclude that these year-to-year differences in gains have great significance. They may only represent random variations on an underlying growth pattern. On the other hand, additional research involving additional years of data may uncover programmatic, generational, or other effects that explain these year-to-year differences.

7. Summary and Recommendations for Future Development

More than one million kindergarten and grade one students took the early literacy assessment over the first three years of its use. Because the CELDT must be administered within 30 days of first enrollment in school, most kindergarten students take the CELDT in August or September. Most grade one students take the test in September or October as part of the annual assessment required of all students identified as English learners.

The data summarized in this report and presented in more detail in the technical reports associated with each of the CELDT editions show that the assessments are reliable and valid. The careful process by which the assessments were developed involved expert judgment throughout each step to ensure that the content was appropriate for the intended purpose. Additional data collected in fall 2010 compared test results of EF students with those of ELs. The differences between the two groups are highly significant, both statistically and educationally.

The domain tests are able to measure the effects of continued instruction in English language skills. Students retested after a year of school show increases in test scores. These gains are evident throughout the K–12 years, but they are greatest in the first few years of instruction—four to six times greater than in later years.

7.1 Recommendation 1: Separate the K–1 tests into distinct grade level exams.

The number of items in each of the current K–1 reading and writing domains is only 20. Consequently, it is difficult to select items that can assess with appropriate rigor the English fluency of students who are at the earliest stages of learning, such as those in kindergarten, and simultaneously, those who have been in school for a year. As Figures 5.1 and 5.2 illustrate, the two distributions of scores are very distinct: scores for kindergarten students tend to bunch at the lower ends of the scales, and scores for grade one students tend to bunch at the top ends of the scales. The differences are most noticeable in reading, where the highest concentration of scores for kindergarten students is about 200 points below that for grade one students.

These data patterns suggest that the items necessary to evaluate grade one students with respect to these skills are too difficult for incoming kindergarten students. Similarly, grade one students largely score correctly on items that are needed to differentiate the skill levels of kindergarten students.

For these reasons, the CDE should consider separating the CELDT into single-grade tests (kindergarten, grade one, and grade two). Although the focus here is on the reading and writing domains of the CELDT, the same argument can be made for the listening and speaking domains because the data are similar.

Should these changes to the tests occur, it would be necessary to convene a new standard-setting panel. The current results are based on a panel that worked with the combined K–1 test forms.

7.2 Recommendation 2: After separating the kindergarten and grade one test, reevaluate the weighting of reading and writing in the overall score calculation.

In the present formulation for K–1 students, reading and writing are given one-ninth the weight of listening and speaking in the calculation of the overall score. The overall score contributes most significantly in weight to classification and reclassification decisions. When the K–1 CELDT is separated into kindergarten and grade one assessments and optimized for use with each population, it would be appropriate to reconsider the weights afforded to reading and writing.

7.3 Recommendation 3: Incorporate common core standards in assessment.

Because California has developed a new set of English -Language Development Standards (ELD) aligned to the California Common Core State Standards (CCSS) for English language arts (ELA), it is necessary to reexamine the CELDT blueprint and the underlying skills currently measured to ensure that they are aligned to the new ELD standards. Item development, which is continuous throughout the life of the test, should be reviewed to confirm that new items created for the test are fully aligned to the new ELD standards and, therefore, to the CCSS for ELA.

Appendix

Table A.1. English Learner Expert Panel Members

Jamal Abedi, Ph.D., Professor
School of Education
University of California at Davis

Alison Bailey, Ed.D., Associate Professor
Psychological Studies in Education
Graduate School of Education and Information Studies
University of California at Los Angeles

Patricia Calabrese, English Learner Coordinator
Robla Elementary School District

Debra Dougherty, Program Manager
Office of Language Acquisition
San Diego Unified School District

Richard Duran, Ph.D., Professor
Gevirtz Graduate School of Education
University of California at Santa Barbara

Margo Gottlieb, Ph.D., Director of Assessment and Evaluation
Illinois Resource Center

Barbara Merino, Ph.D., Professor
Director of Teacher Education
School of Education
University of California at Davis

Edynn Sato, Ph.D., Director, Research and ELL Assessment for Assessment and Standards Development Services and Director, Special Populations for Assessment and Accountability Comprehensive Center
WestEd

Robin Scarcella, Ph.D., Professor
School of Humanities
Director, Program of Academic English/English as a Second Language

Yuuko Uchikoshi Tonkovich, Ed.D., Assistant Professor
School of Education
University of California at Davis

Table A.2. CELDT Blueprint for K–1 Early Literacy Assessment

| | Number of Items | Standard Code | English-Language Development Standard |
|--|-----------------|---------------|---|
| Word Analysis | 4 | 2.12.01 | Recognize English phonemes that correspond to phonemes students already hear and produce in their primary language. |
| | 4 | 2.14.05 | Recognize English phonemes that do not correspond to sounds students already hear and produce (e.g., <i>a</i> as in <i>cat</i> and final consonants). |
| | 4 | 2.17.01 | Recognize and name all uppercase and lowercase letters of the alphabet. |
| | 2 | 2.18.05 | Recognize sound/symbol relationships and basic word-formation rules in phrases, simple sentences, or simple text. |
| Fluency and Systematic Vocabulary Development | 1 | 2.22.09 | Read simple vocabulary, phrases, and sentences independently. |
| | 3 | 2.23.17 | Use decoding skills to read more complex words independently. |
| Comprehension | 2 | 2.35.05 | While reading aloud in a group, point out basic text features, such as the title, table of contents, and chapter headings. |
| | | | |
| Strategies and Applications | 4 | 3.47.01 | Copy the English alphabet legibly. |
| | 4 | 3.48.01 | Copy words posted and commonly used in the classroom (e.g., labels, number names, days of the week). |
| | 4 | 3.49.01 | Write a few words or phrases about an event or character from a story read by the teacher. |
| English Language Conventions | 4 | 3.65.01 | Use capitalization to begin sentences and for proper nouns. |
| | 4 | 3.66.01 | Use a period or question mark at the end of a sentence. |

