LAS Links Third Edition Connecticut Technical Report Form E Contents

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Contents

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

List of Tables

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Purpose and Overview of LAS Links 3rd Edition Technical Manual

The primary purpose of this technical report is to provide detailed evidence of the technical quality, reliability, validity, and fairness of LAS Links 3rd Edition (hereafter referred to as LAS Links 3rd Edition or LAS Links) scores and proficiency information. The evidence supplied in this report is designed, collected, and reported in concordance with the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement (NCME) in *Standards for Educational and Psychological Testing* (2014, hereafter referred to as the Standards). This report is intended to support LAS Links 3rd Edition users such as state departments of education and other organizations that are responsible for meeting the rigorous technical requirements embodied in various technical review processes, including reviews by technical advisory committees and U.S. Federal Peer Review.

Therefore, the organization of this technical report is predicated on descriptions of the intended uses of LAS Links test scores, and detailed descriptions of the reliability, validity, and fairness evidence collected are presented as evidence in support of those intended uses. This information is generally organized by categorizing the sources of validity evidence under the five sources described in the Standards: a) evidence related to the test content, b) evidence related to internal structure, c) evidence related to response processes, d) evidence related to the relationship of LAS Links scores with other variables, and e) evidence related to the consequences of test score use. Specifically, information related to test content including universal design procedures and attention to fairness, test construction, response processes, field test designs and student samples, test administration, scoring processes and quality, psychometric properties of the test scores and the relationship among domain scores, standard setting and validation, and score reporting are presented in this report.

Chapter 1 lays the foundation for how the reliability, validity, and fairness of LAS Links test scores will be demonstrated relative to their intended uses. This chapter discusses the history of LAS Links 3rd Edition development, articulates the intended uses of LAS Links scores, and provides the context for the reliability, validity, and fairness evidence presented in subsequent chapters.

Chapter 2 provides detailed evidence in support of LAS Links content validity, with specific attention to fairness in terms of the application of universal design and through bias review, and evidence of alignment to multiple frameworks and standards. This chapter describes the design and development of the tests, including information about the test standards, test blueprints, item writing, item review process (with and without data), form assembly, and relevant quality control evidence.

Chapter 3 lays out the field test designs, data collection (test admin procedures), and student samples. This chapter also provides details regarding scoring processes and

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

quality and item calibration and equating procedures used to link the 3rd Edition test forms to LAS Links 1st and 2nd Editions.

Chapter 4 provides evidence of test score validity as it relates to the internal structure of LAS Links assessments, including descriptive statistics, reliability, standard error of measurement, decision consistency, and confirmatory factor analysis.

Chapter 5 covers the types of scores and reports offered for LAS Links, as well as the proficiency levels and underlying standard-setting and validation processes that support their validity.

Chapter 6 provides evidence of the relationship of the various test scores to other variables, along with recommendations for how test educational authorities that use LAS Links can demonstrate the construct validity of LAS Links scores in their own unique contexts.

Chapter 7 discusses the consequential validity of LAS Links test scores when used as intended, including special attention to the relationship between consequential validity and fairness.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 1: LAS Links Purpose, History, and Intended Uses

In the last several decades, researchers and policy makers have extensively studied and come to understand the pivotal role of academic language in effective curriculum and instruction (Anstrom et al., 2010). Conceptions of academic language have varied depending on the perspectives and goals of the researchers (e.g., curriculum, assessment, linguistic research); however, many researchers (Bailey & Huang, 2011; Gee, 2008; Gibbons, 1998, 2003; Scarcella, 2003; Schleppegrell, 2004; van Lier & Walqui, 2012) agree that academic language is defined as being

- situated language used within and across specific academic disciplines or content areas;
- embedded in sociocultural contexts that involve activities, practices, and language users;
- characterized by specific discourse and textual features such as genre, register, functions, syntax, and vocabulary; and
- integrated across different mediums and modes of communication used at different levels of complexity across grade spans.

Purpose of LAS Links

LAS Links is a comprehensive assessment system designed for the purpose of measuring the English-language proficiency levels of students in kindergarten through grade 12. LAS Links primarily serves K–12 students who are in the process of developing English language proficiency, and because of its rigor in academic language, LAS Links Forms may be particularly useful in understanding and diagnosing students' language needs for actively participating not only in general instructional settings but in discipline-specific learning as well.

LAS Links comprises tests in four domains: Listening, Speaking, Reading, and Writing. Students earn test scores on each of these four domains, and composite scores are calculated as the truncated average of their respective domain combinations as follows:

- Overall (Listening, Speaking, Reading, & Writing)
- Oral (Listening & Speaking)
- Comprehension (Listening & Reading)
- Literacy (Reading & Writing)
- Productive (Speaking & Writing)

LAS Links test administration takes approximately two hours when all four domains are administered. The tests offer a common scale across five grade-span levels (K–1, 2–3, 4–5, 6–8, and 9–12) so that growth in students' English language learning over time can be evaluated.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

History of LAS Links

LAS Links was originally developed as a response to changes to United States Federal requirements under Titles I and III of the *Every Student Succeeds Act* (ESSA, 2015). ESSA requires the assessment of students' progress toward English language proficiency (ELP) and the determination of students' need for English language instruction.

To address these legal requirements for states, and to support educators, the LAS Links First Edition (hereafter LAS Links 1st Edition) K–12 summative assessment suite, including both English forms (Forms A and B) and their conceptually parallel form in Spanish (Español Form A), was designed and developed by CTB/McGraw-Hill, later acquired by DRC in 2015.

Academic language for the K–12 student population received increased attention with the releases of the Common Core State Standards (CCSS; National Governors Association Center for Best Practices [NGA Center] & Council of Chief State School Officers [CCSSO], 2010) and the CCSSO's framework for English language proficiency development standards (2012), which describe perceived correspondence between language demand and the CCSS content standards.

Therefore, LAS Links 1st Edition was updated in 2013 with the LAS Links Second Edition (hereafter LAS Links 2nd Edition), which enhanced the emphasis on situated language use in school settings. Similar to LAS Links 1st Edition, the 2nd Edition offers test forms in both English (Forms C and D) and Spanish (Español Form B).

More recently, LAS Links 3rd Edition was developed in response to continual improvements in English language teaching, learning, and assessment practices. Specifically, the following updates have been made in the LAS Links 3rd Edition:

- The most current guidelines regarding bias, fairness, sensitivity, and accessibility for English learners were applied.
- The test content and graphics were adjusted to better reflect current English learner experiences, and current learning environments.
- The grades 2–12 Speaking tests were reduced in length by 6 items; the grade 1 Speaking test was reduced in length by 8 items, and the kindergarten Speaking test was reduced in length by 6 items.
- The grades K–1 Writing tests were reduced in length by 2 items.
- The kindergarten Reading test was reduced in length by 2 items.
- The format of items on the Writing and Speaking subtests were updated based on accessibility considerations.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

LAS Links 3rd Edition currently includes Form E, with Form F under development for release in the near future. LAS Links 3rd Edition is aligned to the same English language proficiency (ELP) standards as Forms A–D, with Forms A–E currently available in English. LAS Links Español Forms A and B are available for use in measuring Spanish language proficiency.

Intended Uses of LAS Links Scores

At the most foundational level, the notion of test use implies that there is some need to make claims about what students know and can do. In large-scale educational assessments, such claims are made based on students' test performance and, more specifically, their test scores and the meanings that are assigned to those scores. Since test scores are estimates of student traits that are variable and not directly observable, detailed evidence is required that score estimates are valid for their intended uses.

This is consistent with expectations embodied in the Standards that state the following:

> Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores for proposed uses that are evaluated, not the test itself. (p. 11)

The Standards also state the following:

> Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided. (p. 23)

Accordingly, the validity evidence provided in this report is grounded in the intended uses of LAS Links test results. These uses are described in detail in this section.

Generally, LAS Links test results can be used to identify students' English language proficiency status, instructional needs, and progress in attaining English language proficiency at progressively increasing levels. Such information is useful for making specific instructional, assessment, and accountability determinations as defined within the following specific uses:

- Identify students who require specific English language instruction and support programs,
- Plan instructional programs,
- Evaluate students' English language proficiency growth,

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- Determine student readiness to transition out of English language instruction support programs, and,
- Include student scores and proficiency levels for public reporting and accountability systems.

As with LAS Links 1st and 2nd Editions, the LAS Links 3rd Edition provides many benefits to states, school districts, and local educational agencies (LEAs) who choose to use the assessment in these ways.

Identifying Instructional and Support Needs

Title III of ESSA requires identification and annual assessment of the English proficiency of English learners. Under the Title III requirements, the English language proficiency standards must be based on the four modalities of Listening, Speaking, Reading, and Writing. Additionally, the assessment must measure English language proficiency in the five domains of Listening, Speaking, Reading, Writing, and Comprehension.

LAS Links can be used to identify K–12 students who are (or remain) eligible for Title III instructional programs. The test scores are also valuable for identifying students who may benefit from instructional support to improve their academic English for succeeding in classrooms with rigorous English-medium content learning activities.

Planning Instructional Programs

LAS Links results provide reliable English language proficiency scores and levels needed to make relevant and crucial instructional decisions.

The scores on LAS Links can be used as indicators of proficiency in Listening, Speaking, Reading, and Writing in academic and social English. This information can be used to determine the placement of students in a specific type of instructional program. When determining instructional placement, users are encouraged to consider the decision in conjunction with other available evidence and assessment instruments, including information provided in home language surveys, communication with parents, informal interviews with students, and possibly test scores on content knowledge, depending on the specific purpose of the instructional program.

LAS Links test scores can also assist in diagnosing students' strengths and areas for growth in English, especially their ability to use English in school settings. The test scores in Listening, Speaking, Reading, and Writing provide useful information about what skills students have and help determine their particular language needs in each of the four communicative skills. Using these results, teachers can plan appropriate instruction or remediation for the students.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Evaluating Student Growth

Students' progress from a beginning level to an advanced level of English language proficiency can be reflected by the scores on LAS Links. Because there are five grade spans of the tests (K–1, 2–3, 4–5, 6–8, and 9–12) and multiple test forms available per grade span that cover kindergarten through grade 12, the different grade spans can be used to track changes in English proficiency as the student continues in school across grades, from the beginning of the academic year to the end of the academic year, and from year to year. This feature may be especially useful in schools with bilingual education programs that have as a goal increasing students' English language proficiency over time as a goal.

Determining Readiness to Exit English Language Instruction and Support Programs

The rigor of the LAS Links scales (described in detail in Chapter 4) supports state and local educational authorities in the development of processes that may be used in determining when students are ready to exit specialized English language instruction and supports. Such decisions require that states employ additional and technically rigorous approaches to defining when students have acquired sufficient proficiency in English to succeed academically in classrooms where the language of instruction is English. Educational authorities who use LAS Links results to make program exit decisions about students are further encouraged to use any additional information about student proficiency that is collected through alternate means, including but not limited to the results of other high-quality valid and reliable measures, student evaluations, and other educator inputs.

Supporting Reporting Requirements in Accountability Systems

The rigor of the LAS Links scales (described in detail in Chapter 4) and score reports (described in Chapter 5), which include proficiency levels, supports state and local educational authorities in meeting Title I and III ESSA requirements to publicly report valid and reliable assessment results for all English language learners in K—12 within and across years.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 2: LAS Links Validity Evidence Related to Test Content and Response Processes

The Standards state the following:

> Tests and testing programs should be designed and developed in a way that supports the validity of the interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population. (p. 85)

The Standards further state the following:

> When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified. (p. 26)

Evidence of validity based on test content includes information about the framework's test specifications, including the test design and test blueprints. Test development involves creating a design framework from the statement of the construct to be measured. LAS Links 3rd edition is designed to assess school language that is critical for student intellectual growth in K−12 instructional settings and academic language essential for development within content area learning.

This chapter describes the LAS Links and other relevant English language development frameworks and standards correspondence, blueprints, item development procedures and cycles, content reviews, form construction, and relevant quality control procedures. Each of these components of test construction provides important evidence of content validity.

Framework

The LAS Links Standards Framework reflects a modification of several language development models currently used in guiding the education of English Learners/Multilingual Learners (ELs/MLs). The framework evaluates the receptive and productive control of language by ELs/MLs in social, school, and academic contexts. The standards in the framework are organized into a) language context strands, b) language domains and subtests, c) subskills/objectives, and d) proficiency levels.

The LAS Links Standards Framework meets the requirements for challenging expectations reflected in standards such as the Teachers of English to Speakers of

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Other Languages (TESOL) standards (2006), the Common European Framework of Reference for Languages (CEFR; 2001), and the CCSS (2010). Ensuring the correspondence of the LAS Links standards to these international and national standards is a key step to ensure that the LAS Links standards are equitable and comprehensive enough to assess English Language learners' ability to use English in various school contexts.

Language Context Strands

Although the LAS Links tests include diverse, culturally relevant content coverage, the focus of the assessments is on language and not on content knowledge. The social and school content covers intercultural and instructional communication (e.g., school-related tasks), while the academic content coverage includes communications related to English language arts, mathematics, science, social studies, history, and technical subjects. As previously mentioned, there are four language context strands applicable across all grade levels and language domains in the LAS Links Standards Framework:

- **Strand 1**
  Students are able to listen, speak, read, and write for **Social, Intercultural, and Instructional Communication**.
- **Strand 2**
  Students are able to listen, speak, read, and write for **Language Arts, Social Studies, and History.**
- **Strand 3**
  Students are able to listen, speak, read, and write for **Mathematics, Science, and Technical Subjects**.
- **Strand 4**
  Students are beginning to develop **Foundational Skills** for reading and writing (grades K–3 only).

Language Domains and Subtests

The LAS Links Standards Framework includes four language domains. Listening and Reading domains assess students' **receptive** control of language while the Speaking and Writing domains evaluate students' **productive** control of language.

Listening Test Description

The Listening test consists of two subtests: *Listen for Information* and *Listen for Academic Instruction*. All Listening items are in multiple-choice (MC) format. All instructions, audio passages, questions, and answer choices are delivered online via DRC's testing platform, INSIGHT, for the computer-based tests or delivered via audio files played on the test administrator's computer or device for the paper-based tests. Each question has three answer choices. In grades K–1, all answer choices are

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

pictures. In grades 2–3, there is a mix of both picture- and text–based answer choices. In grades 4–12, all answer choices are text-based.

Listen for Information

In *Listen for Information*, students listen to directions, brief school announcements, content-based discussions, and conversations. Then students answer questions about what they heard. Students are tested on skills such as following common explicit oral directions, identifying main ideas, and making inferences. In some grade spans, students are asked to identify purpose, comprehend idiomatic expressions, and make predictions.

Listen for Academic Instruction

In *Listen for Academic Instruction*, students listen to longer content-based discussions led by a teacher, with comments and contributions provided by class members. In this way, the Listening texts approximate authentic classroom discourse patterns that are co-constructed by the teacher and the class members. Discussions are drawn from two broad academic categories: (a) Language Arts, Social Studies, and History; and (b) Mathematics, Science, and Technical Subjects. Students identify main ideas and supporting details and make inferences. At some grade spans, students also make predictions.

Reading Test Description

The Reading test consists of three subtests in grades K–3: *Read Words, Read School Texts, and Read Academic Texts*; and two subtests in grades 4–12: *Read School Texts and Read Academic Texts*. Reading questions are multiple-choice in format with three answer choices (some picture-based and some text-based) in grades K–3 and four text-based answer choices in grades 4–12.

Read Words (Grades K—1 and 2—3)

In *Read Words*, students in grade K–3 students respond to items addressing word-analysis tasks: identifying rhyming words, applying letter-sound relationships in order to read English words, and applying letter-sound relationships in order to read English phonemes/graphemes. In grades 2–3, students have the additional task of applying knowledge of morphemes and grammar to word meaning.

Read School Texts

In *Read School Texts*, students read a variety of short texts, such as classroom signs, school notices, letters, website postings, emails, and text messages between students. In addition, students in grades 1–12 read texts similar to those they will likely encounter in the content areas of English Language Arts, History, and Social Studies or Mathematics, Science, and Technical Subjects. These texts emulate grade-span appropriate workbook or classroom tasks and measure students' ability to understand the text, not their ability to complete the task being described. All questions are multiple-

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

choice in format and measure students' ability to identify main ideas and supporting details, interpret words and phrases as they are used in text, and identify view, tone, and attitude.

Read Academic Texts

In *Read Academic Texts*, students read extended grade-span appropriate passages drawn from two broad academic categories: (a) Language Arts, Social Studies, and History; and (b) Mathematics, Science, and Technical Subjects. Although both fiction and non-fiction texts are included, there is an emphasis on more complex non-fiction texts.

In grades K–1, each passage has two related questions that measure the student's ability to identify main ideas and important details; or identify view, tone, and attitude. In grades 2–3, each passage has five related questions that measure the student's ability to identify main ideas and important details; identify view, tone, and attitude; and interpret words and phrases as they are used in text. In grades 4–12, each passage has six related questions that measure the student's ability to identify main ideas and important details; identify view, tone, and attitude; and interpret words and phrases as they are used in text.

Speaking Test Description

The Speaking test consists of five subtests: *Make Conversation, Use Academic Words, Describe and Request Information, Present and Explain Information*, and *Tell a Story*. Note that kindergarten students take only the first set of questions in *Present and Explain Information*. All Speaking items are performance-based in format. They measure vocabulary and grammatically correct verbal expressions in social and academic language. Tasks in the Speaking subtest elicit the production of single-word responses as well as multiple sentences related to school-appropriate topics.

Make Conversation

In *Make Conversation*, students answer basic conversational questions that are appropriate in a school or social setting by either providing information or expressing opinions and preferences. Student responses are scored as incorrect (0 points), correct (1 point), or no response (NR).

Use Academic Words

In *Use Academic Words*, students are shown pictures of grade-appropriate vocabulary items, common objects, and objects and concepts they encounter in the classroom. The students are asked to identify the object or concept. Students respond with a single word or short phrases. Student responses are scored as incorrect (0 points), correct (1 point), or no response (NR).

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Describe and Request Information

In *Describe and Request Information*, students are shown a picture depicting an academic or social situation and asked to describe it using sentences. Next, students are required to complete a speech act or function by, for example, saying what a participant in the given academic or social situation might do. Students in grades K–3 students demonstrate their ability to ask questions, request clarification, and negotiate for meaning; while students at all grades demonstrate their ability to make various requests. Student responses are scored on a 0–3 rubric.

Present and Explain Information

In *Present and Explain Information*, student responses are scored on a 0–3 rubric. In grades K–1, students are shown an illustration of an academic or social situation and are asked to describe what is happening in the illustration or explain the purpose, use, or feature of a particular object in the illustration using words, phrases, or sentences. Grade 1 students take an extended section of the same task with two more questions in which they describe or elaborate on an additional set of illustrations.

In grades 2–3, students are shown an illustration of people, a location, or scenery and asked to describe the illustration. Students are then shown a different but related illustration and asked to describe that one as well. Finally, students are asked to compare the information in the two graphics and explain how the information is the same or different.

In grades 4–12, students are shown a slide, map, or other graphic depicting information, such as a chart, and asked to talk about the information in the graphic as if they were giving a presentation to a class. Students are then shown a different but related graphic and asked to present that information as well. Finally, students are asked to compare the information in the two graphics and explain how the information is the same or different.

Tell a Story

In *Tell a Story*, students are shown four related pictures that illustrate a story with a beginning, middle, and end. Pointing to the series of four pictures, the Examiner begins the story by reading a story starter to contextualize the pictures without giving away vocabulary or key content. Students are then asked to complete the detailed story depicted in the series of illustrations using multiple sentences to interpret, narrate, and paraphrase events. Student responses are scored on a 0–4 rubric.

Writing Test Description

The Writing test consists of four sections for Grades K–1: *Start Writing, Use Grammar and Conventions, Write to Express Ideas*, and *Write Academic Texts*. (Kindergarten students do not take the Write Academic Texts section.) For grades 2–12, there are

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

three sections: *Use Grammar and Conventions, Write Academic Texts*, and *Write to Express Ideas*.

The Writing subtest includes both MC and constructed-response (CR) items that assess the student's knowledge of grammar, word order, and word choice and the student's ability to apply that knowledge to produce sentences and paragraphs that are commonly expected of students at their respective grade levels.

Start Writing (Grades K–1 Only)

In *Start Writing*, students in grades K–1 students copy words and sentences and write numbers and letters. In addition, students write English words that identify pictures of common objects.

Use Grammar and Conventions

In *Use Grammar and Conventions*, students in grades K–1 select grammatically correct sentences and indicate whether a sentence has correct use of capital letters, punctuation, articles, singular and plural nouns, pronouns, and subject/verb agreement. In grades 2–12, students select the grammatically–appropriate response in order to complete sentences and paragraphs. Grammatical features are selected according to each grade span and assess the correct use of capitalization, sentence-ending marks, articles, adjectives and adverbs, singular and plural nouns, pronouns, subject/verb agreement, tense and aspect, prepositional phrases, conjunctions, commas, and auxiliary verbs.

Write Academic Texts

In *Write Academic Texts*, students in grades 1–3 students write sentences describing pictures drawn from two broad academic categories: (a) Language Arts, Social Studies, and History; and (b) Mathematics, Science, and Technical Subjects. These tasks approximate common real-world classroom assignments for which students in grades 1–3 students are expected to write about something they see. Students in grades 2–3 also write simple sentences to interpret, analyze, or state opinions regarding what they see.

Students in grades 4–12 are first asked to write a short summary (two to five sentences) of a paragraph selected from a passage they had read earlier in the Read Academic Texts section of the Reading subtest. Next, students are shown a table or diagram and asked to write one or two full sentences explaining the information it contains. Finally, students are asked to compare the paragraph and the information contained in the table or diagram and explain in one or two sentences how they are the same or different. These tasks approximate common real-world classroom assignments where students are expected to summarize, in their own words, course reading material; extract tabular information and express it in prose; and compare and contrast academic content. Responses are scored on a 0–3 rubric to assess the student's ability to communicate effectively using appropriate grammar, vocabulary, and conventions.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Write to Express Ideas

In *Write to Express Ideas*, students are given an opportunity to write for personal communication. Students in grades K–1 write a sentence describing a person. Students in grades 2–3 write a letter. Students in grades 4–12 write extended responses to an email message, note, or blog entry. The writing tasks for grades 2–12 are designed for students to be able to demonstrate their ability to describe, explain, report, compare, narrate, persuade, or express ideas in writing. Responses are scored on a 0–3 (grades K–1) or 0–4 (grades 2–12) holistic rubric to assess the student's use of appropriate grammar and vocabulary and the student's ability to express meaning in a cohesive and coherent manner.

Subskills/Objectives

Table 1 presents a complete list of subskills/objectives within the LAS Links Standards Framework. These subskills/objectives are organized by language domain.

Table 1. LAS Links Standards Framework

Subject and Standard	Substandard
Listening: L1. Follow common, explicit oral directions to participate in diverse academic or social tasks	-
Listening: L2. Respond to idiomatic expressions to participate in diverse academic or social tasks, including phrasal verbs with idiomatic meaning (e.g., give me a hand or settle for)	-
Listening: L3. Demonstrate understanding of academic and social situations that contain diverse language genres, registers, and varieties	L3.1. Identify purpose L3.2. Identify main ideas L3.3. Identify supporting details L3.4. Relate to practical issue
Listening: L4 Interpret layers of meaning using critical listening skills and learning strategies in academic and social situations that contain diverse language genres, registers, and varieties	L4.1. Make predictions based on known information L4.2. Make inferences based on known information

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Subject and Standard	Substandard
Reading: R1. Analyze words	R1.1. Identify rhyming words R1.2. Apply letter-sound relationships to read English words R1.3. Apply letter-sound relationships to read English phonemes R1.4. Apply knowledge of morphemes and syntax to word meaning
Reading: R2. Understand word meaning	R2.1. Associate words with their representation R2.2. Classify words R2.3. Interpret words and phrases as they are used in a text, including determining technical, connotative, and figurative meanings.
Reading: R3. Comprehend written material	R3.1. Identify main ideas R3.2. Identify supporting details R3.3. Identify important literary features of text R3.4. Analyze the structure of texts, including how specific sentences, paragraphs, and larger portions of the text (e.g., a section, chapter, scene, or stanza) relate to each other and the whole R3.5. Identify point of view, tone, and attitude R3.6. Make predictions based on known information R3.7. Make inferences based on known information

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Subject and Standard	Substandard
Speaking: S1. Participate in diverse academic or social conversations, with attention to appropriate register, grammar, vocabulary, and pronunciation	S1.1. Provide information S1.2. Describe information S1.3. Interpret and analyze information S1.4. Relate information to personal experience or practical issue S1.5. Express opinions and preferences S1.6. Make requests S1.7. Ask questions, request clarification, and negotiate for understanding S1.8. Conduct transactions
Speaking: S2. Demonstrate knowledge related to diverse academic or social settings, with attention to appropriate register, grammar, vocabulary, and pronunciation	S2.1. Identify an object (inanimate and animate) or concept S2.2. Describe purpose, use, or feature, using words, phrases, or sentences S2.3. Identify an academic or social situation and describe it, using sentences
Speaking: S3. Describe ideas, experiences, and immediate surroundings in diverse academic and social settings, with attention to appropriate register, grammar, vocabulary, and pronunciation	S3.1. Describe process S3.2. Describe people, locations, and scenery
Speaking: S4. Speak persuasively in diverse academic or social situations, with attention to appropriate register, grammar, vocabulary, and pronunciation	S4.1. Explain process S4.2. Explain ideas and opinions
Speaking: S5. Talk in depth and with detail about diverse academic or social events, with attention to appropriate register, grammar, vocabulary, and pronunciation	S5.1. Interpret, narrate, and paraphrase events, using visual information S6.1. Present with integrated information from multiple sources
Speaking: S6. Present with integrated information	
Writing: W1. Copy words and sentences	-
Writing: W2. Write letters, numerals, and words	-

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Subject and Standard	Substandard
Writing: W3. Use appropriate grammar and style	W3.1. Use articles W3.2. Demonstrate correct use of singular and plural W3.3. Use subject/verb agreement W3.4. Demonstrate correct use of tense and aspect W3.5. Use conjunctions W3.6. Use pronouns correctly W3.7. Distinguish adjectives and adverbs W3.8. Use prepositional phrases W3.9. Use auxiliary verbs W3.10. Use nominalization W3.11. Use parallel structure
Writing: W4. Use appropriate capitalization and punctuation	W4.1. Use appropriate capitalization W4.2. Use appropriate sentence-ending marks W4.3. Use commas appropriately W4.5. Use semi-colons appropriately W4.6. Use colons appropriately
Writing: W5. Use appropriate sentence structure	W5.1. Differentiate complete sentences from fragments W5.2. Differentiate complete sentences from run-ons W5.3. Form statements and questions W5.4. Use various types of clauses W5.5. Use various types of phrases (noun, verb, adjectival, adverbial, participial, prepositional, and absolute)
Writing: W6. Write sentences to summarize, describe, narrate, interpret, analyze, state opinion, relate, or explain	W6.1. Write sentences to summarize W6.2. Write sentences to describe or narrate W6.3. Write sentences to interpret or analyze W6.4. Write sentences to state opinions W6.5. Write sentences to relate to personal experience or practical issue W6.6. Write sentences to explain
Writing: W7. Write expository compositions	W7.1. Write to describe, explain, report, compare, narrate, persuade, or express
Writing: W8. Write with integrated information	W8.1. Write with integrated information from multiple sources

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Levels

The LAS Links Standards Framework represents a continuum of English language development in social, school, and academic contexts. LAS Links has the following five levels, and each level builds on to the next level (see Appendix D for a full description of the LAS Links Proficiency Level Definitions):

1. Beginning,
2. Early Intermediate,
3. Intermediate,
4. Proficient,
5. Above Proficient.

To describe these levels, the following list represents the progression of the development of ELs/MLs' receptive and productive control of lexical, syntactic, phonological, and discourse features in English:

1. Developing,
2. Emerging,
3. Exhibiting limited range,
4. Exhibiting control,
5. Commanding a high degree of control.

In terms of the complexity or the difficulty of texts that ELs/MLs can comprehend and analyze, the following progression applies:

1. Very basic level,
2. Familiar topics,
3. Range of grade-level appropriate,
4. Across and within disciplines (grade-level appropriate),
5. Wide range.

Finally, in terms of the communicative skills of ELs/MLs, the following progression applies:

1. Developing the ability and using familiar topics,
2. Developing the ability to communicate effectively,
3. Refining the ability to communicate effectively and using context clues,
4. Communicating effectively and beginning to express in creative forms,
5. Communicating effectively, skillfully organizing and explaining information, and expressing subtle nuances.

Correspondence with CCSS, TESOL, CEFR, ELPA21, and WIDA Standards

The alignment of the LAS Links Framework with the most commonly used content standards has been examined and the results show a strong correspondence in each case discussed in this section: CCSS, TESOL, the Common European Framework of

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Reference for Languages (CFER), English Language Proficiency Assessment (ELPA21), and the 2012 WIDA Consortia English Language Development Standards correspondences.

CCSS Correspondence

Correspondence between the LAS Links Standards Framework and the CCSS (2010) is highly valuable because the CCSS serves as a set of overarching educational standards for K–12 students in the United States. The correspondence ensures that LAS Links shares the expectations of language proficiency in Listening, Speaking, Reading, and Writing, as well as knowledge and skills in various content areas relevant to social, school, and academic contexts.

Table 2 represents a sample correspondence of LAS Links standards to the CCSS (2010). This sample focuses on students' skills in terms of summarizing, determining/identifying main ideas, and explaining/supporting with details.

Table 2. Sample Correspondence of LAS Links Standards with the CCSS

LAS Links Standards Framework	CCSS
W6: Write sentences to summarize, describe, narrate, interpret, analyze, state opinion, or explain. ***W6.1****: Write sentences to summarize.* ***W6.2****: Write sentences to describe or narrate.* ***W6.6****: Write sentences to explain.* **W7**: Write expository compositions. ***W7.1****: Write to describe, explain, report, compare, narrate, persuade, or express.* **W8**: Write with integrated information. ***W8.1****: Write with integrated information from multiple sources.*	**SL.5.2**: Summarize a written text read aloud or information presented in diverse media and formats, including visually, quantitatively, and orally.
L3: Demonstrate understanding of academic and social situations that contain diverse language genres, registers, and varieties. ***L3.1****: Identify purpose.* ***L3.2****: Identify main ideas.* ***L3.3****: Identify supporting details.*	**RL.5.2**: Determine a theme of a story, drama, or poem from details in the text, including how characters in a story or drama respond to challenges or how the speaker in a poem reflects upon a topic; summarize the text.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

LAS Links Standards Framework	CCSS
R3: Comprehend written material. ***R3.1****: Identify main ideas.* ***R3.2****: Identify supporting details.* ***R3.5****: Identify point of view, tone, and attitude.*	**RI.5.2**: Determine two or more main ideas of a text and explain how they are supported by key details; summarize the text.
S1: Participate in diverse academic or social conversations, with attention to appropriate register, grammar, vocabulary, and pronunciation. ***S1.1****: Provide information.* ***S1.2****: Describe information.* **S3**: Describe ideas, experiences, and immediate surroundings in diverse academic and social settings, with attention to appropriate register, grammar, vocabulary and pronunciation. ***S3.1****: Describe process.* ***S3.2****: Describe people, locations, and scenery.* **S4**: Speak persuasively in diverse academic or social situations, with attention to appropriate register, grammar, vocabulary, and pronunciation. ***S4.1****: Explain process.* ***S4.2****: Explain ideas and opinions.* **S5**: Talk in depth and with detail about diverse academic or social events, with attention to appropriate register, grammar, vocabulary, and pronunciation. ***S5.1****: Interpret, narrate, and paraphrase events, using visual information.* **S6**: Present with integrated information. ***S6.1****: Present with integrated information from multiple sources.*	**SL.5.3**: Summarize the points a speaker makes and explain how each claim is supported by reasons and evidence.

TESOL Correspondence

It is valuable to compare the LAS Links context strands with the TESOL standards (2006) because the TESOL standards play a critical role in developing ESL standards for teachers of K–12 students in the United States. LAS Links examines its correspondence with TESOL because the target audience of both LAS Links and TESOL includes students who use languages other than English and who need to learn English to be successful inside and outside of the U.S. classroom. Table 3 shows how LAS Links reporting strands correspond to the TESOL standards (2006). LAS Links has

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

combined the target language use skills into three strands for practical use and reporting.

Table 3. LAS Links Strands and the TESOL Standards (2006)

LAS Links	TESOL
Strand 1: Students are able to listen, speak, read, and write for Social, Intercultural, and Instructional Communication.	**Standard 1**: English language learners communicate for social, intercultural, and instructional purposes within the school setting.
Strand 2: Students are able to listen, speak, read, and write for Language Arts, Social Studies, and History.	**Standard 2**: English language learners communicate information, ideas, and concepts necessary for academic success in the area of Language Arts. **Standard 5:** English language learners communicate information, ideas, and concepts necessary for academic success in the area of Social Studies.
Strand 3: Students are able to listen, speak, read, and write for Mathematics, Science, and Technical Subjects.	**Standard 3**: English language learners communicate information, ideas, and concepts necessary for academic success in the area of Mathematics. **Standard 4**: English language learners communicate information, ideas, and concepts necessary for academic success in the area of Science.
Strand 4: Students are beginning to develop Foundational Skills for reading and writing (only applicable for grades K–3).	-

CEFR Correspondence

The CEFR (2001) serves as an influential source in the development of language and education policies in Europe and beyond. Many language testers and education/examination boards refer to the CEFR to help define language proficiency levels and analyze language qualifications. Table 4 shows how LAS Links proficiency levels conceptually align with the CEFR (2001).

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 4. Correspondence of LAS Links Proficiency Levels with the CEFR (2001)

LAS Links	CEFR
Beginning Level: Is beginning to develop receptive and productive uses of English	**Breakthrough Level**: Understand and use familiar, everyday expressions; very basic phrases
Early Intermediate: Can identify, describe, and discuss simple pictorial or text prompts; can interpret language related to familiar topics; and can draw simple inferences and make simple comparisons	**Waystage**: Understand sentences and frequently used expressions; simple and routine tasks
Intermediate: Can compare, contrast, summarize, and relate text to graphic organizers and uses coherent language use but lacks elaboration	**Threshold**: Understand main points; produce simple, connected text; briefly give reasons
Proficient: Can adequately express ideas and organize responses in logical and sequenced order; can distinguish nuances of meaning; and can interpret, analyze, and evaluate written and oral information	**Vantage**: Understand main ideas of complex text; produce clear, detailed text
Above Proficient: Can critically evaluate and synthesize written and oral information; can draw sophisticated inferences and explain their reasoning; can skillfully organize information; and can express subtle nuances of meaning	**Effective Operational Fluency – Level C1 and Master – Level C2**: Understand wide range, longer texts; recognize implicit meaning; produce clear, detailed, and well-structured text

ELPA21 Correspondence

An alignment study between LAS Links 2nd Edition and the ELPA21 Standards was conducted in 2015 (Lotfi & Houston, 2015). The alignment process found that 100% of the LAS Links items were aligned to an ELPA standard. For the purposes of this alignment, only a principal alignment was chosen, although the three standards ELPA refers to as supportive might provide a secondary alignment in some cases. In a few instances, standards 8, 9, or 10 were selected as the principal standard.

ELPA standards 5 and 6 refer to activities best undertaken in the classroom or measured by performance assessment. No LAS Links item was aligned to either of these two standards. Similarly, standard 7 was not aligned directly to any LAS Links item but an argument could be made that this task is an implicit part of many items. However, the reviewer also noted that while it is possible to directly assess this standard, it may also be best left to classroom practices.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Because LAS Links contains CR items, the tables that report alignment are reporting on a score point basis as opposed to an item basis. This is important as this shows an alignment distribution not only by standard, but by proficiency level. This is an enhanced feature of this alignment study as most alignment studies do not take this data point into consideration.

Table 5 shows the results of the study of the correspondence between LAS Links and ELPA21 Standards. LAS Links shows strong coverage of the ELPA21 standards.

Table 5. LAS Links Score Points Aligned to ELPA21 Standards

ELPA21 Standards	Subject	Score Point Distribution: Listening	Score Point Distribution: Reading	Score Point Distribution: Speaking	Score Point Distribution: Writing
construct meaning from oral presentations and literary and informational text through grade appropriate listening, reading, and viewing	Listening; Reading	39%	15%	-	-
participate in grade-appropriate oral and written exchanges of information, ideas, and analyses, responding to peer, audience, or reader comments and questions	Listening Reading Speaking Writing	57%	61%	51%	23%
speak and write about grade- appropriate complex literary and informational texts and topics	Speaking Writing	-	-	25%	31%
construct grade-appropriate oral and written claims and support them with reasoning and evidence	Speaking Writing	-	-	24%	14%
conduct research and evaluate and communicate findings to answer questions or solve problems	Listening Reading Speaking Writing	-	-	-	-
analyze and critique the arguments of others orally and in writing	Listening Reading Speaking Writing	-	-	-	-
adapt language choices to purpose, task, and audience when speaking and writing	Speaking Writing	-	-	Rubric	Rubric
determine the meaning of words and phrases in oral presentations and literary and informational text	Supportive	4%	25%	-	2%
create clear and coherent grade-appropriate speech and text	Supportive	-	-	-	2%
make accurate use of standard English to communicate in grade appropriate speech and writing	Supportive	-	-	-	29%

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Review of ELPA21 Alignments

In April 2020, an independent third-party reevaluation of the 2015 Middlebury Interactive Alignment Study of LAS Links 2nd Edition (Lotfi & Houston, 2015) to the Connecticut English Language Proficiency (CELP) Standards, was adopted directly from the ELPA21 Standards. This reevaluation study also evaluated the alignment between the LAS Links 2nd Edition assessments and the CELP Standards for each of the grade bands K—1, 2—3, 4—5, 6—8, and 9—12.

The study focused on alignment to the Primary CELP Standards 1, 2, 3, and 4, and the Supporting CELP Standards 7, 8, 9, and 10. The Connecticut State Department of Education's reevaluation revealed a moderate alignment between the CELP Standards and LAS Links 2nd Edition. CELP Standards 1, 2, 3, 7, 9, and 10 showed adequate to strong coverage on LAS Links 2nd Edition, while CELP Standards 4 and 8 showed a need for better coverage on LAS Links.

This review included educators from Connecticut as well as across the United States. Members of the review panel and their affiliations and roles are noted in Table 6.

Table 6. Review Panel Participants

Name	District/Organization	Role
Jean Borrup	Berlin	English Learner Assessment Coordinator
Nadia Gonzalez	North Haven	English Learner Assessment Coordinator
Kimberly Lebrun	Hartford	English Learner Teacher
Darcy Lockwood	RSD 15	English Learner Assessment Coordinator
Mary Parady	CTECS	English Learner Assessment Coordinator
Marilyn Rosario-Cosme	East Hartford	English Learner Teacher
Celmia (Sally) Vernaglia	East Lyme	English Learner Assessment Coordinator
Karen Lapuk	LEARN	English Learner Assessment Coordinator

The CELP Standards found to require greater coverage on LAS Links are as follows:

- *CELP Standard 4: Construct grade-appropriate oral and written claims and support them with reasoning and evidence.*
- *CELP Standard 8: Determine the meaning of words and phrases in oral presentations and literary and informational text*.

DRC considered the need for greater coverage of CELP Standards 4 and 8 during its LAS Links 3rd Edition item development and review processes. In an effort to provide better coverage of these CELP Standards in the 3rd Edition, DRC developed a number of field test items aligning to CELP Standards 4 and 8. An item review with CT educators and other educators from around the country was conducted in Summer

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

2020. In addition to reviewing the items for bias, fairness, sensitivity, and suitability for inclusion, CT educators also identified each item's degree of alignments to the CELP Standards.

CT educators determined that a number of field test items at each grade band were aligned to CELP Standards 4 and 8. DRC test development specialists with expertise in English language learning and second language acquisition also reviewed the CELP Standards 4 and 8 for correspondence to the LAS Links Standards. Table 7 shows the correspondence between CELP Standards 4 and 8 and the LAS Links Standards, informed by CT educator item alignments and DRC English language learner test development specialists' reviews. Items aligned to these standards comprised a pool of field test items from which LAS Links 3rd Edition content was selected with an effort to improve the alignment of LAS Links 3rd Edition to the CELP Standards.

Table 7. CELP Standards 4 and 8 Correspondence to LAS Links Standards

CELP Standard	Subject	LAS Links Standards
CELP Standard 4: Construct grade-appropriate oral and written claims and support them with reasoning and evidence	Writing Standards	Write Academic Texts: WR.2/3.C.4.e: Write sentences to summarize, describe, narrate, interpret, analyze, state an opinion, relate, or explain; Write simple sentences to state opinions.
CELP Standard 4: Construct grade-appropriate oral and written claims and support them with reasoning and evidence	Writing Standards	Write to Express Ideas: WR.1.B.3.a: Write expository compositions; Write to describe, explain, report, compare, narrate, persuade, or express.
CELP Standard 4: Construct grade-appropriate oral and written claims and support them with reasoning and evidence	Speaking Standards	Make Conversation: SP.1.A.1.b: Participate in diverse school interactions, with attention to appropriate register, grammar, vocabulary, and pronunciation; Express opinions and preferences.
CELP Standard 8: Determine the meaning of words and phrases in oral presentations and literary and informational text	Reading Standards	Read School Texts: RD.1/2/3.B.2.a: Understand word meaning; Interpret words and phrases as they are used in text.
CELP Standard 8: Determine the meaning of words and phrases in oral presentations and literary and informational text	Reading Standards	Read Academic Texts: RD.2/3.C.2.a: Understand word meaning; Interpret words and phrases as they are used in text.
CELP Standard 8: Determine the meaning of words and phrases in oral presentations and literary and informational text	Listening Standards	Listen for Information: LI.1.B.4.a: Respond to idiomatic expressions to participate in diverse tasks, including phrasal verbs with idiomatic meaning (e.g., give me a hand or settle for); Respond to idiomatic expressions.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

WIDA Correspondence

An alignment study was conducted in February 2018 for the LAS Links Forms C & D assessment and the 2012 WIDA Consortia English Language Development Standards. Eight reviewers analyzed the WIDA Standards and the LAS Links assessment according to Dr. Norman Webb's alignment methodology. To appropriately align the LAS Links assessment to the WIDA Standards, all elements of the standards were included. Two alignment studies were conducted to account for the interaction between the domains, standards, and levels: LAS Links to WIDA Standards and LAS Links to WIDA Levels.

Data on the alignment of the LAS Links assessment were collected from the eight reviewers following the methodology developed by Norman Webb, as modified by H. Gary Cook (2007), for English Language proficiency assessments. The data collected were then analyzed to determine whether the LAS Links assessment met the criteria established by the alignment model. The major difference between the Webb methodology and the modification by Cook is the substitution of Linguistic Difficulty Level (LDL) for Webb's Depth of Knowledge (DOK).

The reviewers were English learner curriculum and test development specialists. Eight reviewers participated in the study, with six reviewers participating in the study for each grade band. The reviewers began the alignment process by first reviewing and analyzing the 2012 WIDA Standards for linguistic difficulty level. Subsequently, reviewers analyzed the LAS Links operational items for linguistic difficulty level and then aligned the LAS Links operational items to the 2012 WIDA Standards. Once reviewers determined the primary and/or secondary alignment and linguistic difficulty level for each item, they analyzed the entire assessment for linguistic difficulty level consistency, categorical concurrence, range-of-knowledge correspondence, and balance of representation.

The reviewers' judgments were statistically analyzed according to Webb's model of alignment. The model's statistical alignment criteria were applied, and the results were reviewed along with written responses to a debriefing questionnaire. Alignment of the test content with the 2012 WIDA Standards and the 2012 WIDA Levels was examined for this study.

Alignment to WIDA Standards

The following list shows the alignment of the Listening items to WIDA Standards 1 through 5 across all the LAS Links grade bands. It shows a strong alignment across all alignment indices for the Listening domain. Recall that, within each of the four alignment dimensions, there are three classifications for an item: Strong, Moderate, and Limited. Also, the five standards for WIDA are the following:

- Standard 1: Social and Instructional Language
- Standard 2: The Language of Language Arts

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- Standard 3: The Language of Mathematics
- Standard 4: The Language of Science
- Standard 5: The Language of Social Studies

Alignment to WIDA Levels

WIDA also assigns language proficiency levels in the standards. Following the Cook (2007) alignment study, DRC also aligned the LAS Links assessment to the WIDA Levels. Using the same four alignment criteria of linguistic difficulty level consistency, categorical concurrence, range-of-knowledge correspondence, and balance of representation, items were evaluated for alignment to the five levels used by WIDA:

- Level 1: Entering
- Level 2: Emerging
- Level 3: Developing
- Level 4: Expanding
- Level 5: Bridging

Taken as a whole, the alignment of the LAS Links items to the WIDA Standards and Levels is strong. Given the results, the use of LAS Links as a tool for progress monitoring and student attainment of English language proficiency is recommended and would help teachers and administrators assess their students well with regard to the WIDA Standards and Levels.

LAS Links 3rd Edition Updates

At the conclusion of the series of meetings to collect stakeholder input regarding LAS Links content needs, the following updates were specified for inclusion in the LAS Links 3rd Edition:

- The most current guidelines regarding bias, fairness, sensitivity, and accessibility for English learners have been applied.
- The test content and graphics have been adjusted to better reflect current English learner experiences and current learning environments. The format of some item types in Writing and Speaking were updated based on educator feedback.
- Blueprint modifications were made to select content areas and grade levels.
- Estimated administration times for operational tests were included.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Bias, Fairness, Sensitivity, and Accessibility

At every stage of the item and test development process for the LAS Links items, passages, and graphics, DRC employs procedures that are designed to ensure that items and tests meet Standard 7.4 of the *Standards for Educational and Psychological Testing* (AERA et al, 2014).

> Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain. (126)

Item and passage development, as well as internal and external item reviews, focused especially on consideration of bias, fairness, sensitivity, and accessibility for the diverse and multicultural populations of EL/ML students. This population of students has a diverse range of experiences with schooling, technology, and academic language, and special consideration was given to ensure the passages and items were free from bias and were fair and accessible for EL/ML students. Additional information about this process can be found in the Content and Bias/Sensitivity Review section in Chapter 2 of this report.

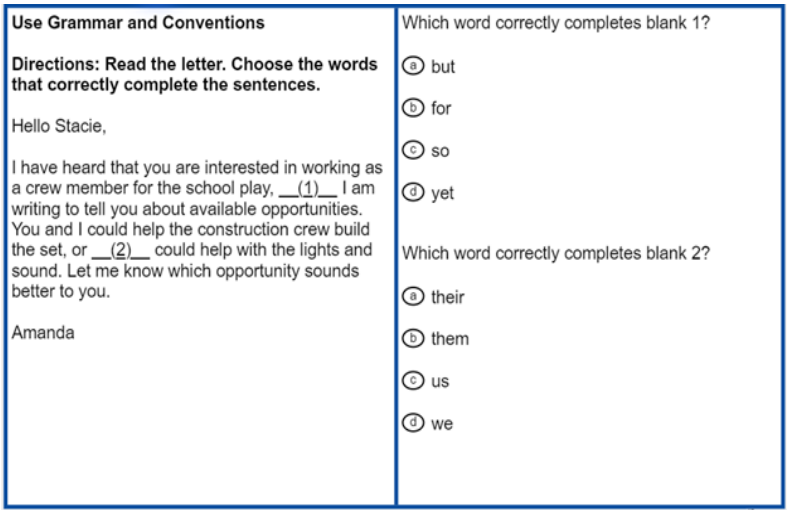
Test Content, Graphics, and Item Formats

The test content and graphics have been developed to reflect current English learner experiences and current learning environments. The test content includes experiences that are universal and accessible to English learner students, such as common social situations and classroom settings. For example, one of the Speaking practice items asks student to identify a chair, an object commonly found in classroom and home settings.

The format of some item types in Writing and Speaking were updated based on educator feedback. An example of the revised format used for pairs of writing items using the same stimulus paragraph has been provided in Figure 1 below. This example reflects the content included in a pair of items in grades 9–12. This item demonstrates the enhanced student view that includes both the stimulus paragraph and items on the same screen.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Figure 1. Example of Updated Item format



Update for Grades K–1 Writing Assessment

In the Start Writing section of the grades K–1 assessment, there will be a difference between the print and the online forms for those items that ask students to “write” a number or letter. Online, students will be asked to “type” the number or type the letter. In print, they will continue to be asked to “write” the number or letter.

- The Start Writing items that ask students to look at a picture and write the word for that picture will have the same difference between print and online. Online, students will be asked to “type” the word for the picture in the box. In print, they will continue to be prompted to “write” the word on the line.
- The Start Writing items that currently ask students to copy a word or sentence will continue to use the word “copy” in both print and online.
- In all other sections of the Writing assessment (including at other grade bands), the word “write” will continue to be used for both print and online, as we are referring to the cognitive process of crafting a response rather than the physical act of putting that response in the response area (in print or online).

Blueprint Modifications

Update for Grade K Reading Assessment Blueprint

The item count for the kindergarten Reading subskill area Read School Texts has been reduced from ten to eight items. This change was made based on educator feedback indicating that some skills that are important to assess for first-grade students are beyond the appropriate expectations for kindergarten students. Additionally, student response data indicate that kindergarten students can be assessed reliably with fewer items in this subskill area.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Update for Grades K—1 Writing Assessment Blueprint

The item count for the grades K−1 Writing subskill area Foundational Skills has been reduced from nine to seven items. This change was made as student response data indicate that seven items are sufficient to measure student performance accurately and reliably in this subskill area.

Update for Grades K—12 Speaking Assessment Blueprint

The item count for the grades K−12 Speaking subskill area of Make Conversation has been reduced from three to one item, and the item count for the grades K−12 Speaking subskill area Use Academic Words has been reduced from four to two items. Additionally, the item count for the Speaking subskill area of Present and Explain Information has been reduced from eight to six items for the assessments in grades 2−12, reduced from eight to four items for the grade 1 assessment, and reduced from four to two items in the kindergarten assessment. These changes were made as student response data indicate that fewer items are sufficient to measure student performance accurately and reliably in these subskills area.

Estimated Time to Complete Administration of Each Subtest for LAS Links 3rd Edition

Although the LAS Links is an untimed test, Table 8 shows the estimated time to complete the administration of each subtest across grade bands of the 3rd Edition. At grades K−1, for example, the inclusion of field-test items in the online forms is balanced by the reduction in operational items; therefore, estimated testing times remain consistent with the 2nd Edition tests. Listening, Speaking, and Writing all have the same estimated administration times as the previous forms. Only Reading has an increased estimated administration time (increased by five minutes).

Table 8. LAS Links 3rd Edition Testing Times

Subtest Area	Estimated Administration Time
Speaking	15 minutes—all grades
Listening	35 minutes—grades K−1, 6−8, 9−12, 30 minutes—grades 2−3, 4−5
Reading	40 minutes—grades K−1, 2−3, 4−5, 50 minutes—grades 6−8, 9−12, CR items were replaced with MC items.
Writing	30 minutes—grades K−1, 40 minutes—grades 2−12

For scheduling purposes, these time allocations are recommended to complete the administration of each subtest and do not include setup, logins, etc.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

LAS Links 3rd Edition Test Design

As in the 1st and 2nd Editions, the LAS Links 3rd Edition is organized by grade and skill area. Core operational items are also distinguished from embedded field test items for test item review purposes. Embedded field-test items facilitate ongoing improvement and monitoring of the tests and enable the development of future LAS Links forms. Table 9 provides the test design details for LAS Links 3rd Edition.

Table 9. LAS Links Forms E and F Operational (OP) Field Test Design

Grade Level	Skill Area	Number of Practice Items	Number of OP Items	Number of Embedded FT Items	Total Number of Items (Practice + OP + FT)
K-1	Listening	2	20	3	25
K-1	Speaking	4	10 (8 for K)	3	17 (15 for K)
K-1	Reading	3	30 (24 for K)	4	37 (31 for K)
K-1	Writing	3	18 (14 for K)	3 (2 for K)	24 (19 for K)
2-3	Listening	2	20	3	25
2-3	Speaking	4	12	3	19
2-3	Reading	3	30	5	38
2-3	Writing	2	17	3–4	22–23
4-5	Listening	2	20	3	25
4-5	Speaking	4	12	3	19
4-5	Reading	2	30	6	38
4-5	Writing	2	17	3–4	22–23
6-8	Listening	2	23	3	28
6-8	Speaking	4	12	3	19
6-8	Reading	2	30	6	38
6-8	Writing	2	17	3–4	22–23
9-12	Listening	2	23	3	28
9-12	Speaking	4	12	3	19
9-12	Reading	2	30	6	38
9-12	Writing	2	17	3–4	22–23

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Blueprint

The operational blueprint for LAS Links 3rd Edition is shown in Table 10. The blueprint is organized by skill area, language context strand, and subskill area. Operational item counts for each grade are shown, and MC items are distinguished from CR items. All new test forms for LAS Links will retain the same score scale as 1st and 2nd Editions.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 10. LAS Links 3rd Edition Blueprint: Listening Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (20 total)	2–3 # of Items (20 total)	4–5 # of Items (20 total)	6–8 # of Items (23 total)	9–12 # of Items (23 total)
Social, Intercultural, and Instructional Communication	Listen for Information	MC	8	8	8	9	9
Language Arts/Social Studies/History	Listen for Academic Instruction	MC	2	3	3	3	3
Language Arts/Social Studies/History	Listen for Information	MC	4	3	3	4	4
Mathematics/Science/Technical Subjects	Listen for Academic Instruction	MC	2	3	3	3	3
Mathematics/Science/Technical Subjects	Listen for Information	MC	4	3	3	4	4

Table 11. LAS Links 3rd Edition Blueprint: Speaking Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (10 total; 8 for K)	2–3 # of Items (12 total)	4–5 # of Items (12 total)	6–8 # of Items (12 total)	9–12 # of Items (12 total)
Social, Intercultural, and Instructional Communication	Make Conversation	CR	1	1	1	1	1
Social, Intercultural, and Instructional Communication	Describe and Request Information	CR	2	2	2	2	2
Social, Intercultural, and Instructional Communication	Tell a Story	CR	1	1	1	1	1
Language Arts/Social Studies/History	Use Academic Words	CR	1	1	1	1	1
Language Arts/Social Studies/History	Present and Explain Information	CR	2 (1 for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Academic Words	CR	1	1	1	1	1
Mathematics/Science/Technical Subjects	Present and Explain Information	CR	2 (1 for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 12. LAS Links 3rd Edition Blueprint: Reading Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (30 total; 24 for K)	2–3 # of Items (30 total)	4–5 # of Items (30 total)	6–8 # of Items (30 total)	9–12 # of Items (30 total)
Foundational Skills	Read Words	MC	12	6	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Read School Texts	MC	10 (8 for K)	10	14	14	14
Language Arts/Social Studies/History	Read Academic Texts	MC	2	5	6	6	6
Language Arts/Social Studies/History	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2
Mathematics/Science/Technical Subjects	Read Academic Texts	MC	2	5	6	6	6
Mathematics/Science/Technical Subjects	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2

Table 13. LAS Links 3rd Edition Blueprint: Writing Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (18 total; 14 for K)	2–3 # of Items (17 total)	4–5 # of Items (17 total)	6–8 # of Items (17 total)	9–12 # of Items (17 total)
Foundational Skills	Start Writing	Auto CR CR	2 5	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Use Grammar and Conventions	MC	4	6	6	6	6
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–3)	1	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–4)	N/A	1	1	1	1
Language Arts/Social Studies/History	Use Grammar and Conventions	MC	1	2	2	2	2
Language Arts/Social Studies/History	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Grammar and Conventions	MC	1	2	2	2	2
	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Item Development Process

According to the most recent edition of the *Standards for Educational and Psychological Testing* (AERA et al, 2014), “validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11). Essential validation evidence supporting the LAS Links assessments is produced during the item and test development process. Content-related validation evidence supports inferences from a sample of observations (i.e., the test) to a domain of observations (i.e., English language proficiency within the content domain). A substantial source of this validation evidence is gathered from expert judgement of whether the test items and tasks are an adequate and representative sample of the domains of content being measured. For LAS Links Forms E and F, expert judgement included, for example, educators involved in the item and test development process, members of the state departments of education where LAS Links is used for summative purposes, DRC test development specialists, and national consultant experts in English language proficiency.

As stated, content-related evidence of the validity of the intended test score interpretation in summative assessments is supported by a correspondence between test content and a specification of the content domain or reporting category. For LAS Links, evidence of content-related validity is demonstrated through consistent adherence to test blueprints and through high-quality item and test development processes that include reviews of items for their alignment to the LAS Links Standards Framework, appropriateness for the grade span and population of English learners, and freedom from issues of bias, fairness, or sensitivity.

The item and test development process requires a cohesive development approach blending what may appear to be discrete processes into a single, seamless development cycle. Those discrete processes include the development of test and item specifications and content blueprints, item writing, item editing, passage and/or stimulus creation, item reviews (by internal DRC reviewers, external consultants, and committees of educators), field test administrations, and data reviews, and the processes must be understood as a whole to understand the relationships between the parts. DRC’s model for the LAS Links development follows the *Standards for Educational and Psychological Testing* (AERA et al, 2014) since items are developed to reflect the range of cognitive ability inherent in the standards, resulting in reliable and instructionally valid tests. In addition, the item and test development process adheres to the Principles of Universal Design, and it reflects a clear understanding of how items and tests must lend themselves to accessibility by diverse groups of students and must function appropriately across a broad range of test administration accommodations.

Table 11 provides a summary of the major item and test development activities that occurred in order to develop the operational LAS Links Forms E and F test forms. This section also provides information regarding how DRC item and test developers engaged educators from states including Connecticut, Mississippi, Texas, Florida, New York, and

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

California in the process and followed rigorous procedures to develop and subsequently select items to be administered on the LAS Links Form E and F assessments.

This section is particularly relevant to addressing AERA, APA, and NCME (2014) Standards 4.0, 4.1, and 4.7 from Chapter 4 of the AERA, APA, and NCME (2014) Standards, "Test Design and Development." AERA, APA, and NCME (2014) Standard 4.0 states the following:

> Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population. (p. 85)

This section also addresses Standards 3.1, 3.2, 3.9, 4.12, and 7.4.

Table 14. Development Process for the LAS Links Forms E and F

Step	Description
1. Item and Test Development Planning (including development of test designs, blueprints, and draft item specifications)	DRC received feedback from potential state partners and advisory committees that were formed to review LAS Links Forms C and D to consider new forms development and make recommendations for the direction of future forms. Based on this feedback, DRC drafted a development plan, including test designs and desired blueprints from which draft item specifications and the item development plan were determined.
2. Development of Passages	DRC item and test development specialists provided LAS Links specific training to experienced passage developers who submitted passages for review and approval. Those passages were then reviewed by DRC item and test development specialists and editors, including bias, fairness, and sensitivity experts, and edited as necessary. Graphics and audio files were also developed and approved during this process. Acceptable passages with associated graphics were moved forward to the item writing phase.
3. Item Writing	DRC provided LAS Links specific training to item writing specialists who then wrote and submitted items; submitted items were entered into DRC's Item Banking System (IDEAS).
4. Editorial and Content Review of the Items; Graphics Creation	DRC item and test development specialists and editors, including bias, fairness, and sensitivity experts, reviewed and edited items as needed. Items were prepared for item review.
5. Item Review	Items were reviewed by experts in English language acquisition and assessment, including bias, fairness, and sensitivity experts. Each reviewer recommended that items be accepted as is, accepted with specific revisions, or rejected from the item pool.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Step	Description
6. Items Selected for Field Testing	The feedback from all reviewers was then reconciled by DRC item and test development specialists, and edits were incorporated as needed. Final content and editorial reviews were completed. DRC item and test development specialists then selected items to be placed into forms for the standalone field test.
7. Embedded Pilot of Newly Developed Items	The pilot of the items took place. Participating students were administered either Form C or Form D for operational scores as well as a pilot session for each participating domain. The performance of the items was analyzed, and the results were provided to DRC item and test development specialists.
8. Review Meeting with Connecticut Educators	DRC facilitators led a committee of educators through a review of items from LAS Links Forms C and D. Educators provided both general and item-specific feedback that was used to inform plans for additional item development and revisions to test directions.
9. Item Development Plan for Round 2 Finalized	Based on the results of the pilot and feedback from educators, including those participating in the review meeting in Connecticut, DRC item and test development specialists created item development plans for the second round of item development. These item development plans were reviewed and approved by senior members of the item and test development staff at DRC.
10. Development of Passages	DRC item and test development specialists provided LAS Links specific training to experienced passage developers who submitted passages for review and approval. Those passages were then reviewed by DRC item and test development specialists and editors, including bias, fairness, and sensitivity experts, and edited as necessary. Graphics and audio files were also developed and approved during this process. Acceptable passages with associated graphics were moved forward to the item writing phase.
11. Item Writing	DRC provided LAS Links specific training to item writing specialists who then wrote and submitted items; submitted items were entered into DRC's Item Banking System (IDEAS).
12. Editorial and Content Review of the Items; Graphics Creation	DRC item and test development specialists and editors, including bias, fairness, and sensitivity experts, reviewed and edited items as needed. Items were prepared for item review.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Step	Description
13. Item Review Meeting with Educators	DRC facilitated a virtual item review meeting with educators from Connecticut, Mississippi, Texas, Florida, New York, and California. Committee members participated in one grade-span committee: K−1, 2−3, 4−5, 6−8, or 9−12. Items were reviewed for content alignment (educators from Connecticut were asked to align items to Connecticut English Language Proficiency standards as well as LAS Links standards); grade-level appropriateness; level of difficulty; Universal Design, including bias, fairness, and sensitivity and appropriate language demand for the grade span and content area; and elements of technical design (including reviews of the correct answer and distractors as well as graphics for each item). Committees came to consensus regarding the status of each item: accepted as is, accepted with revisions (specified by the committee), or rejected.
14. Items Selected for the LAS Links Form E Operational Field Test	All feedback from the item review meeting was reviewed; edits suggested by the educators were incorporated as needed; final content and editorial reviews were conducted. DRC item and test development specialists selected items to be placed into embedded field test forms within the Form E Operational Field Test. Additionally, adhering to the approved test designs, blueprints, item specifications, and guidelines for item analysis and forms construction, items previously used on Forms C or D or in field test positions within the pilot were selected for the operational positions upon which student scores were based.
15. Administration of LAS Links Form E Operational Field Test	The LAS Links Form E Operational Field Test was prepared for administration in Spring 2021. These forms were administered in Mississippi in Spring 2021. The forms were ready for administration in Spring 2021 in Connecticut; however, due to lingering concerns about the pandemic, the administration of the Form E Operational Field Test was delayed until Spring 2022.
16. Data Review Meeting with Educators	DRC facilitated a virtual data review meeting with educators from Connecticut and Mississippi. Committee members participated in one grade-span committee: K−5, 6−8, or 9−12. Educators reviewed item content and alignment, informed by student response data, to determine whether each item was acceptable for use on a LAS Links assessment or should be rejected. For certain items that had appeared on the K−1 assessment, educators could also determine that the item was appropriate for first grade students while rejecting it for kindergarten students.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Step	Description
17. Items Selected for the LAS Links Form F Operational Field Test	Adhering to the approved test designs, blueprints, item specifications, and guidelines for item analysis and forms construction, items previously used on Forms C or D or in field test positions within LAS Links Operational Field Test Form E were selected for the operational positions upon which student scores were based. Additional items were selected for research purposes to fill the field test positions.
18. Administration of LAS Links Form F Operational Field Test	The LAS Links Form F Operational Field Test was administered in Connecticut in Spring 2023.
19. Items Selected for the LAS Links Form E	Minor adjustments were made to the operational selection of items for LAS Links Form E per educator feedback received during the data review. The operational selection was again confirmed to adhere to the approved test designs, blueprints, item specifications, and psychometric guidelines for item analysis and forms construction. Additional items were included in embedded field test positions in order to facilitate future form maintenance.
20. Administration of LAS Links Form E	The LAS Links Form E was administered in Connecticut in Spring 2024.

Listening Test Considerations

Brown (1995) provides a very useful set of Cognitive Load Principles, which we have adapted below.

- Less is more. It is easier to understand a text involving fewer individuals, characters, or objects. As the number of people or things involved in a Listening passage increases so does the likelihood of confusion, even for native speakers, as these details must be retained in short-term memory.
- Distinguish between interactants. It is less cognitively demanding to understand a text (e.g., narrative, description, instruction) involving individuals and objects that are easily distinguishable from one another. It would be easier to understand and remember story details about a dump truck, an ambulance, and an SUV than one about three sedans. “The more individuals and objects are similar and the more they are described in similar terms, the more likely they are to be misidentified . . .” (Brown, 1995, p. 63).
- There’s no “there there.” It is easier to understand texts that involve uncomplicated spatial relations. When we listen to a story, we construct a mental model of the scene and use this model as a stage on which to place the people and things and observe their actions. The simpler the spatial relations, the easier it is to visualize them. The same can be said for temporal relationships.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- Straight talk. It is easier to understand texts when the order of telling matches the order in which the events occurred. As they listen to a narrative, listeners assume the events happened in the order reported. Such narratives are easier to understand because they require less manipulation of the listeners' mental model and do not overburden short-term memory.
- Be clear. It is easier to understand a text if relatively few inferences are necessary to relate each sentence to the preceding text. In other words, avoid ambiguity and obscurity, and be clear with respect to orderliness. In the English spoken in the United States, the rhetorical style is for more general details to precede more specific ones and for causes to precede effects to avoid non-linear narratives. On the other hand, a related pitfall we find is that in attempting to "simplify" texts, some writers make the mistake of eliminating detail to shorten sentences. An analysis by Beck, McKeown, Sinatra, & Loxterman (1991) has shown that texts that present only facts with little explanation of their relationship are more difficult to comprehend than texts that provide more elaboration on how the material is connected. One reason for this may be that the lack of elaboration puts the onus of drawing all such inferences on the reader. Likewise, some syntactically simple texts are difficult to comprehend because the text is poorly organized. The following excerpt from a reading passage exemplifies this point:

  > A house on stilts is high above the ground. People build houses on stilts in Thailand. Thailand is a country in Southeast Asia. The weather in Thailand is very warm. Air can move around well in a house on stilts. In the summer the country has heavy rainstorms. These heavy storms are called monsoons. One area of Thailand gets the most rain.

- The sentences are syntactically straightforward and, with the exception of "stilts," the vocabulary consists of mostly common words; however, the text lacks coherence, making it very difficult for even a skillful reader to understand. As a result, the excerpt is easy to read but not easy to comprehend.
- Expect the expected. It is easier to understand a text if the information is consistent and fits with the listener's pre-existing knowledge. Thus, it is easier to follow a narrative about a topic we already know well than one we know nothing about. It is, for example, a standard gambit to open a conversation by setting a common point of reference, such as, "Remember the time we went to Lake Revelstoke?" This strategy ensures that everyone involved in the conversation starts from the same point of reference and listeners can retrieve the shared background knowledge necessary for comprehension. If the information that follows is new but compatible with the old, it is easier for listeners to incorporate it into their knowledge system. According to Brown (1995), problems arise for listeners when the incoming information is ambiguous, expressed vaguely, or is not compatible with the listener's existing knowledge.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Speaking Test Considerations

- Specifications to develop stimuli for items in the Present and Explain Information subtest provided guidance to ensure test authenticity.
- For items in the Tell a Story subtest, art development was specified so that
    - distinct events occurred in each illustration.
    - actions were easily understood visually.
    - the depicted events and actions had a beginning, middle, and end.
    - interpretation did not rely on facial expressions, gestures, or body stance to convey action or meaning, for example, “He is looking at the pear and the apple but can’t decide which one to choose.” These subtle clues could be tied to a particular culture unknown or unfamiliar to the student. If the narrative relied on these types of clues, students may not understand what is happening and may stop their telling of the story. This could lead the test administrator to erroneously assume that the student lacks the language skills necessary to tell the story when the cause could in fact be more a matter of the pictures not telling the story.

Reading Test Considerations

- The LAS Links Reading test was designed to reflect the expectations that all K—12 students read and comprehend more complex texts, including English Learners/Multilingual Learners (ELs/MLs). Thus, the criteria for developing extended passages were to ensure that texts were comparable to mainstream classroom texts currently in use throughout the United States in terms of construction, complexity, and appearance.
- ACT, Inc (2006) provided guidance on defining degrees of text complexity. It categorized texts as being Uncomplicated, More Challenging, or Complex. For the purposes of text development for LAS Links, the extended texts were written to align most closely to the More Challenging category, which is defined by the following text features: implicit relationships, detailed richness, involved structure, and a context-dependent use of some more complex vocabulary.
- Reading dichotomous constructed-response items consist of a chart, table, or diagram with missing information for students to complete. To avoid raters having to make a judgment as to the veracity of the response, students are required to enter the information exactly as it appears in the text to improve reliability. The information required to respond is contained in the passage to ensure that students will not need to rely on background knowledge to complete the table.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Writing Test Considerations

- As there are far more assessable features of language than there are test items, developers were asked to target the essential aspects of syntax and mechanics. For guidance, they referred to the skills explicitly noted in the CCSS (2010) for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.
- The texts in the Writing prompts are designed to be grade-level appropriate and not require new background knowledge. This is crucial to assess students' language proficiency skills, instead of their content knowledge of specific topics.

Item Selection and Form Assembly

Items selected for the 3rd Edition came primarily from the operational field tests for Forms E and F, with a small number of items used from Forms C and D. Any items that failed to meet the established psychometric criteria were not selected. The detailed content and psychometric criteria are listed in Table 15.

Table 15. Item Selection and Form Assembly Criteria

Aspects	Criteria
Test Blueprint	adhere to sub-skill category quotas to ensure content coverage.
Item Difficulty	minimize the number of items with *p*-values <=0.20 or >= 0.95.
Item-Total Correlation	minimize the number of items with item-total correlations <0.15 and MC items with any of the distractor point-biserial >0.05.
Item Omit Rate	minimize the number of items with omit rates >=5%.
Test Information	maximize test information at and around the LAS Links cut scores with the target test information equal to or greater than that on the operational Forms A and B.
Standard Error of Measurement	minimize standard error of measurement for the target student ability span at each grade span.
IRT Model Fit	minimize the number of poor-fitting items.
DIF	avoid inclusion of items with C+ or C- DIF
Distribution of MC Answer Key Positions	distribute MC answer key positions should be evenly distributed throughout the form in general and should avoid the same position being repeated consecutively.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Aspects	Criteria
Item Selection Review	cues as to the correct answer from one item to another, context redundancy, presence of clang (distractors not unique from one another), diversity of names and artwork for gender and ethnicity, time to complete test consistent with established recommendations.

Quality Control Evidence

Items were reviewed for adherence to the item-writing specifications, which included developmental appropriateness, item difficulty, freedom from areas of potential bias, and appropriate answer choices and distractors on the basis of both content considerations (e.g., expert reviews) and statistical evidence as noted in Table 12. Additionally, items went through a thorough review by internal and external review panels for bias and sensitivity. During form development, items and the overall test construction were reviewed for considerations of Universal Design principles, including equitable use, flexibility in use, simple and intuitive use, perceptible information, tolerance for error, low physical effort, and size and space for approach and use.

Item Review Criteria

To ensure appropriate cognitive demands and readability, DRC developed items with the following criteria in mind:

- Vocabulary that is consistently at or below the targeted grade level, verified by graded word lists such as *EDL Core Vocabularies* and *The Living Word*.
- Syntax that is grade-level appropriate.
- Information that is necessary for assessing the skill or knowledge being tested is the only information given.
- Detailed directions or large amounts of text divided into steps, sections, or bulleted lists to help students understand the task.
- Key words or phrases that are presented in a consistent style to make the task clear for the student.
- After items were written, DRC content specialists reviewed the material for standards alignment, grade-level-appropriateness, item difficulty, freedom from areas of potential bias, and appropriate answer choices and distractors. Every item underwent at least two reviews by the content specialists to ensure the following:
    - Item correspondence to the identified standard and construct.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- Relevance of each item to the purpose of the test.
- Correspondence to the principles of quality item development.
- Appropriate item difficulty.
- Accuracy of content presented in the item.
- Appropriateness of language, graphics, artwork, charts, and figures.

LAS Links items must demonstrate a match to the LAS Links Standards Framework, high technical quality, and appropriate difficulty; the items must provide appropriate alternative choices (distractors) in MC items and the item writers must provide complete answers for open-ended questions.

Match to LAS Links Standards Framework

Each item had to demonstrate a specific match to the selected standard. The item writers were required to establish the close correspondence between the standards and the test questions clearly; this correspondence was verified by DRC content editors and development supervisors. This step represented the first verification of the content validity of each item.

Technical Quality

For MC items, technical quality included fully stated stems (i.e., the stem states a complete question so that the student understands what is asked before reading the response options); balanced response options (no answer choice is conspicuous due to length, syntax, tone, level of specificity, or other reason); plausible and reasonable distractors; absence of cueing between stem and answer choices; brevity; and clarity. For OE items, technical quality included precisely and fairly stated prompts that yield appropriate responses and well-formed and effective scoring rubrics and sample student responses.

Difficulty Level

Items were reviewed to ensure an appropriate difficulty level for the purpose of the test. DRC's development team kept a record of the estimated difficulty of each item to ensure that items were written within a specific range of difficulty in any given test.

Appropriate Distractors for Multiple-Choice Items

Item writers submitted answer keys with their MC items. Writers were directed to double check distractors to verify that no ambiguous or misleading incorrect response options existed, that there was only one clear correct answer per item, and that answer choices did not include outliers. DRC content editors and development supervisors then verified the correct responses.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Complete Answers for Open-Ended Items

When writing OE items, the writers provided a correct and complete answer, as well as a range of answers possible for each item. In addition, both the writers and the reviewers examined every item to ensure that none invited a discussion of the personal beliefs or practices of a student or student's family. Any such items were immediately revised or rejected.

Development supervisors, content editors, and item writers further refined items collaboratively until all items met or exceeded both DRC's high standards and the criteria in the specifications. All items developed for the LAS Links pool went through this exacting process.

Universal Design

Assessments that are universally designed allow the widest possible range of students to participate, resulting in more valid inferences about students' performance. Universally designed assessments may reduce the need for accommodations by decreasing or eliminating access barriers associated with the tests themselves.

The Principles of Universal Design were incorporated throughout the item development process to allow participation of the widest possible range of students who would be taking LAS Links® Forms E or F. During the item writing and subsequent item review, educators and test development specialists were provided with information regarding the Principles of Universal Design and how items and eventually operational assessments need to adhere to the principles.

Table 16 provides an overview of the information regarding the Principles of Universal Design and the guidelines provided by the National Center for Educational Outcomes (NCEO) (Thompson, Johnstone & Thurlow, 2002). As an integral part of the item writer training, this information was provided to the educator and test development specialist item writers and reviewers. The elements of universal design are relevant to both item development and form construction. Through the use of the principles, validity evidence is provided to document that the items and subsequently the assessments have been designed to measure the knowledge and skills across the full achievement continuum described in the content standards and that the assessments are fair for all students at all levels of proficiency.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 16. Elements of Universal Design

Element	Explanation
Inclusive Assessment Population	Tests designed for state, district, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field-testing procedures.
Precisely Defined Constructs	The specific constructs tested must be clearly defined so that all construct-irrelevant cognitive, sensory, emotional, and physical barriers can be removed.
Accessible, Nonbiased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amendable to Accommodations	The test design facilitates the use of needed accommodations.
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and presented in understandable language.
Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words are kept to a minimum) to produce readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, tables, figures, illustrations, and response formats.

In addition to the Principles of Universal Design and the NCEO guidelines, DRC also considers the Frameworks for Universal Design for Computer-Based Testing (UD-CBT) and Universal Design for Learning in providing training to educators and test development specialists contributing to the development of the LAS Links Forms E and F. These guidelines specify how digital technologies can be used to create tests that more accurately assess students who possess a diverse range of physical, sensory, and cognitive abilities and challenges. UD-CBT has been found to level the playing field for students with disabilities and for English learners.

In adherence with the federal Individuals with Disabilities Education Act (IDEA) of 2004, DRC trained item writers to adhere to the Principles of Universal Design, and DRC's team of content item and test development staff also incorporated the principles into the design and development of the assessment delivery system, DRC INSIGHT, ensuring that the system used to deliver LAS Links Forms E and F is also accessible to the widest possible range of students.

<u>Considerations for the Speaking Test</u>

- Specifications to develop stimuli for items in the Present and Explain Information subtest provided guidance to ensure test authenticity.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- For items in the Tell a Story subtest, art development was specified so that:
    - distinct events occurred in each illustration.
    - actions were easily understood visually.
    - the depicted events and actions had a beginning, middle, and end; and,
    - art specifications did not rely on facial expressions, gestures, or body stance to convey action or meaning. These subtle clues could very well be tied to a particular culture unknown or unfamiliar to the student. If the narrative relied on these types of clues, students may not understand what is happening, and may stop their telling of the story. This could lead the test administrator to erroneously assume that the student lacks the language skills necessary to tell the story when the cause could in fact be more a matter of the pictures not telling the story.

Considerations for the Reading Test

- Given the expectation that all K−12 students read and comprehend more complex texts, including English learners (ELs), the LAS Links Reading test was designed to reflect this trend. Thus, the criteria for developing extended passages were to ensure that texts were comparable to mainstream classroom texts currently in use throughout the United States in terms of construction, complexity, and appearance.
- ACT, Inc. (2006) provided guidance on defining degrees of text complexity. It categorized texts as being Uncomplicated, More Challenging, or Complex. For the purposes of text development for LAS Links, the extended texts were written to align most closely to the More Challenging category, which is defined by the following text features: implicit relationships, detailed richness, involved structure, and a context-dependent use of some more complex vocabulary.

Considerations for the Writing Test

- As there are far more assessable features of language than there are test items, developers were asked to target the essential aspects of syntax and mechanics. For guidance, they referred to the skills explicitly noted in the CCSS (2010) for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.
- The texts in the Writing prompts are designed to be grade-level appropriate and not requiring new background knowledge. This is crucial in order to assess students' language proficiency skills instead of their content knowledge of specific topics.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Content and Bias/Sensitivity Review

At every stage of the item and test development process, DRC employs procedures that are designed to ensure that items and tests adhere to Standard 7.4 of the *Standards for Educational and Psychological Testing* (AERA et al, 2014).

> Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain. (p. 82)

To meet Standard 7.4, DRC employs a series of internal and external quality steps. DRC provides specific training for test developers, graphic artists, editors, item writers, and reviewers on how to write, review, revise, and edit items for issues of bias, fairness, and sensitivity (and for technical quality).

Training also includes an awareness of and sensitivity to issues of cultural diversity. In addition to providing internal training in reviewing items to eliminate potential bias, DRC's guidelines for bias, fairness, and sensitivity include instruction concerning how to eliminate language, symbols, words, phrases, and content that might be considered offensive by members of racial, ethnic, gender, or other groups. Areas of bias that are specifically targeted include, but are not limited to, stereotyping, gender, regional/geographic, ethnic/cultural, socioeconomic/class, religious, and biases against a particular age group (ageism) or persons with disabilities. In the development of the LAS Links assessments, DRC also noted topics that should be avoided and maintained a balance of gender and ethnic emphasis within the pool of available items and passages. DRC's guidelines were provided at all stages of the item development cycle, from item writing to editing and reviews.

Below are the criteria used by external reviewers for evaluating content and potential bias/sensitivity issues.

- The item is appropriate to the grade span.
- The item is aligned to the indicator.
- The item has accurate content.
- The item has a single correct answer (selected-response (SR) items only).
- The item has an accurate rubric (CR items only).
- The item has accurate art (graphics only).
- The item has clear graphics (graphics only).
- The item adheres to Universal Design principles.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- The item is free from bias and sensitivity issues.

Below is a list of the major areas of assessment that were reviewed.

- passages
- artwork
- item questions
- distractors in SR items
- cognate “Say words” items (potentially favoring Spanish speakers)
- interchangeable items (where items need to be scored together and answers are interchangeable)
- scoring rubrics/sample answers
- item/test directions
- standards alignment

The LAS Links items were also reviewed by national bias, fairness, and sensitivity experts. The national English Learner bias, fairness, and sensitivity reviewers DRC uses to review its LAS Links items collectively have a vast array of experience in education, providing them with diverse perspectives. All reviewers are experienced in the review of passages and items for bias, fairness, and sensitivity issues and for adherence to the Principles of Universal Design. Their perspectives and experiences include knowledge of special populations, such as English language learners, students with disabilities, and ethnically and culturally diverse students. The reviewers have backgrounds in the following professions: English Learner classroom teacher, English Learner curriculum specialist, content-area instructional specialist, test development editor, and university professor. To provide a national and diverse perspective, reviewers are also located around the country in California, Connecticut, Florida, Minnesota, New York, and Texas.

The national reviewers of the LAS Links items also identified issues that could negatively affect a student’s ability to elicit valid evidence about an assessment target or to access passages or items. During the bias, fairness, and sensitivity review, for example, reviewers were specifically tasked with identifying content in the passages or items that could negatively affect a student’s ability to produce a correct response because of the student’s background. In addition, reviewers also checked the accuracy of the content, answer keys, and scoring materials. Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised by DRC item and test development specialists to address the issues identified by the reviewers, or the items were removed from the pool of items.

Data Review

The main purpose of the data review was to use item level statistics derived from the operational field testing to identify items requiring additional content review before they might be considered for selection on the LAS Links 3rd Edition forms. The following

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

presents information about the process used for data review, how items were identified for further review, and the results and outcomes of the data review.

The data review process employs three approaches to statistically evaluate flagged items for review: evaluation of item difficulty, discrimination, and Differential Item Functioning (DIF). Collectively, these analyses are referred to as classical item analyses (CIA) and their results, unlike IRT item statistics, are straightforward for content reviewers to understand and apply in their review of items. Readers are referred to Chapter 4 for detailed descriptions of the item statistics used here.

Criteria for Identifying Items

All field test items were analyzed statistically using conventional item-analysis methods. For MC items, classical item statistics included the corrected point-biserial correlation (Pt. Bis.) for the correct and incorrect responses (distractors), percent correct (p-value), and the percent responding to incorrect responses. For CR items, the statistical indices included the item-test correlation, the point-biserial correlation for each score point, the percent of responses at each score point, and the percent of non-scoreable responses.

In general, more capable students are expected to respond correctly to easy items and less capable students are expected to respond incorrectly to difficult items. If either of these situations does not occur, the item will be reviewed by DRC test development staff and committees of educators to determine the nature of the problem and the characteristics of the students affected. The primary way of detecting such conditions is through the point-biserial correlation. In each case the statistic will be positive if the total test mean score is higher for the students who respond correctly to MC items (or attain a higher CR item score) and negative when the reverse is true.

The following set of criteria was used to identify items for additional review.

For an MC item to be flagged, the criteria included any of the following:

- Percent correct (p-value) less than 0.2 or greater than 0.95,
- Point-biserial correlation for the correct response less than 0.15,
- Point-biserial correlation for any incorrect response greater than 0.0,
- Gender DIF code of B-, B+, C-, or C+.

For a CR item to be flagged, the criteria included any of the following:

- Adjusted p-value less than 0.2 or greater than 0.95,
- Point-biserial correlation less than 0.15,
- Gender DIF code of B-, B+, C-, or C+.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Review of Items with Data

In the preceding section, it was stated that test development content-area specialists used certain statistics from CIA and DIF analyses of the 2021—2022 field test to identify items for review by educators. Items not identified for this review had good statistical characteristics and, consequently, were entered into the eligible pool for future item selection. Likewise, items of extremely poor statistical quality were regarded as unacceptable and needed no additional review. DRC content-area test development specialists and DRC psychometric specialists identified the remaining items for further review by a committee of educators. The intent was to capture all items that needed a closer review; thus, the criteria employed tended to over-identify rather than under-identify items.

The review of the items with associated data was conducted by 15 educators (teachers, curriculum specialists, and administrators) broken out into grade-span committees. Demographic and background information for the participants in each committee is included in Tables 17 and 18 below.

Table 17. Grades K—5 Committee Participants Demographic and Background Information

Reviewer Number	Gender	Race/ Ethnicity	Background
1	Female	White	Classroom teacher for English Learners
2	Female	White	Instructional coach and teacher consultant for EL acquisition
3	Female	White	English Learner/Bilingual Coordinator
4	Female	White	English Learner/Bilingual Coordinator and Special Education Coordinator
5	Female	White	Instructional coach and teacher consultant for EL acquisition
6	Male	White	Consultant for a State Department of Education

Table 18. Grades 6—12 Committee Participants Demographic and Background Information

Reviewer Number	Gender	Race/ Ethnicity	Background
1	Female	White	Classroom teacher for English Learners and Special Education
2	Female	White	English Learner/Bilingual Coordinator

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Reviewer Number	Gender	Race/ Ethnicity	Background
3	Female	Hispanic	Classroom teacher for English Learners
4	Female	White	Classroom teacher for English Learners
5	Female	White	Instructional coach and teacher consultant for English language acquisition
6	Female	White	Classroom teacher for English Learners
7	Female	White	English Learner/Bilingual Coordinator
8	Female	Hispanic	District Coordinator for English Learners
9	Female	White	English Language Arts Consultant for a State Department of Education

The review took place in June 2022 virtually using Zoom. The committee reviewing the items for the K—1, 2—3, and 4—5 grade bands met over the course of three full-day meetings within the same week. The committee reviewing the items for the 6-8 and 9-12 grade bands met for two full-day meetings within the same week. In these sessions, committee members were first introduced to the processes of developing the LAS Links assessments and were then trained by a representative from DRC's psychometrics staff with regard to the statistical indices used in item evaluation. This was followed by a discussion with examples concerning reasons that an item might be retained regardless of the statistics.

The committee review process involved a brief exploration of possible reasons for the statistical profile of an item (e.g., possible bias, grade appropriateness, instructional issues) and a decision regarding acceptance. DRC content-area test development specialists facilitated the review of the items. Each committee reviewed the flagged items from the pool of field tested items and made recommendations on each item and/or passage. Note that items were flagged separately at grades K and 1, but if they were flagged at one grade, the data were reviewed for both grades. The educator committee had the options of accepting an item for both grades, accepting the item at only grade 1, or rejecting the item at both grades. No items could be accepted for kindergarten if they were not also accepted at grade 1. For all other grade spans, data from all grades within the span were combined and a single set of item parameters was reviewed. Data review details and results are shown in Tables 16 and 17.

Results

DIF analyses were conducted on the 2021−2022 field test items. The number of items from each subject and grade span that were assigned to each severity code is shown in Table 16. In Listening, there were a total of five items identified as having more than

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

negligible (B) DIF for gender. In Reading, there were eight items identified as having more than negligible (B) DIF and one item with sizable (C) DIF for gender. No Speaking items were identified as having either more than negligible (B) or sizable (C) DIF for gender. In Writing, four items were identified as having more than negligible (B) DIF and two items with sizable (C) DIF for gender.

Note that items flagged for p-value and/or point-biserials are summarized together in Table 19.

Table 19. 2022 Data Review Committee Results

Subject	Grade Span	Total No. of Items in FT	Reviewed MC	Reviewed CR	Reviewed DIF Only	Total Reviewed	No. of Items Rejected by Committee	No. of Items Classified as Rejected[1]
Listening	K	23	8	NA	0	8	2	6
	1	23	8	NA	0	8	2	6
	2–3	24	8	NA	1	8	0	0
	4–5	27	5	NA	1	5	1	2
	6–8	30	9	NA	0	9	4	4
	9–12	24	3	NA	2	3	0	0
Reading	K	42	31	NA	0	31	7	7
	1	42	31	NA	0	31	3	3
	2–3	43	3	NA	1	3	1	1
	4–5	58	9	NA	4	9	3	3
	6–8	57	16	NA	1	16	8	9
	9–12	58	3	NA	3	3	0	2
Speaking	K	15	NA	2	0	15	0	1
	1	15	NA	2	0	15	0	1
	2–3	16	NA	0	0	0	0	0
	4–5	16	NA	0	0	0	0	0
	6–8	16	NA	0	0	0	0	0
	9–12	16	NA	1	0	1	1	1

[1] *Items classified as "Rejected" from 2021—2022 field test (all sources: data review committee, rejected prior to committee review)*

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Subject	Grade Span	Total No. of Items in FT	Reviewed MC	Reviewed CR	Reviewed DIF Only	Total Reviewed	No. of Items Rejected by Committee	No. of Items Classified as Rejected[1]
Writing	K	22	5	8	0	13	1	1
	1	22	5	8	1	13	1	1
	2–3	21	2	1	0	3	0	0
	4–5	27	0	1	1	1	0	0
	6–8	27	0	1	1	1	0	0
	9–12	28	1	3	3	4	1	2

Table 20. DIF Summary for Male/Female

Subject	Grade Span	B+	B-	C+	C-	Total
Listening	K	-	-	-	-	0
	1	-	-	-	-	0
	2–3	2	-	-	-	2
	4–5	-	1	-	-	1
	6–8	-	-	-	-	0
	9–12	1	1	-	-	2
Reading	K	-	-	-	-	0
	1	-	-	-	-	0
	2–3	-	-	1	-	1
	4–5	2	2	-	-	4
	6–8	-	1	-	-	1
	9–12	1	2	-	-	3
Speaking	K	-	-	-	-	0
	1	-	-	-	-	0
	2–3	-	-	-	-	0
	4–5	-	-	-	-	0
	6–8	-	-	-	-	0
	9–12	-	-	-	-	0
Writing	K	-	-	-	-	0
	1	1	-	-	-	1
	2–3	-	-	1	-	1
	4–5	1	-	-	-	1
	6–8	1	-	-	-	1
	9–12	1	-	1	-	2

Validity Based on Linguistic Processes

Knowledge of the theory and practice of cognitive labs has grown substantially since the original development of LAS Links. Meanwhile, the body of evidence demonstrating the validity of LAS Links test scores has been collected over time, and over the

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

development of new forms. Although the exploration of response processes has not been conducted through the application of cognitive labs, several components of the larger body of evidence of LAS Links test score validity provide relevant evidence of the validity of linguistic processes.

First, item discrimination as measured by point-biserial correlations and DIF are measures that are sensitive to unexpected response patterns. Where unexpected response patterns are present, items would likely be flagged for low discrimination or DIF. As noted, selecting items during the form construction processes involves a purposeful focus on selecting items that meet criteria for well-discriminating items that do not favor or disfavor particular student groups. Refer to Chapter 4 for details regarding the results of CIA and DIF analyses for the final LAS Links 3rd Edition item selections.

Further, items that were flagged during item analyses were subsequently evaluated by content experts and by US educators for their appropriateness to include in the LAS Links 3rd Edition assessments during the data review previously described. The items that did not pass this data review were not considered for inclusion.

Also, as noted previously, panels of English language development and assessment experts have been continually consulted over time to provide recommendations for improvement during new forms development. The focus of these efforts was to identify opportunities to improve the content and format of items such that the responses collected from students provide accurate representations of their English language proficiency without interference from construct irrelevant features. As previously discussed, findings from these targeted reviews have been addressed directly in the test design and blueprints for the LAS Links 3rd Edition.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 3: Field Testing: Design, Samples, Administration, Scoring, Item Calibration and Equating

This chapter provides information about field testing the content for LAS Links 3rd Edition. Field test designs, samples, test administration, scoring processes, scoring quality, and item calibration and equating are described.

Operational Field Testing

The new LAS Links content developed as described in Chapter 2 was administered as two operational field tests, in two states, over three consecutive Spring test administrations. Operational field testing in this case allowed for the collection of valid and reliable test data while simultaneously implementing the intended overall enhancements to the test design. This design ultimately supported optimization of the balance between test length for individual students and the time required to deliver LAS Links 3rd Edition on a useful and reasonable timeline following the impacts of the COVID-19 pandemic on schools and large-scale assessment.

The use of an operational field test approach for the development of the LAS Links 3rd Edition also helped avoid some well-known challenges associated with student motivation in standalone field test contexts and with the time required to collect sufficient data for scaling and equating through embedded field testing of new content.

Field Test Designs

The LAS Links Form E operational field test consisted of existing items that had been modified based on educator feedback and newly developed items following the processes described in Chapter 2. The field tests also necessarily included items from LAS Links 2nd Edition for the purpose of linking the 3rd Edition content to the existing LAS Links score scales described in the LAS Links 1st and 2nd Edition Technical Manuals (CTB, 2006 & 2013). Refer to Chapter 4 for technical details regarding the equating procedures applied.

Form E Operational Field Test Design

The LAS Links Form E operational field test was organized by grade and skill area. Core operational items were distinguished from embedded field test items for test item review purposes. Embedded field test items facilitated ongoing improvement and monitoring of the tests and the development of the Form F operational field test.

Tables 18—21 depict the overall test designs and blueprints for both the Forms E and Form F operational field tests, from which the final 3rd Edition content was selected.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 21. LAS Links Form E Operational (OP) Field Test Design

Grades	Skill Area	Number Practice Items	Number OP Items	Number Embedded FT Items	Total Number Items (Practice + OP + FT)
K-1	Listening	2	20	3	25
	Speaking	4	10 (8 for K)	3	17 (15 for K)
	Reading	3	30 (26 for K)	4	37 (33 for K)
	Writing	3	18 (14 for K)	3 (2 for K)	24 (19 for K)
2-3	Listening	2	20	3	25
	Speaking	4	12	4	20
	Reading	3	30	5	38
	Writing	2	17	2–3	21–22
4-5	Listening	2	20	3	25
	Speaking	4	12	4	20
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23
6-8	Listening	2	23	4	29
	Speaking	4	12	4	20
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23
9-12	Listening	2	23	3	28
	Speaking	4	12	4	20
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23

**Note: All tests are untimed; estimated administration times are provided for planning purposes.*

LAS Links Form E Operational Field Test Blueprint

The operational blueprint for the LAS Links Form E Operational Field Test is shown in Table 18. The blueprint is organized by skill area, language context strand, and sub-skill area. Operational item counts for each grade are shown, and MC items are distinguished from CR items.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 22. LAS Links Form E Operational Field Test Blueprint: Listening Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (20 total)	2–3 # of Items (20 total)	4–5 # of Items (20 total)	6–8 # of Items (23 total)	9–12 # of Items (23 total)
Social, Intercultural, and Instructional Communication	Listen for Information	MC	8	8	8	9	9
Language Arts/Social Studies/History	Listen for Academic Instruction	MC	2	3	3	3	3
Language Arts/Social Studies/History	Listen for Information	MC	4	3	3	4	4
Mathematics/Science/Technical Subjects	Listen for Academic Instruction	MC	2	3	3	3	3
Mathematics/Science/Technical Subjects	Listen for Information	MC	4	3	3	4	4

Table 23. LAS Links Form E Operational Field Test Blueprint: Speaking Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (10 total; 8 for K)	2–3 # of Items (12 total)	4–5 # of Items (12 total)	6–8 # of Items (12 total)	9–12 # of Items (12 total)
Social, Intercultural, and Instructional Communication	Make Conversation	CR	1	1	1	1	1
Social, Intercultural, and Instructional Communication	Describe and Request Information	CR	2	2	2	2	2
Social, Intercultural, and Instructional Communication	Tell a Story	CR	1	1	1	1	1
Language Arts/Social Studies/History	Use Academic Words	CR	1	1	1	1	1
Language Arts/Social Studies/History	Present and Explain Information	CR	2 (1 for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Academic Words	CR	1	1	1	1	1
Mathematics/Science/Technical Subjects	Present and Explain Information	CR	2 (1 for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 24. LAS Links Form E Operational Field Test Blueprint: Reading Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (30 total; 24 for K)	2–3 # of Items (30 total)	4–5 # of Items (30 total)	6–8 # of Items (30 total)	9–12 # of Items (30 total)
Foundational Skills	Read Words	MC	12	6	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Read School Texts	MC	10 (8 for K)	10	14	14	14
Language Arts/Social Studies/History	Read Academic Texts	MC	2	5	6	6	6
Language Arts/Social Studies/History	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2
Mathematics/Science/Technical Subjects	Read Academic Texts	MC	2	5	6	6	6
Mathematics/Science/Technical Subjects	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2

Table 25. LAS Links Form E Operational Field Test Blueprint: Writing Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (18 total; 14 for K)	2–3 # of Items (17 total)	4–5 # of Items (17 total)	6–8 # of Items (17 total)	9–12 # of Items (17 total)
Foundational Skills	Start Writing	Auto CR CR	2 5	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Use Grammar and Conventions	MC	4	6	6	6	6
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–3)	1	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–4)	N/A	1	1	1	1
Language Arts/Social Studies/History	Use Grammar and Conventions	MC	1	2	2	2	2
Language Arts/Social Studies/History	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Grammar and Conventions	MC	1	2	2	2	2
	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Form F Operational Field Test Design

The LAS Links Form F Operational Field Tests were organized by grade and skill area. Core operational items were distinguished from embedded field test items for test item review purposes. These distinctions are depicted in Table 20.

Table 26. LAS Links Form F Operational Field Test Design

Grades	Skill Area	# of Practice Items	# of OP Items	# of Embedded FT Items	Total # of Items (Practice + OP + FT)
K-1	Listening	2	20	3	25
	Speaking	4	10 (8 for K)	3	17 (15 for K)
	Reading	3	30 (24 for K)	4	37 (31 for K)
	Writing	3	18 (14 for K)	3 (2 for K)	24 (19 for K)
2-3	Listening	2	20	3	25
	Speaking	4	12	3	19
	Reading	3	30	5	38
	Writing	2	17	3–4	22–23
4-5	Listening	2	20	3	25
	Speaking	4	12	3	19
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23
6-8	Listening	2	23	3	28
	Speaking	4	12	3	19
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23
9-12	Listening	2	23	3	28
	Speaking	4	12	3	19
	Reading	2	30	6	38
	Writing	2	17	3–4	22–23

LAS Links Operational Forms E and F Blueprint

The operational blueprint for LAS Links Forms E and F is shown in Table 21. The blueprint is organized by skill area, language context strand, and subskill area. Operational item counts for each grade are shown, and MC items are distinguished from CR items.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 27. LAS Links Form E and F Operational Blueprint: Listening Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (20 total)	2–3 # of Items (20 total)	4–5 # of Items (20 total)	6–8 # of Items (23 total)	9–12 # of Items (23 total)
Social, Intercultural, and Instructional Communication	Listen for Information	MC	8	8	8	9	9
Language Arts/Social Studies/History	Listen for Academic Instruction	MC	2	3	3	3	3
Language Arts/Social Studies/History	Listen for Information	MC	4	3	3	4	4
Mathematics/Science/Technical Subjects	Listen for Academic Instruction	MC	2	3	3	3	3
Mathematics/Science/Technical Subjects	Listen for Information	MC	4	3	3	4	4

Table 28. LAS Links Form E and F Operational Blueprint: Speaking Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (10 total; 8 for K)	2–3 # of Items (12 total)	4–5 # of Items (12 total)	6–8 # of Items (12 total)	9–12 # of Items (12 total)
Social, Intercultural, and Instructional Communication	Make Conversation	CR	1	1	1	1	1
Social, Intercultural, and Instructional Communication	Describe and Request Information	CR	2	2	2	2	2
Social, Intercultural, and Instructional Communication	Tell a Story	CR	1	1	1	1	1
Language Arts/Social Studies/History	Use Academic Words	CR	1	1	1	1	1
Language Arts/Social Studies/History	Present and Explain Information	CR	2 (1 for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Academic Words	CR	1	1	1	1	1
Mathematics/Science/Technical Subjects	Present and Explain Information	CR	2 (1 for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 29. LAS Links Form E and F Operational Blueprint: Reading Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (30 total; 24 for K)	2–3 # of Items (30 total)	4–5 # of Items (30 total)	6–8 # of Items (30 total)	9–12 # of Items (30 total)
Foundational Skills	Read Words	MC	12	6	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Read School Texts	MC	10 (8 for K)	10	14	14	14
Language Arts/Social Studies/History	Read Academic Texts	MC	2	5	6	6	6
Language Arts/Social Studies/History	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2
Mathematics/Science/Technical Subjects	Read Academic Texts	MC	2	5	6	6	6
Mathematics/Science/Technical Subjects	Read School Texts (Gr. 1–12 only)	MC	2 (N/A for K)	2	2	2	2

Table 30. LAS Links Form E and F Operational Blueprint: Writing Skill Area

Language Context Strand	Subskill Area	Item Type	K–1 # of Items (18 total; 14 for K)	2–3 # of Items (17 total)	4–5 # of Items (17 total)	6–8 # of Items (17 total)	9–12 # of Items (17 total)
Foundational Skills	Start Writing	Auto CR CR	2 5	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Use Grammar and Conventions	MC	4	6	6	6	6
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–3)	1	N/A	N/A	N/A	N/A
Social, Intercultural, and Instructional Communication	Write to Express Ideas	CR (0–4)	N/A	1	1	1	1
Language Arts/Social Studies/History	Use Grammar and Conventions	MC	1	2	2	2	2
Language Arts/Social Studies/History	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3
Mathematics/Science/Technical Subjects	Use Grammar and Conventions	MC	1	2	2	2	2
	Write Academic Texts (Gr. 1–12 only)	CR	2 (N/A for K)	3	3	3	3

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Samples

The operational field tests were administered to all students required to take an English language proficiency assessment under Title III in two U.S. states with sufficiently large populations of students classified as English language learners. The Form E operational field test was administered in Mississippi in 2021 and in Connecticut in 2022. The Form F operational field test was administered in Connecticut 2023.

Sample sizes for valid student records collected are provided in Table 22. Note that the Form F Operational Field Test included most of the Form E operational field test items embedded in unscored positions.

Table 31. Operational Field Test, Forms E and F Case Counts by Grade Span

Grade Span	Form E Listening	Form E Speaking	Form E Reading	Form E Writing	Form F Listening	Form F Speaking	Form F Reading	Form F Writing
K—1	12,775	12,750	12,768	12,756	10,229	10,186	10,210	10,241
2—3	12,659	12,644	12,649	12,639	9,726	9,701	9,721	9,730
4—5	10,768	10,760	10,767	10,769	9,180	9,165	9,184	9,193
6—8	11,164	11,140	11,154	11,144	10,002	9,972	10,000	10,025
9—12	10,725	10,669	10,722	10,701	10,581	10,486	10,581	10,659
Total	**58,091**	**57,963**	**58,060**	**58,009**	**49,718**	**49,510**	**49,696**	**49,848**

Administration

Testing Modes and Accessibility

The LAS Links Operational Field Test Forms E and F were offered in computer- and paper-based formats. Large print and Braille versions were also available for Form E of LAS Links 3rd Edition. A full discussion of the accessibility features that were available for the LAS Links Operational Field Tests Form E and F, and that are available in LAS Links 3rd Edition is provided in Chapter 7. Refer to the LAS Links 3rd Edition Test Administration Guide (DRC, 2024) for detailed test administration protocols.

Testing Times

The estimated administration time for each skill area is shown (note that all LAS Links tests are untimed). At K–1, for example, the inclusion of field test items is balanced by the reduction in operational items; therefore, estimated testing times remain consistent with LAS Links 2nd Edition. Listening, Speaking, and Writing all have the same estimated administration times as the previous forms. Only Reading had an increased estimated administration time of five minutes.

Table 23 shows the estimated time to complete the administration of each subtest across grade bands of the Form E and F operational field tests.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 32. LAS Links Form E and F Operational Field Testing Times

Subtest Area	Estimated Administration Time
Speaking	15 minutes - all grades
Listening	35 minutes - grades K–1, 6–8, 9–12 30 minutes - grades 2–3, 4–5
Reading	40 minutes - grades K–1, 2–3, 4–5 50 minutes - grades 6–8, 9–12
Writing	30 minutes - grades K–1 40 minutes - grades 2–12

For scheduling purposes, these time allocations are recommended to complete the administration of each subtest and do not include setup, logins, etc.

Response Scoring

DRC has developed and maintained rigorous scoring processes in concordance with the Standards, which state the following:

> Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented. (p. 119)

The standards further state the following:

> Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected. (p. 119)

Scoring Multiple-Choice Items

The scoring process included the scoring of MC items against the answer key and the aggregation of raw scores from the open-ended responses. A student's raw score is the actual number of points achieved by the student for tested elements of an assessment. From the raw scores, the scale scores were calculated.

The student file was scored against the final and approved MC answer key. Items were scored as right, wrong, omitted, or double-gridded (more than one answer was selected for an item). Sections of the test were evaluated as a whole and an attempt status was determined for each student for each skill area. The score program defined all data elements at the student level for reporting.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Overview of DRC Hand Scoring

All DRC scoring procedures are designed to ensure consistency of scoring in both rater qualification and ongoing operational scoring activities. Specifically, DRC hand scoring uses a process in which 40 to 50 student responses per item are selected during a range-finding activity and scored by experts. These papers and their expert assigned scores are then used to train and qualify raters.

To qualify for participation in operational scoring, raters must reach established agreement criteria. The value of those criteria is dependent on the number of score points an item is worth. All DRC Scoring Directors and raters that participated in the Forms E and F operational field tests have been scoring LAS Links for at least 10 years.

Once raters are qualified, DRC hand scoring experts continue to monitoring rater performance throughout operational scoring through a) read-behinds, whereby an expert rater provides periodic reviews of rater scoring of student responses; b) second reads, whereby a certain percentage of randomly selected student responses are passed to two raters and their scores are compared; and c) additional rater monitoring using sets of the expert scored responses.

DRC performs a 10 percent second read for LAS Links and documents the resulting inter-rater reliability (IRR) in terms of the percentages of exact, adjacent, and discrepant agreement. For LAS Links 2nd Edition, DRC scoring has averaged exact agreement rates of 99% for Reading, 90% for Writing and 86% for Speaking. Inter-rater reliability statistics for all hand scored items on the Forms E and F operational field tests are provided in Appendix E.

Range-finding

After student answer documents were received and processed, DRC's Performance Assessment Services (PAS) staff assembled groups of 40 to 50 responses that exemplified the different score points for each domain and item type.

Response copies were made for each range-finding participant. Range-finding committees consisted of DRC Test Development staff and DRC Performance Assessment Services staff.

Each range-finding activity began in a joint session with a review of the history of the assessment and a discussion of the subject/grade-specific groups. Sets of student responses were presented to the committees one item at a time. Each committee initially reviewed and scored student responses as a group to ensure that everyone was interpreting the scoring guidelines consistently. Committee members then went on to score responses independently. For each student response, committee members' scores were discussed until a consensus was reached. Only those responses for which there was strong agreement among committee members were chosen for inclusion in training materials for DRC raters.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Discussions of student responses included the mandatory use of scoring guideline language. This ensured that committee members remained focused on the specific requirements of each score level. DRC PAS staff took notes addressing how and why the committees arrived at score point decisions, and this information was used by the scoring directors in rater training.

DRC and PDE discussed scoring guideline edits suggested by the range-finding committees. Changes approved by PDE were then incorporated into the scoring guidelines by DRC Test Development staff. The edited scoring guidelines were used in the preparation of materials and the training of raters.

Rater Recruitment/Qualifications

DRC retains a number of LAS Links raters from year to year. This pool of experienced raters was drawn on to staff the operational field test scoring activities. To complete the rater staffing for this project, DRC placed advertisements in local newspapers and utilized a variety of web sites. Open houses were held and applications for rater positions were screened by DRC's recruiting staff. Candidates were personally interviewed by DRC staff. In addition, each candidate was required to provide proof of a four-year college degree. In this screening process, preference was given to candidates with previous experience scoring large-scale assessments and degrees emphasizing expertise in English language development. Thus, the rater pool consisted of educators and other professionals with content-specific backgrounds. These individuals were valued for their content-specific knowledge, but they were required to set aside their own biases about student performance and accept the scoring standards outlined for LAS Links.

Leadership Recruitment/Qualifications

Scoring directors and team leaders were selected by content specialists from a pool of employees who had displayed expertise as raters and leaders on previous DRC projects. These individuals had strong backgrounds in organization, leadership, and management. A majority of scoring directors and team leaders had at least five years of leadership experience working on large-scale assessments, including LAS Links. All scoring directors, team leaders, and raters were required to sign confidentiality agreements before handling secure materials.

Each room of raters was assigned a scoring director. This individual led all hand scoring activities for the duration of the project. Scoring directors assisted in range-finding, worked with supervisors to create training materials, conducted team leader training, and were responsible for training the raters. The scoring directors made sure that reports were available and interpreted those reports for the raters. The scoring directors also supervised the team leaders. All scoring directors were monitored by the project director, the project manager, and the content specialists.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Team leaders assisted the scoring director with rater training by leading their teams in small group discussions and answering individual questions that raters may not have felt comfortable asking in a large group. Once raters were qualified, team leaders were responsible for maintaining the accuracy and workload of each team member. Ongoing monitoring identified those individuals having difficulty scoring accurately. These raters received one-on-one retraining from the team leader. Any rater who could not be successfully retrained had his/her scores purged and was released from the project.

Training

As part of preparation for the scoring activities, scoring guidelines and scored student responses approved by range-finding committees were assembled into sets used for training raters. The item-specific scoring guidelines served as the raters' constant reference. Responses that were relevant in terms of the scoring concepts they illustrated were annotated and included in a representative sample of items, or "anchor set." The full range of each score point was clearly represented and annotated in the anchor set, which was used for reference by raters throughout the project.

Training sets and qualifying sets contained student responses consensus-scored by range-finding committee members. Raters were instructed on how to apply the scoring guidelines and were required to demonstrate a clear comprehension of each anchor set by performing well on the associated training materials. Responses were selected for training to show raters the range of each score point (e.g., high, mid, and low 2s). Examples of 0s were included for all items. This process helped raters recognize the various ways that a student could respond in order to earn each score point outlined and defined in the scoring guidelines.

The scoring director conducted a team leader training session before training the raters. This session followed the same procedures as rater training, but qualifying standards were more stringent due to the extra responsibilities required of team leaders. During team leader training, all materials were reviewed and discussed. Team leaders were required to annotate all their training materials with committee justifications from the range-finding meetings. To facilitate scoring consistency, it was imperative that all team leaders imparted the same rationale for each response. Once the team leaders were qualified, leadership responsibilities were reviewed, and team assignments were given. A ratio of one team leader per 8–10 raters ensured sufficient monitoring rates for team members.

Rater training began with the scoring director providing an intensive review of the scoring guidelines and anchor papers. Next, raters practiced by independently scoring the responses in the training sets. After each training set, the scoring director or team leaders led a thorough discussion of the responses, either in a large-group or small-group setting.

Once the scoring guidelines, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Raters who failed to achieve 70 percent exact agreement on the first qualifying set were given additional, individual training. Raters who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential raters in DRC's imaging system and released from the project.

Hand Scoring Process

Student responses were scored independently. All responses were scored once, and ten percent of the responses were scored a second time. The data collected from the ten-percent double-read portion were used to calculate the exact and adjacent agreement rates provide in Appendix E. The responses that were used for the ten percent read behind were randomly chosen at the item level. Additional read behinds by the team leaders and scoring directors were done to further ensure score reliability. Raters were only provided with student responses that they were qualified to score. Scores were keyed into DRC's imaging system.

To handle possible alerts (i.e., student responses indicating potential issues related to students' safety and well-being that sometimes require attention at the state or local level), DRC's system allows raters to forward responses needing attention to the scoring director. These alerts are reviewed by project management, who then notifies the students' schools of the occurrences. At no time in the alerts process do raters, or other DRC staff, acquire any knowledge concerning a student's personal identity.

Item Calibration and Equating

Following data collection for the Form E operational field test, all items were calibrated using item response theory and linked to LAS Links assessment scales using a randomly equivalent groups design during the Spring and Summer of 2022. Specifically, stratified random samples of 2021 (LAS Links Form D) and 2020 (LAS Links Form C) test data were drawn and used to target scale score means and standard deviations for Form E test data. The transformation constants were obtained using a linear equating procedure (Crocker & Algina, 1986). Details of these procedures are discussed in the following sections.

Data Preparation

The following procedures were used to sample and clean examinee scale scores from the 2020 (Form C) and 2021 (Form D) data.

Observed valid Form E test records distribution (%) by grade within a grade span on each test domain (Listening, Reading, Speaking, and Writing) by the time of the 2022 calibration and equating (March 2, 2022) were used to set the sampling target for the 2020 (Form C) and 2021 (Form D) test data.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

As student performance may vary by grade, stratified sampling by grade ensured that the target sample was consistent with the observed Form E data in grade distribution. In addition, the same N counts of sampling target were used for each year (2020 and 2021) to ensure equal weight of data from the two years. Table 24 shows the resultant sampling target, taking the above factors into consideration.

Using the sampling target, samples were drawn separately from 2020 (Form C) and 2021 (Form D). The samples were then combined to reduce potential impact of COVID-19 (2020 vs. 2021) and form effect (C vs. D). The combined sample was used in the following linear equating procedure (described in the Calibration and Equating Section below).

Table 33. Stratified Random Sample Target Definition

Content	Test Level	Grade	Sampling Target
Listening	K	K	3,712
	1	1	3,649
	2	2	4,041
	2	3	4,064
	4	4	3,709
	4	5	3,156
	6	6	2,478
	6	7	2,185
	6	8	1,898
	9	9	2,085
	9	10	1,563
	9	11	1,333
	9	12	1,041

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Content	Test Level	Grade	Sampling Target
Reading	K	K	3,640
	1	1	3,583
	2	2	3,887
	2	3	3,924
	4	4	3,611
	4	5	3,125
	6	6	2,416
	6	7	2,190
	6	8	1,837
	9	9	2,101
	9	10	1,536
	9	11	1,319
	9	12	1,019
Speaking	K	K	3,949
	1	1	3,725
	2	2	4,016
	2	3	4,043
	4	4	3,706
	4	5	3,083
	6	6	2,581
	6	7	2,381
	6	8	2,035
	9	9	2,037
	9	10	1,564
	9	11	1,330
	9	12	1,049

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Content	Test Level	Grade	Sampling Target
Writing	K	K	3,917
	1	1	3,731
	2	2	4,147
	2	3	4,130
	4	4	3,653
	4	5	3,152
	6	6	2,527
	6	7	2,401
	6	8	2,103
	9	9	2,101
	9	10	1,564
	9	11	1,340
	9	12	1,076

Before the samples were drawn, the following exclusions were applied to the data from Forms C, D, and E data:

- Invalidated records
- Off grade records
- Records with no scale scores for the reported domain
- Duplicate records

Item Response Theory Models

Item response theory (IRT) models were used to calibrate and scale the LAS Links Form E Operational Field Test items. Since both MC and CR items are included on the tests, both were placed on the LAS Links Score Scale using a combination of the 3-parameter logistic (3PL) model (Lord & Novick, 1968; Lord, 1980), and the 2-parameter partial credit (2PPC) models (Muraki, 1992; Yen, 1993).

Under the 3PL model, the probability that a student with the ability θ correctly responds to item *i* is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp\left[-1.7a_i\left(\theta - b_j\right)\right]}$$

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

where a_i is the item discrimination, b_j is the item difficulty, and c_i is the probability of a correct response by a very low-scoring student.

For analysis of CR items, the 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with the ability θ having a score at the k-th level of the j-th item is

$$P\left(x_j = k-1 \mid \theta\right) = \frac{\exp Z_{jk}}{\sum_{i-1}^{mi} \exp Z_{ji}}, k = 1 \ldots m_j$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk},$$

and A_{jk} is the slope of the k-th level and and C_{jk} is its intercept.

For the special case of the 2PPC model used here, the following constraints are used:

$$A_{jk} = \alpha_j(k-1)$$

and

$$C_{jk} = -\sum_{i=0}^{k-1} \Upsilon_{ji}$$

where $\Upsilon_{i0} = 0$ and α_j and Υ_{ji} are the free parameters to be estimated from the data. The first constraint implies that higher item scores reflect higher ability levels and that the items may vary in their discriminations. For the 2PPC model, each item consists of $m_j - 1$ 1 independent Υ_{ji} parameters and one α_j parameter; a total of m_j individual item parameters are estimated for each item.

Equating the Form E Operational Field Test

An equivalent groups equating design was used whereby two randomly equivalent groups of students took the Form E Operational Field Test and Forms C and/or D. Recall from the previous discussion on sampling that equivalent groups were produced by using the sampling target to draw samples separately from 2020 (Form C) and 2021 (Form D). The samples were then combined to reduce the potential impact of COVID-19 (2020 vs. 2021) and form effect (C vs. D). The obtained target sample was used to set the target distribution of scales scores by applying the procedure specified in Crocker and Algina (1986) such that

$$Y^* = a(X - c) + d,$$

(Angoff, 1971), where X denotes a score on Form C or D, and Y^* denotes the transformed scale score on the Form E Operational Field Test that is equivalent in score meaning to X. In this expression,

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

$$a = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$$

$$c = \hat{\mu}_X$$

$$d = \hat{\mu}_Y$$

where c is the scale score mean of the target sample, d is the theta score mean of the Spring 2022 Form E data, and a represents a standard deviation of the 2022 theta score divided by the standard deviation of the target sample. The linear transformation constants were derived from a, c, and d and applied to Form E.

Scale Evaluation

The scaling and equating results for the Form E operational field test were evaluated by checking summary scale score statistics, impact data, and test characteristic curves for reasonableness in terms of their alignment with expectations based on Connecticut data from the preceding three years (2019, 2020, and 2021). Results indicated good alignment with historical patterns.

Item Parameter Updates

The additional Form E operational field test data from Mississippi were used to further update the item parameters and scale the field test items. Since these data were not available on the timeline for Spring 2022 equating, they were used in this subsequent step to update the item parameters after the Form E Operational Field Test equating window.

After applying the same exclusion rules applied to the data in March 2022, the combined data were used to update the scored item parameters and to place all embedded field test items on the LAS Links scale. This was done through the application of the Stocking and Lord (1983) equating procedure, whereby the existing IRT item parameters for Operational Field Test Form E scored items (using the Form E Operational Field Test data from 2021 and 2022) were used as anchors in a common item, non-equivalent groups equating design.

The Stocking and Lord (1983) procedure employs a linear transformation to minimize the average squared difference between anchor item characteristic curves (ICC). Thus, a second concurrent calibration of items by test domain over all test levels using the 2-PL and 2PPC models was conducted, and the linear transformation that minimized this difference was applied. This minimization is defined by *F,* which is a function of transformation constants M_1 and M_2:

$$F = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\gamma}_j - \hat{\gamma}_j^*\right)^2$$

where N is the number of examinees in a group, $\hat{\gamma}_j$ is the estimated true score obtained from the base test form, and $\hat{\gamma}_j^*$ is the estimated true score obtained from the equated

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

test form after it has been transformed to the previous scale. It was transformed to the previous scale using

$$\hat{\gamma}_j = \hat{\gamma}(\theta_j) \sum_{i=1}^{N} P_i(\theta_j; a_i, b_i, c_i)$$

$$\hat{\gamma}_j^* = \hat{\gamma}(\theta_j) \sum_{i=1}^{N} P_i\left(\theta_j; \frac{a_i}{M_1}, M_1 b_i + M_2, c_i\right)$$

where a_i, b_i, c_i are the IRT discrimination, location (difficulty), guessing parameters for item i from equation (1). Equating was performed using IRTEQ (Han, 2009).

Scale Evaluation

The equating results were evaluated by comparing test characteristic curves between the March scaling and the June item parameter updates. Further, correlations between the discrimination, difficulty, and guessing parameters (MC only) for all items on the scored test were evaluated, as well as the root mean square deviation (RMSD) of the F-parameters.

As expected, the item parameter correlations were very high (no item parameter correlations were below 0.91), the equated TCCs showed near perfect alignment, and RMSDs of the F-parameters were very small (none were greater than 0.004).

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 4: Validity Evidence Related to Internal Structure

Chapter 4 provides evidence of test score validity as it relates to the internal structure of the LAS Links assessments, including detailed technical information about test scale development, multiple forms equating to the LAS Links common scale, descriptive score statistics, reliability, standard error of measurement, decision consistency, confirmatory factor analysis, and scale maintenance.

Classical Item Analysis

CIA and the evaluation of DIF of field test items were conducted to assess the quality of the test items and to identify items for data review. Data from the Form E Operational Field Test were collected in the Spring of 2022.

Item Difficulty

At the most general level, an item's difficulty is indicated by its mean score in some specified group (e.g., grade level):

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$$

In the mean score formula above, the individual item scores (x_i) are summed and then divided by the total number of students (n). For MC items, student scores are represented by 0s and 1s (0 = wrong, 1 = right). With 0–1 scoring, the equation above also represents the number of students correctly answering the item divided by the total number of students. Therefore, this is also the proportion correct for the item, or the p-value. In theory, p-values can range from 0.00 to 1.00 on the proportion-correct scale. For example, if an item has a p-value of 0.89, it means 89 percent of the students tested answered the item correctly. Additionally, this value might suggest that the item was relatively easy and/or that the students who attempted the item were relatively high achievers. In other words, item difficulty and student ability are somewhat confounded. Note that for MC items with four response options, pure random guessing would lead to an expected p-value of 0.25.

For CR items, mean scores can range from the minimum possible score (usually zero) to the maximum possible score (e.g., four points in the case of some Speaking and Writing items). Often, a pseudo p-value is provided for a CR item. This is done by dividing the mean item score by the maximum possible item score. The minimum and maximum extremes of the difficulty scale are typically not seen in practice. However, understanding the extremes helps illustrate that relatively lower values correspond to more difficult items and that relatively higher values correspond to easier items.

Item difficulty is an important consideration for the LAS Links assessments because of the range of language proficiency levels of English learners (Beginning, Early Intermediate, Intermediate, Proficient, and Above Proficient). Items that are either very hard or very easy provide little information about student differences in language

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

development. However, an item answered correctly by a high percentage of students would suggest that the knowledge or skill the item taps has been mastered by most students. Conversely, an item answered incorrectly by a high percentage of students would suggest few students have mastered the knowledge or skill the item measures. On a language proficiency assessment such as LAS Links, a test development goal is to include a wide range of item difficulties. Item difficulty information can be found in Appendix D.

Item Discrimination

At the most general level, item discrimination indicates an item's ability to differentiate between high and low achievers. It is expected that students with high language proficiency (i.e., those who perform well on the LAS Links assessment overall) would be more likely to answer any given item correctly, while students with low language proficiency (i.e., those who perform poorly on the LAS Links assessment overall) would be less likely to answer the same item correctly.

Pearson's product-moment correlation coefficient between item scores and test scores is used to indicate discrimination. The item score is removed from the total score such that the resulting correlations will not be spuriously high. The correlation coefficient can range from -1.0 to +1.0 The expectation is that high-ability students tend to answer the item correctly while low-ability students tend to answer the item incorrectly. If this expectation is met, the correlation between the item score and the total test score will be both positive and noticeably large in its magnitude (i.e., well above zero). This would indicate the item discriminates well between high- and low-ability students. The correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly. In other words, the relationship between student test performance and item performance is expected to be consistent.

Reliability

According to the *Standards for Educational and Psychological Testing* (AERA et al, 2014), the general notion of reliability/precision refers to

> the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported. (p. 33)

Frisbie (2005) highlighted several elements of reliability. First, reliability is a property of test scores, not of a test itself. Many may appreciate this distinction, but in casual usage, individuals frequently refer to a reliable test. While reliability concerns test scores (and not the test specifically), it is important to appreciate the fact that test scores can be affected by characteristics of the instrument. For example, all other things being equal, tests with more items/points tend to be more reliable than tests with fewer items/points. Second, reliability coefficients are group specific. Reliabilities tend to be higher in populations that are more heterogeneous in terms of score ranges and lower

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

in populations that have more restricted score ranges. Consequently, both test length and population score ranges should be considered when evaluating reliability.

There is a reliability consideration that may be less evident from the *Standards'* definition yet still important for test users to understand. While freedom from measurement error is very important, reliability is specifically concerned with random sources of error. Indeed, the degree of inconsistency due to random error sources is what determines reliability: less consistency is associated with lower reliability and more consistency is associated with higher reliability. Of course, systematic error sources also exist which can artificially increase reliability and decrease validity.

Another noteworthy issue is that multiple sources of error exist (e.g., the day of testing, the items used, the raters who score the items). However, most widely used reliability indices only reflect a single type of error. Consequently, it is important for test users to understand what specific type of error is being considered in a reliability study and equally (if not more) importantly, what types are not.

Understanding the distinction between relative error and absolute error is also important as many reliability indices only reflect relative error. Relative error is of interest whenever the relative ordering of individuals respective to their test performance is of interest.

Understanding examinee rank-order stability is important; however, such stability might be well achieved even when the specific score values are considerably different. When specific score values are considered important (e.g., if cuts scores are used), then absolute error is too. Generally, there is more error variance when considering the absolute scores of examinees, which in turn suggests lower reliability.

As suggested, reliability is a complex, nonunitary notion that cannot be adequately represented by a single number. Therefore, consistency of scores is examined, as is their standard error. Additionally, the consistency of decisions about the level of proficiency that students have attained based on LAS Links scores is examined.

Reliability Indices

As shown below, the reliability coefficient expresses the consistency of test scores as the ratio of true score variance to total score variance. The total variance contains two components: 1) the variance in true scores and 2) the variance due to the imperfections in the measurement process. Put differently, total variance equals true score variance plus error variance, as shown below.

$$\rho_x^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

Reliability coefficients indicate the degree to which differences in test scores reflect true differences in the attribute being tested rather than random fluctuations. Total test score variance (i.e., individual differences) is partly due to real differences in the attribute (true

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

variance) and partly due to random error in the measurement process (error variance). A covariance term is not required as true scores and error are assumed to be uncorrelated in classical theory.

Reliability coefficients range from 0.0 to 1.0. If all test score variances were true, the index would equal 1.0. The index will be 0.0 if none of the test score variances were true. Such scores would be pure random noise (i.e., all measurement error). If the index achieved a value of 1.0, scores would be perfectly consistent (i.e., contain no measurement error). Although values of 1.0 are never achieved in practice, larger coefficients are more desirable because they indicate that test scores are less influenced by random error.

As noted in the introduction, there are several different indices that can be used to estimate this ratio. One approach is referred to as internal consistency. This is derived from analyzing the performance consistency of individuals over the items within a test. As discussed below, these internal consistency indices do not take into account other sources of error, for example, variations due to random errors associated with the linking process, day-to-day variations (student health, testing environment, etc.), and rater inconsistency.

Coefficient Alpha

Although a number of reliability indices exist, perhaps the most frequently reported for large scale assessments is Coefficient Alpha (Cronbach, 1951). Consequently, this index is reported for LAS Links. Alpha indicates the internal consistency over the responses to a set of items measuring an underlying trait, in this case, English language proficiency in domains: Listening, Reading, Speaking, and Writing.

Alpha is an internal consistency index. It can be conceptualized as the extent to which an exchangeable set of items from the same domain would result in a similar rank ordering of students. Note that relative error is reflected in this index. Variation in student performance from one sample of items to the next should be of particular concern for any achievement test user. Consider two hypothetical vocabulary tests intended for the same group of students. Each test contains different sets of unique words that are believed to be randomly equivalent, perhaps like the ones shown below in Table 25.

Table 34. Two Hypothetical Vocabulary Tests

Test One	Test Two
Abase	Abate
Boon	Bilk
Capricious	Circuitous
Deface	Debase
….	….

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

If a representative group of students could take both tests and the correlation between the scores could be obtained, then that result would represent the parallel forms reliability of the test scores. However, such data-collection designs are impractical in large-scale settings and experimental confounds like fatigue and practice effects are likely to affect the results. Internal-consistency reliability indices arose in part to provide reliability measures using the data from just a single test administration. So if students only took Test One and the Coefficient Alpha index for those test scores was high, then this would suggest that Test Two would provide a very similar rank ordering of the students if they had taken it instead. If Coefficient Alpha were low, dissimilar rank orderings would likely be observed—again, relative-error variance is reflected in Alpha. It should also be noted that Coefficient Alpha is algebraically identical to a Person × Item design under Generalizability Theory when relative error variance is assumed.

Consider the following data matrix displayed in Table 26 below. The rows represent examinees, labelled “Person” (or “P”) 1 through N, and columns represent test items, labeled “Item” (or “I”) 1 through k. The score Y for each examinee on each item is represented by the cells in the matrix.

Table 35. Person × Item Score ($\mathbf{X}_{pi}$) Infinite (Population-Universe) Matrix

Person	Item 1	Item 2	Item…I	Item…K
1	Y_{11}	Y_{12}	$\ldots Y_{1i}$	$\ldots X_{1k}$
2	Y_{21}	Y_{22}	$\ldots Y_{2i}$	$\ldots X_{2k}$
… …	…	…	…	…
P	Y_{p1}	Y_{p2}	$\ldots Y_{pi}$	$\ldots X_{pk}$
… …	…	…	…	…
N	Y_{N1}	Y_{N2}	$\ldots Y_{ni}$	$\ldots X_{nk}$

Note. Adapted from Cronbach and Shavelson (2004).

Then, a general computational formula for Alpha is as follows:

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma_{Yi}^2}{\sigma_X^2}\right)$$

where N is the number of parts (items or testlets), σ_X^2 is the variance of the observed total test scores, and σ_{Yi}^2 is the variance of part i.

Interpretation

What reliability value is considered high enough? What values are considered too low? Although frequently asked for, any rules of thumb for interpreting the magnitude of reliability indices are mostly arbitrary. Another approach is to research the reliabilities from similar testing instruments to see what values are commonly observed. For LAS

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Links, comparisons to tests of similar lengths that were administered to similar student populations from other large-scale assessment programs would be relevant. For many state academic achievement assessment programs, reliabilities in the low 0.90s are usually the highest ever observed and reliabilities in the high 0.80s are very common. Shorter domain tests in English language proficiency tests tend to have lower reliabilities than these longer achievement tests. However, there is no firm guideline regarding how low is too low. As an informative point of reference, a reliability coefficient of 0.50 would suggest that there is as much error variance as true-score variance in the scores.

Is Alpha a Lower Limit to Reliability?

According to Brennan (1998), "the conventional wisdom that Coefficient Alpha is a lower limit to reliability is based largely on a misunderstanding." In reflecting on the 50th anniversary of his seminal 1951 article, Cronbach—in Cronbach and Shavelson (2004)—expressed similar misgivings about this conventional wisdom:

> One could argue that alpha was almost an unbiased estimate of the desired reliability…the almost in the preceding sentence refers to a small mathematical detail that causes the alpha coefficient to run a trifle lower than the desired value. This detail is of no consequence and does not support the statement made frequently in textbooks or in articles that alpha is a lower value to the reliability coefficient. That statement is justified by reasoning that starts with the definition of the desired coefficient as the expected consistency among measurements that had a higher degree of parallelism than the random parallel concept implied.

The assumptions for three common parallelism models are presented in Table 27. Alpha's assumptions come from the Essentially Tau Equivalent model, which does not require equal means or equal variances across test parts. Based on this, Brennan (1998) asserts that the lower-limit issue, as conceptualized by many, provides an answer to a question that is of minimal importance. Reframed differently, the goal of selecting a reliability coefficient is not to find the one that provides the highest coefficient but the one that most accurately reflects the test data under study.

It is important to note that there are factors encountered in practice that may legitimately make Coefficient Alpha an underestimate of reliability. However, there are also factors that might make Coefficient Alpha an overestimate of reliability. Both possibilities are discussed further below and generally arise when the Essentially-Tau Equivalent assumptions are strained.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 36. Summary of Expectations/Observable Relationships for Different Parallelism Models

Relationship	Degree of Measurement Parallelism*: Classically Parallel	Degree of Measurement Parallelism*: Essentially-Tau Equivalent	Degree of Measurement Parallelism*: Congeneric
Content Similarity	Yes	Yes	Yes
Equal Means across Parts	Yes	No	No
Equal Variances across Parts	Yes	No	No
Equal Covariances across Parts	Yes	Yes	No
Equal Covariances with Other Variables	Yes	Yes	No

Note. Other models exist but are not considered here due to their limited application in practice.

Biases That Might Make Alpha an Underestimate of Reliability

There are factors that might negatively bias Coefficient Alpha, making the apparent reliability lower than it may actually be. Two situations frequently encountered in practice that might cause this include tests that are composed of mixed item types (e.g., MC and OE items) and tests that include a planned stratification of the test items according to topics or subdomains.

Although both situations strictly violate the assumptions on which Coefficient Alpha is derived (i.e., the tests are not based on equal part lengths in the former case and are not randomly parallel in the latter case), neither necessarily guarantees that the reliability will be markedly lower. In the latter case, reliability will be underestimated only when strand items are homogeneous enough for the average covariance within strata to exceed the average covariance between strata.

Biases That Might Make Alpha an Overestimate of Reliability

As emphasized in earlier sections, Coefficient Alpha only considers measurement error that arises from the selection of items used on a particular test form. There are other sources of random inaccuracy. One is the occasion of testing. Various random conditions that might affect students on any particular testing occasion include illness, fatigue, and anxiety.

Also, when a test includes OE items, as LAS Links does, another source that can cause random fluctuation is the OE item scorers. In a sense, Alpha may be positively biased because it does not consider these other important sources of random error. Any internal consistency reliability index could understate the overall problem of measurement error because it ignores such sources of random error.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Another positive bias can occur when items are associated (clustered) with a common stimulus. Item bundles and testlets are other frequently used terms for this situation. One concrete example is when multiple reading comprehension items are associated with a common passage selection. Again, such a situation does not guarantee that the reliability estimate will be markedly affected, but the potential exists.

Standard Error of Measurement

The reliability coefficient is a unit-free indicator that reflects the degree to which scores are free of measurement error. It always ranges between 0.0 and 1.0 regardless of the test's scale. Reliability coefficients best reflect the extent to which measurement inconsistencies may be present or absent in a group. However, they are not that useful for helping users interpret test scores. The standard error of measurement (SEM) is another indicator of degree of consistency for the scores obtained by individual examinees. A relatively large SEM indicates relatively low reliability. The conditional SEMs (CSEM) discussed further below are SEM at that score level.

Traditional Standard Error of Measurement

A precise, theoretical interpretation of the SEM is somewhat unwieldy. A beginning point for understanding SEM is to first understand the concept of a true score. A true score is the score a person would receive if the measurement process were perfect. However, as educational and psychological assessments measure latent traits (traits that are not directly observable), some error is expected due to random factors such as differences in attention during instruction, concentration during testing, and the sampling of test items to name a few. Such random factors create a circumstance where, even if everyone being tested had the same true score, there would still be some variation in observed scores due to imperfections in the measurement process.

The standard error, then, is defined as the standard deviation of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents very important information for test score users.

The SEM formula is provided below.

$$SEM = SD\sqrt{1 - \text{ reliability}}$$

This formula indicates that the value of the SEM depends on both the reliability coefficient and the standard deviation of test scores. For normal distributions, the standard deviation is a measure of the dispersion of the observations about the mean, where approximately 16 percent of the observations are more than one standard deviation above the mean. If the reliability were equal to 0.00 (the lowest possible value) the SEM would be equal to the standard deviation of the test scores. If test reliability were equal to 1.00 (the highest possible value) the SEM would be 0.0. In other words, a perfectly reliable test has no measurement error (Harvill, 1991). Additionally, the value of the SEM takes the group variation (i.e., score standard deviation) into account.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Consider that an SEM of 3 on a 10-point test would be very different than an SEM of 3 on a 100-point test.

Traditional Standard Error of Measurement Confidence Intervals

The SEM is an index of the random variability in test scores in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the precision of individual test scores. SEMs help place “reasonable limits” (Gulliksen, 1950) around observed scores through the construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, X, and adding and subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between +/-1 SEM about two-thirds of the time. For +/-2 SEM confidence intervals, this increases to about 95 percent.

Further Interpretations

One Standard Error of Measurement for All Test Scores

The SEM approach described above only provides a single numerical estimate for constructing the confidence intervals for examinees regardless of their score level. In reality however, such confidence intervals vary according to a student’s score. Consequently, care should be taken using the SEM for students with extreme scores.

Group Specific

As noted in the introduction, reliabilities are group specific. The same is true for SEMs because both score reliabilities and score standard deviations vary across groups.

Raw-Score Metric

The SEM approach is calculated using raw scores, and as such, the resulting confidence interval bands are on the raw score metric. Error bands on the scaled score metric are considered in the next section.

Type of Error Reflected

The interpretation of the SEM should be driven by the type of score reliability that underpins it. The LAS Links SEMs involve the same source of error relevant to internal consistency indices. In other words, if a student were tested an infinite number of times, the +/-1 SEM confidence intervals constructed for each score would capture the student’s true score about 68 percent of the time.

One simpler description is that a confidence interval represents the possible score range one would observe if a student could be tested twice with the same instrument. Taking the same test on a different day implies the only source of random error being considered is related to the occasion of testing, such as a student might be sleepier one day than another or may be sick or did not get a good breakfast. There is a reliability index that captures this source of random error, and it is referred to as the test-retest

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

reliability coefficient. This is not the type of reliability computed for LAS Links. When internal consistency reliability estimates are used, such an explanation blurs the fact that random error based on the occasion of testing is not considered.

When SEMs are derived from internal consistency reliability estimates, a better approach is to describe the confidence interval as providing reasonable bounds for the range of scores that a student might receive if he or she took an equivalent version of the test; that is, if the student took a test that covered exactly the same content but included a different set of items. If an infinite number of tests with equivalent content were taken, the student's true score would lie within the constructed confidence intervals 68 percent of the time. As an example, if the LAS Links score was 500 and the SEM band was 475 to 525, then a student would be likely to receive a score somewhere between 475 and 525 if a different version of the test had been taken.

Results

Table 28 provides raw score summaries, reliabilities, average discrimination and difficulty, and SEM estimates for each domain grade span tested. Results show that reliabilities for students who use accommodations tend to be lower than for students who do not, however, reliabilities for male and female students and Hispanic and non-Hispanic are very similar with patterns of difference that appear to be random.

It is also noted that the reliabilities for the Listening domain are lower than those for Reading, Speaking, and Writing. Although increasing the test length would be expected to increase the Listening reliabilities, the impact to overall testing time across the four domains is considered impractical at this time.

Table 37. Raw Score Summaries, Reliabilities, Average Item Discrimination and Difficulty, and SEM

Content	Grade	N	Total Score Points	Mean	SD	Average Discrimination	Average Difficulty	Alpha	SEM
Listening	K	6,532	20	12.12	3.97	0.38	0.75	0.75	1.99
	1	6,238	20	14.95	3.83	0.38	0.75	0.81	1.67
	2-3	12,616	20	13.12	3.91	0.35	0.66	0.79	1.79
	4-5	10,737	20	12.46	4.18	0.34	0.62	0.78	1.96
	6-8	11,160	23	14.08	4.62	0.35	0.61	0.8	2.07
	9-12	10,714	23	12.17	4.35	0.29	0.53	0.74	2.22

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Content	Grade	N	Total Score Points	Mean	SD	Average Discrimination	Average Difficulty	Alpha	SEM
Reading	K	6,527	26	12.57	4.70	0.40	0.71	0.76	2.3
	1	6,236	30	20.12	6.43	0.40	0.68	0.88	2.23
	2-3	12,606	30	16.90	6.83	0.40	0.57	0.88	2.37
	4-5	10,737	30	14.85	6.39	0.37	0.50	0.85	2.47
	6-8	11,148	30	13.93	6.14	0.35	0.47	0.84	2.46
	9-12	10,712	30	15.53	6.83	0.40	0.53	0.88	2.37
Speaking	K	6,516	19	9.21	4.97	0.64	0.69	0.84	1.99
	1	6,229	25	16.07	6.29	0.65	0.68	0.88	2.18
	2-3	12,601	31	21.06	6.66	0.65	0.67	0.91	2
	4-5	10,727	31	17.36	6.42	0.61	0.58	0.89	2.13
	6-8	11,134	31	16.46	7.29	0.66	0.57	0.91	2.19
	9-12	10,657	31	14.34	7.71	0.67	0.49	0.91	2.31
Writing	K	6,519	18	8.07	3.68	0.44	0.71	0.72	1.95
	1	6,232	30	15.86	7.66	0.52	0.62	0.88	2.65
	2-3	12,596	32	16.66	8.43	0.52	0.54	0.89	2.79
	4-5	10,739	32	16.18	7.63	0.57	0.58	0.9	2.41
	6-8	11,138	32	13.67	6.69	0.50	0.49	0.88	2.32
	9-12	10,689	32	12.64	7.11	0.50	0.47	0.88	2.46

Decision Consistency and Accuracy

Another way to evaluate the score consistency and accuracy is in terms of the decisions made about examining the proficiency levels. Classification consistency refers to the degree to which the achievement level for each student can be replicated upon retesting using an equivalent form (Huynh, 1976).

Decision consistency answers this question: what is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test? If two parallel forms of the test were given to the same students, the consistency of the measure would be reflected by the extent to which the classification decisions made from the first set of test scores matched the decisions based on the second set of test scores.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Classification accuracy refers to the agreement of the observed classifications of students with the classifications made on the basis of their true scores. An observed score contains measurement error while a true score is free of measurement error. A student's observed score can be formulated by the sum of his or her true score plus measurement error. Decision accuracy is an index to determine the extent to which measurement error causes a classification different than expected from the true score.

Since true scores are unobserved and since it is not feasible to repeat LAS Links testing to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model needs to be imposed on the data to estimate the true scores and to project the consistency and accuracy of classifications solely using data from the available administration (Hambleton & Novick, 1973).

Although a number of procedures are available, one well-known method was developed by Livingston and Lewis (1995) utilizing a specific True Score Model. In this case, classifications across the five LAS Links proficiency levels were examined using the program BB-Class (Brennan, 2004) which applies the Livingston and Lewis (1995) method.

Several factors might affect decision consistency and accuracy. One important factor is the reliability of the scores. More reliable test scores tend to result in more similar reclassifications and less measurement error. Another factor is the location of the cut score in the score distribution. More consistent and accurate classifications are observed when the cut scores are located away from the mass of the score distribution.

For example, when scores are close to being normally distributed, the mass is concentrated in the middle of the distribution, and thus, classifications tend to become more consistent when cut scores go up from 70 percent to 80 percent to 90 percent or, alternatively, go down from 30 percent to 20 percent to 10 percent. The number of performance levels is also a consideration.

Consistency and accuracy indices for four performance levels should be lower than those based on two categories. This is not surprising since classification and accuracy using four levels would allow more opportunity to change achievement levels. Hence, there would be more classification errors and less accuracy with four achievement levels, resulting in lower consistency indices.

LAS Links classification consistency and accuracy reflect the expected pattern, where consistency and accuracy are much higher when based on the Proficient cut score only versus all four cut scores. Refer to Tables 29—36 for details.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 38. Listening Decision Consistency and Accuracy Based on the Proficient Cut Score

Grade	Consistency	Accuracy
K	0.90	0.94
1	0.84	0.89
2	0.87	0.91
3	0.85	0.90
4	0.83	0.89
5	0.82	0.87
6	0.84	0.89
7	0.83	0.88
8	0.83	0.88
9	0.84	0.89
10	0.84	0.89
11	0.81	0.88
12	0.80	0.86

Table 39. Reading Decision Consistency and Accuracy Based on the Proficient Cut Score

Grade	Consistency	Accuracy
K	0.93	0.96
1	0.87	0.91
2	0.91	0.94
3	0.88	0.92
4	0.87	0.91
5	0.85	0.90
6	0.90	0.93
7	0.89	0.92
8	0.87	0.91
9	0.96	0.97

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Consistency	Accuracy
10	0.94	0.96
11	0.92	0.95
12	0.92	0.95

Table 40. Speaking Decision Consistency and Accuracy Based on the Proficient Cut Score

Grade	Consistency	Accuracy
K	0.92	0.96
1	0.89	0.93
2	0.91	0.94
3	0.89	0.92
4	0.87	0.91
5	0.87	0.90
6	0.87	0.91
7	0.87	0.91
8	0.87	0.91
9	0.91	0.94
10	0.90	0.93
11	0.90	0.93
12	0.90	0.93

Table 41. Writing Decision Consistency and Accuracy Based on the Proficient Cut Score

Grade	Consistency	Accuracy
K	0.99	0.99
1	0.97	0.98
2	0.92	0.95
3	0.93	0.96
4	0.89	0.92

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Consistency	Accuracy
5	0.88	0.92
6	0.87	0.91
7	0.86	0.90
8	0.85	0.89
9	0.93	0.95
10	0.91	0.94
11	0.89	0.93
12	0.89	0.93

Table 42. Listening Decision Consistency and Accuracy Based on 4 Cut Scores

Grade	Consistency	Accuracy
K	0.53	0.62
1	0.51	0.60
2	0.49	0.59
3	0.50	0.61
4	0.48	0.59
5	0.46	0.57
6	0.47	0.59
7	0.47	0.58
8	0.48	0.58
9	0.46	0.56
10	0.43	0.54
11	0.52	0.62
12	0.51	0.60

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 43. Reading Decision Consistency and Accuracy Based on Four Cut Scores

Grade	Consistency	Accuracy
K	0.54	0.62
1	0.52	0.64
2	0.55	0.65
3	0.51	0.62
4	0.51	0.62
5	0.51	0.62
6	0.53	0.63
7	0.52	0.62
8	0.50	0.60
9	0.66	0.74
10	0.61	0.71
11	0.60	0.70
12	0.60	0.70

Table 44. Speaking Decision Consistency and Accuracy based on Four Cut Scores

Grade	Consistency	Accuracy
K	0.52	0.65
1	0.62	0.73
2	0.73	0.82
3	0.73	0.81
4	0.68	0.77
5	0.67	0.77
6	0.69	0.78
7	0.66	0.76
8	0.67	0.77
9	0.43	0.65

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Consistency	Accuracy
10	0.74	0.86
11	0.44	0.67
12	0.73	0.84

Table 45. Writing Decision Consistency and Accuracy Based on Four Cut Scores

Grade	Consistency	Accuracy
K	0.73	0.80
1	0.71	0.80
2	0.54	0.64
3	0.58	0.70
4	0.58	0.69
5	0.57	0.68
6	0.53	0.65
7	0.52	0.64
8	0.51	0.64
9	0.60	0.70
10	0.58	0.69
11	0.56	0.68
12	0.57	0.69

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was conducted to assess the degree to which the intended construct for each test explains performance on the operational test items within each of the four domains: Listening, Reading, Speaking, and Writing. Although causal structures can never be proven, it is useful to assess the "…degree to which the plausibility of factor models is empirically confirmed" (Kim & Mueller, 1978, p 46). Further, there are several conditions that can affect the empirical confirmation of a theorized factor structure, including the number of variables and common factors. In the case of LAS Links, we might therefore expect less certainty in the level of empirical confirmation noted for the shorter domain tests of Speaking and Writing.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

A single factor model was specified for each domain and level of the LAS Links test as

$$y_i = \tau + \Lambda\eta_i + \varepsilon_i$$

where y_i is the outcome vector, τ is the intercept vector, Λ is the factor loading matrix, η_i is the common factor score, and ε_i represents the unique factor scores. For each model, the factor variance was fixed to 1.0 for model identification purposes. As the indicators in these models are ordered categorical variables and likely violate the assumption of multivariate normality required for maximum likelihood estimation, the models were fit using robust weighted least squares estimation.

Results were evaluated for the significance of factor loadings, the consistency of standardized variances of the unique factor scores, and the overall fit of the single factor models to the data. Model fit was evaluated for each model using adjusted Chi-Square tests of fit (Satorra & Bentler, 1994; Asparouhov & Muthén, 2010), Root Mean Square Error of Approximation (RMSEA), and the comparative fit index (CFI). RMSEA values below 0.06 and CFI values of 0.90 and above are generally considered to represent good fit (Hu & Bentler, 1999).

Results for the four single factor models in each grade span show that factor loadings for all items are statistically significant. This indicates that the construct is explaining individual item performance reasonably well.

The Chi-square test of fit results for the CFAs show that, for all tests, the model does not fit perfectly in the population with p-values < 0.000. MacCallum (2001) notes that this is often the finding with larger sample sizes. The RMSEA results shows good fit for most tests, with values ranging between 0.019 and 0.116. The CFI results also show good fit for most tests, with values ranging from 0.794 to 0.997. All tests show good fit for at least one of the fit measures applied. Overall, the factor analysis results suggest that a single factor (construct) is explaining the variance in the LAS Links test well in each of the domains, supporting the four-domain reporting structure for LAS Links. Tables 37-40 provide fit details.

Table 46. Model Fit for Listening

Grade	No. Indicators (Items)	RMSEA	CFI	χ^2 df	P-Value Adj χ^2	N
K	20	0.030	0.960	190	< 0.000	4748
1	20	0.026	0.981	190	< 0.000	4486
2-3	20	0.030	0.966	190	< 0.000	9288
4-5	20	0.023	0.975	190	< 0.000	8344
6-8	23	0.023	0.977	253	< 0.000	8636
9-12	23	0.022	0.966	253	< 0.000	8607

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 47. Model Fit for Reading

Grade	No. Indicators (Items)	RMSEA	CFI	χ^2 df	P-Value Adj χ^2	N
K	26	0.047	0.794	325	< 0.000	4721
1	30	0.042	0.944	435	< 0.000	4467
2-3	30	0.026	0.971	435	< 0.000	9230
4-5	30	0.019	0.983	435	< 0.000	8275
6-8	30	0.024	0.972	435	< 0.000	8603
9-12	30	0.022	0.985	435	< 0.000	8643

Table 48. Model Fit for Speaking

Grade	No. Indicators (Items)	RMSEA	CFI	χ^2 df	P-Value Adj χ^2	N
K	8	0.038	0.996	28	< 0.000	4646
1	10	0.030	0.997	45	< 0.000	4386
2-3	12	0.049	0.991	66	< 0.000	8559
4-5	12	0.116	0.962	66	< 0.000	7880
6-8	12	0.110	0.964	66	< 0.000	7687
9-12	12	0.086	0.982	66	< 0.000	7495

Table 49. Model Fit for Writing

Grade	No. Indicators (Items)	RMSEA	CFI	χ^2 df	P-Value Adj χ^2	N
K	14	0.082	0.926	91	< 0.000	4574
1	18	0.077	0.971	153	< 0.000	4358
2-3	17	0.053	0.989	136	< 0.000	9058
4-5	17	0.048	0.989	136	< 0.000	8058
6-8	17	0.041	0.989	136	< 0.000	8311
9-12	17	0.041	0.989	136	< 0.000	8276

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 5: Scores, Interpretations, and Validating Standards

Chapter 5 covers the types of scores and reports offered for LAS Links, as well as the proficiency levels and underlying standard-setting and validation processes that support claims of their validity.

Scale Scores, Proficiency Levels, and Their Interpretations

As discussed in Chapters 1 and 2, LAS Links is a comprehensive assessment system designed to measure English-language proficiency in students from kindergarten through grade 12. The test comprises four domains (Reading, Writing, Listening, and Speaking). Students earn test scores on each of these four domains, and composite scores are calculated as the average of their respective domain combinations as follows:

- Overall (the average of the test scores earned by the student in all four domains),
- Oral (the average of the test scores earned by the student in Listening and Speaking),
- Comprehension (the average of the test scores earned by the student in Reading and Listening),
- Literacy (the average of the test scores earned by the student in Reading and Writing),
- Productive (the average of the test scores earned by the student in Writing and Speaking),

Additionally, LAS Links percentile ranks and normal curve equivalents are provided to help users understand their students' performance relative to national student performance.

Scale score descriptive statistics can be referenced in Appendix A. Detailed performance level descriptions can be found in Appendix B. Complete raw score to scale score conversions, conditional standard errors of measurement (CSEM), percentile ranks, and normal curve equivalents are provided in Appendix C.

Standards Validation

On January 17–27, 2023, DRC sponsored a standards validation workshop for LAS Links. The purpose of the workshop was to evaluate the current cut scores (passing scores) for LAS Links and to determine whether they were valid for continued use in light of recent updates to the test design for Forms E and F. A committee of 42 language educators were engaged in the Bookmark Standard Setting Procedure (Lewis, Mitzel, & Green, 1996) to evaluate the LAS Links cut scores.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

During the online, six-day workshop, educators (a) discussed the expectations for students in each LAS Links proficiency level and (b) evaluated cut scores for each of the four domains: Reading, Writing, Listening, and Speaking. Participants worked in five groups of 4–9 participants each, and each group focused on a different grade span: Kindergarten and grades 2, 4, 7, and 11.

Standards Validation Purpose

When test designs are modified, The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education's Standards for Educational and Psychological Testing (2014) require that proficiency level cut scores (also commonly referred to as "standards") be evaluated for the purpose of determining whether or not they are still valid for continued use in defining the proficiency levels that are used to make decisions about students. This is commonly done through an educator workshop facilitated by experts in standard setting processes. In this case, DRC psychometricians applied the Bookmark Procedure (Lewis et al., 1996).

Standards Validation Results Summary

During DRC's application of the Bookmark Procedure (Lewis et al., 1996), educators made strong recommendations to adjust the Proficient and Above Proficient cut scores for kindergarten and grade 1 Writing. The cut scores for the Overall, Literacy, and Productive composites were also adjusted due to their inclusion of student Writing scores.

Table 41 summarizes the adjustments that will be made to LAS Links cuts scores for Forms E and F beginning in the Spring of 2023:

Table 50. Cut Score Adjustments for LAS Links Forms E/F, Kindergarten and Grade 1

Grade	Cut Score	Writing	Overall	Literacy	Productive
Kindergarten	Early Intermediate	347	389	351	388
	Intermediate	417	425	399	439
	Proficient	*437	*456	*429	*466
	Above Proficient	*452	*499	*463	*501
Grade 1	Early Intermediate	355	394	357	393
	Intermediate	435	433	410	448
	Proficient	*470	*466	*446	*483
	Above Proficient	*498	*512	*488	*524

Note. Adjusted Cut Scores are indicated with Asterisks.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Background

This section provides details regarding the uses and format of LAS Links, describes changes which have been made to LAS Links since the publication of Forms C and D, and presents the history of the LAS Links cut scores.

Changes to the Test Design and Blueprint of LAS Links

Although the updated tests are aligned to the same English language proficiency (ELP) standards as Forms C and D, the following changes were implemented in the development of LAS Links Forms E and F:

- The most current guidelines regarding bias, fairness, sensitivity, and accessibility for English learners have been applied.
- Content and graphics have been adjusted to better reflect current English learner experiences and current learning environments.
- The kindergarten Reading test was reduced in length by 2 items.
- The format for items in the Writing subtest was updated to allow both the stimulus paragraph and the item to appear on the same screen in the online environment.

History of LAS Links Cut Scores

Defensible, skills-based cut scores have been a component of LAS Links since its inception. The standards validation re-evaluated the cut scores that were originally set in 2005 on Forms A and B in light of the changes to the test design and blueprint described above.

2005: Original LAS Links Standard Setting

In June 2005, CTB/McGraw-Hill sponsored a standard setting study for the original two forms of LAS Links, Forms A and B. The standard setting had two primary components: the bookmark standard setting and cut score creation.

On June 27–29, 2005, 109 educators from across the country convened in Scottsdale, Arizona, to recommend cut scores for LAS Links. Participants were divided into 10 groups: two groups each were assigned to kindergarten and grades 2, 4, 7, and 11. Of the two groups assigned to each grade, one group focused on Speaking and Listening, and the other group focused on Reading and Writing. Each group comprised 10–12 participants. For each of these tests, participants considered cut scores associated with the Early Intermediate and Proficient proficiency levels. To make their recommendations, participants engaged in the Bookmark Standard Setting Procedure.

After the standard setting, CTB interpolated (and extrapolated) cut scores for the remaining cut scores and grades of LAS Links. Specifically, CTB used participants' recommendations to calculate cut scores for Intermediate and Above Proficient and cut scores for grades 1, 3, 5, 6, 8, 9, 10, and 12. In addition, cut scores were calculated for

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

three composite scores: Oral, Comprehension, and Overall. Researchers and language acquisition experts at CTB then reviewed the cut scores and implemented them on the new product.

2013: Standards Validation for LAS Links Forms C/D

In 2013, two new forms were created—Forms C and D—the cut scores established in 2005 were applied to the new forms. The existing cut scores were applied to the new forms because (a) the new forms were designed to be substantively parallel to Forms A and B, (b) states and school systems would be encouraged to use any of the four forms of LAS Links interchangeably, and (c) the same test scales would be used to report scores from the four forms.

At the standards validation, item maps were created using items from Forms C and D to support an update to the proficiency level descriptors (PLDs) for the product. Educators then convened to review the item maps and validate the cut scores.

2023: Standards Validation for LAS Links Forms E/F

At the present standards validation, DRC sought to (a) evaluate the existing LAS Links cut scores using the test items developed for Forms E/F, (b) make sure the cut scores linked students' scores on LAS Links to these expectations, and (c) were appropriate for each domain and grade. Four cut scores were evaluated that defined five proficiency levels: Beginning, Early Intermediate, Intermediate, Proficient, and Above Proficient.

Standards Validation Methodology

The Bookmark Procedure (Lewis, Mitzel, & Green, 1996) is one of the most frequently implemented methods to establish proficiency standards on educational assessments. The Bookmark Procedure was previously used on LAS Links at the original 2005 standard setting and at the 2013 standards validation.

A modification of the Bookmark Standard Setting Procedure (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) was implemented to validate the cut scores for LAS Links on January 17–27, 2023. This method has a history of use on large-scale assessments across the U.S. and abroad. The process was comprised of three phases:

- A committee of 42 educators from all over the U.S. convened in a standards validation workshop, where they engaged in a modification of the Bookmark Procedure to review the existing cut scores. These educators reviewed the revised tests, considered the existing cut scores as applied to the revised tests, and engaged in content-based conversations about the extent to which the cut scores remained valid for continued use.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- A committee of five educators and administrators convened in a standards validation review workshop to consider the recommendations made during the standards validation workshop. This committee considered the percentage of students classified in each performance level on the tests (the impact data) as well as the consistency of the existing cut scores and the recommendations made at the standards validation workshop.
- DRC content and standards validation experts reviewed committee recommendations from the preceding phases and made final cut score determinations.

Workshop Materials

The materials used at the standard setting workshop were based on empirical test data from LAS Links collected in 2021 and 2022.

LAS Links Proficiency Level Descriptors

Proficiency level descriptors (PLDs) are a key input into the standards validation process. PLDs summarize the knowledge and skills expected of students in each proficiency level. Egan, Schneider, and Ferrara (2012) suggest a framework of four types of PLDs, described here.

- Policy PLDs summarize the general definition for each proficiency level, providing information to stakeholders on the suggested interpretation of each level. They are typically not specific to any given grade or domain.
- Range PLDs summarize the knowledge and skills expected of students in a given proficiency level on a specific test (e.g., grade 2 Reading, grade 8 Speaking). The range PLDs show the types of English-language skills, as informed by the LAS Links standards, that should be mastered by students in each proficiency level on the test at hand.
- Threshold PLDs are based on the range PLDs and summarize the English-language skills expected of students who are at the point-of-entry (the threshold) of each proficiency level. For any given test, these descriptors show the types of skills needed to be just classified in a given proficiency level (e.g., to be just to be classified as Proficient on grade 4 Listening).
- Reporting PLDs are the version of the PLDs used for score reporting. Typically, versions of the policy or range PLDs are used, and the language in the reporting PLDs is adjusted to be accessible to a wide audience that may not have in-depth specialist knowledge. (Reporting PLDs were not part of the scope of this workshop.)

The policy and range PLDs were developed in 2005 and updated in 2013. Staff members from Test Development reviewed the PLDs and considered whether any

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

adjustments should be made prior to the workshop. Some adjustments included removing references to English language skills no longer measured by LAS Links and amplifying references to skills now more prominently measured by LAS Links. The updated PLDs were presented to the standards validation participants during the workshop. At the workshop, participants used these PLDs to develop threshold descriptors. Refer to Appendix B for a detailed presentation of the PLDs.

Ordered Item Booklets

Ordered item booklets (OIBs) lay at the heart of the Bookmark Procedure. An OIB of 38 to 87 items/score points and an accompanying item map were created for each of the 20 grade-span and domain combinations for LAS Links. Within each of the 20 OIBs, test items are ordered by ascending difficulty. Item difficulty, as defined by its scale location adjusted by a response probability (RP) value of RP67GA, is based on data from the sample of examinees used to create the LAS Links test scales.

Participants used the OIB to evaluate cut scores. Accordingly, it was important that the items included in the OIB spanned the difficulty continuum—from easy to hard—and that items were found around the points on the test scale where cut scores currently appeared. The items in the OIB reflected the test blueprint, mirroring the range of skills measured by the test.

To create the OIBs, DRC used the following plan:

- Start with the operational/field test items from Forms E and F. The items selected for the operational/field test Forms E and F comprised the core of each OIB.
- Augment with additional items as needed. Additional items were used to augment the OIBs in cases where (a) there were significant gaps between RP-adjusted scale locations in the OIB, (b) there were few items around the existing cut scores, and (c) there were few items that separate the existing cut scores from each other.
- Check the test blueprints for similarity. The items in each OIB should mirror the test blueprints in a meaningful way. Specifically, the proportion of items in each OIB should mirror the test blueprints at the Language Context Strand level (e.g., “Reading: Foundational Skills,” “Speaking: Language Arts/Social Studies/History”). The OIB proportions should match those from the blueprint within 5%.

Most items on LAS Links are worth one point. As participants studied these items, they considered the English language skills that students needed to answer the item correctly.

The test also contains CR items worth multiple points. As participants studied these items, they used the items and associated scoring guides (rubrics) to consider the English language skills that students needed in order to earn the first score point, and

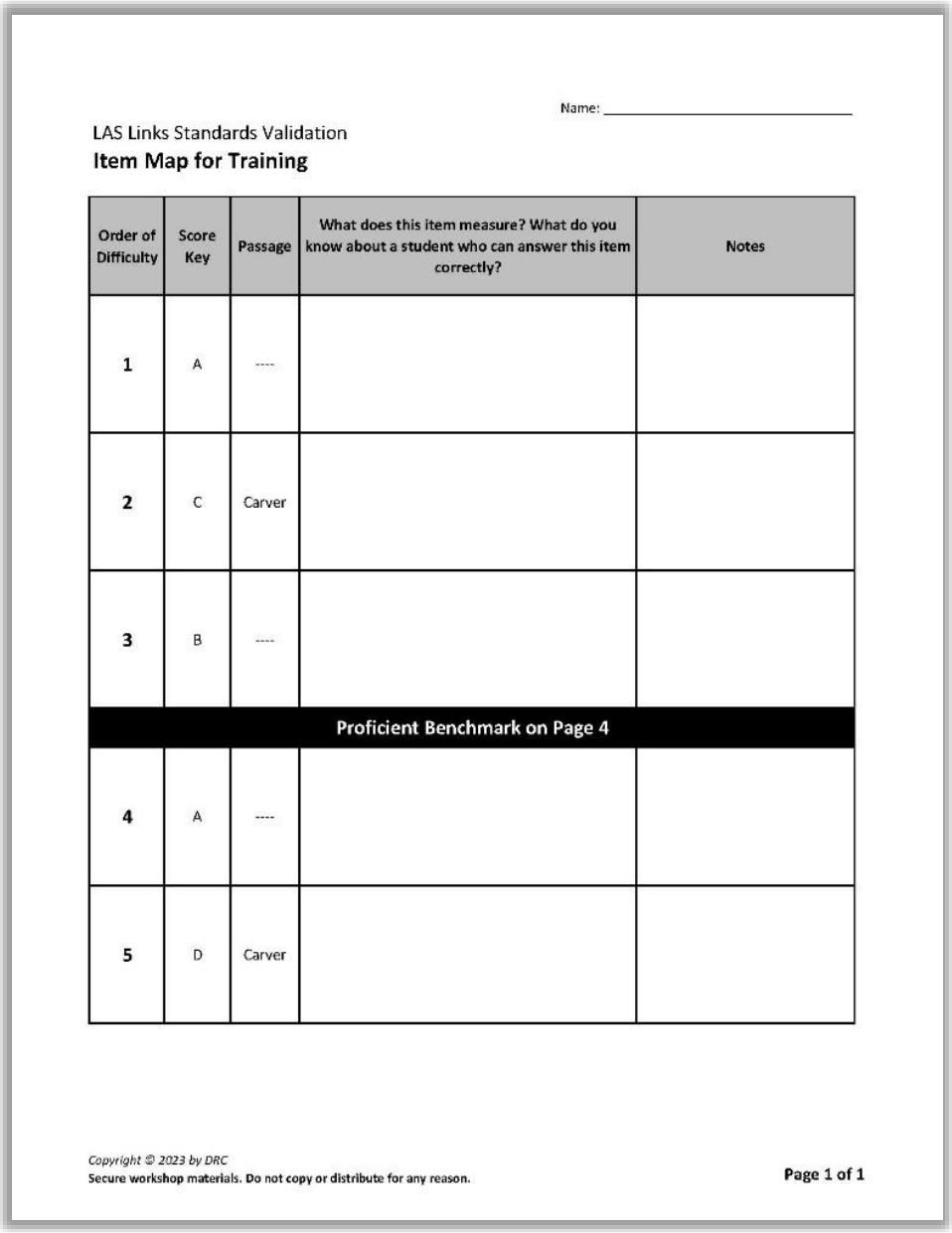
Copyright © 2026 by Data Recognition Corporation. All rights reserved.

then they considered the additional knowledge and skills needed to earn two points, and so on.

Item Maps

The item map summarizes information about the items in an OIB. For each item, the item map indicates the item sequence number and answer key. The workshop item maps incorporate secure test information and are not included in this document. However, Figure 2 shows the item map that was used during the participant training session and is included for illustration.

Figure 2. Training Item MapFi



Name: _____

LAS Links Standards Validation

Item Map for Training

Order of Difficulty	Score Key	Passage	What does this item measure? What do you know about a student who can answer this item correctly?	Notes
1	A	----		
2	C	Carver		
3	B	----		
Proficient Benchmark on Page 4				
4	A	----		
5	D	Carver		

Copyright © 2023 by DRC
Secure workshop materials. Do not copy or distribute for any reason.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Standard Setting Hub

Each participant could access the Hub, a specially designed website that contained materials accessible to workshop participants. Participants used the Hub to access selected materials (e.g., item maps), view test items, and enter standard setting recommendations. Access to the Hub was limited to workshop participants by DRC, and access was only allowed during the workshop.

DRC recognized that participants would benefit from having certain frequently referenced materials (e.g., the LAS Links proficiency level descriptors) available to them in hardcopy format. These materials were provided to participants on paper prior to the workshop and were also made available on the Hub.

Staff and Participants

Staff members from DRC served as facilitators and in support roles on all aspects of the standard setting workshop. These staff members did not contribute to the cut score recommendations during the workshop.

DRC Staff

Staff members from DRC trained participants in the standards validation methodology, facilitated workshop discussions, and answered participants' questions. Staff members from DRC Psychometric Services included Ricardo Mercado, Sr. Research Director; Sara Kendallen, Sr. Research Analyst; Michelle Boyer, Ph.D., Sr. Psychometrician; Julie Pointner, Research Coordinator; Chalin Walters, Statistical Analyst; Jess Smith, Ph.D., Sr. Psychometrician; Lee McKenna, Statistical Analyst; Keith Boughton, Ph.D., Sr. Psychometrician; Joanna Tomkowicz, Ph.D., Sr. Psychometrician; Christie Plackner, Sr. Research Director; Kim Hudson, Ph.D., Psychometrician; Daisy Ye, Statistical Analyst; Scott Li, Statistical Analyst; and Jenni Miska, Research Analyst. Huan Wang, Ph.D., Sr. Psychometrician, provided additional support with test data analyses.

Experts in language testing from DRC Test Development worked with participants to provide specialist support. These staff members included Jen Heller, Sr. Test Development Specialist; Megan Ormseth, Sr. Project Manager; Brittany Weber, Associate Test Development Specialist; and Kara Courtney, Vice President, ELA Curriculum and Assessment.

Project management for the workshop was provided by Jana Hilleren, Sr. Director of Language Solutions; Jonica Backes, Sr. Director, State Assessment Programs; and Maggie Frye, Manager of Meeting Planning.

Workshop Participants

The committee comprised a purposeful mix of educators with a variety of backgrounds. Special care was taken to promote diversity among participants in terms of background and location. Participants were asked to self-report their demographic characteristics

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

(e.g., ethnicity, number of years in the profession) as part of the pre-session participant survey.

Configuration of the Committee

The workshop committee was composed of a total of 42 educators. Forty-one completed the pre-workshop survey. Thirty-eight participants were female and three were male. Six were Black, 31 were white, one was American Indian and Alaska Native and Native Hawaiian or Other Pacific Islander, one was Asian, and three were of other races. Twenty-seven participants were teachers (three were special educators, and 21 were EL teachers), eight were consultants, and six were administrators of a district or other jurisdiction. 78% of participants worked in education more than 10 years, and 41% of them worked in education for more than 20 years.

Standards Validation Workshop

The workshop took place on Zoom. Before the workshop, participants were invited to a pre-workshop session to meet the workshop staff and test their computer equipment. The workshop was then conducted in three sessions:

January 17–19: Standards validation for Reading and Writing

January 24–26: Standards validation for Listening and Speaking

January 27: Review of the recommendations

Nearly all participants took part in the first two sessions, and five participants took part in the last. This section describes the standards validation method used in the first two sessions. Table 42 illustrates the committee configuration used each week of the standards validation. Table 43 provides a broad overview of the weekly agenda.

Table 51. Committee Configuration

Week 1: Reading & Writing	Week 2: Listening & Speaking
Kindergarten Reading & Writing	Kindergarten Listening & Speaking
Grade 2 Reading & Writing	Grade 2 Listening & Speaking
Grade 4 Reading & Writing	Grade 4 Listening & Speaking
Grade 7 Reading & Writing	Grade 7 Listening & Speaking
Grade 11 Reading & Writing	Grade 11 Listening & Speaking

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 52. Summary Agenda

Day	Time	Activity
Day 1	AM	Orientation and training, begin standards validation for Reading/Listening
	PM	Continue standards validation for Reading/Listening
Day 2	AM	Complete standards validation for Reading/Listening
	PM	Begin standards validation for Writing/Speaking
Day 3	AM	Continue standards validation for Writing/Speaking
	PM	Complete standards validation, review recommendations

Opening Session and Participant Training

The workshop began for all participants on January 17, 2023. All participants began the workshop with an opening session led by DRC. During this session, Ms. Hilleren from DRC welcomed the participants to the workshop and described the purpose of the workshop. Ms. Hilleren summarized the structure of LAS Links, described how DRC would soon publish LAS Links Forms E/F, and noted the goals of the workshop.

Mr. Mercado from DRC then introduced the standards validation methodology. Participants were introduced to the materials that would be used during the rest of the workshop. Participants were instructed that their goal for the workshop was to evaluate cut scores for the tests to determine whether they were still valid for continued use and, if not, to recommend more appropriate cut scores. Participants understood that they would consider the English language skills expected of students in each proficiency level and that they would engage in the Bookmark Procedure to make recommendations.

Discussion of the Threshold Students

DRC instructed participants to read the threshold proficiency level descriptors (PLDs) to consider the English language skills that students were expected to demonstrate at the threshold of each proficiency level.

Participants engaged in discussions about the English language skills they expected to be demonstrated by each of the four threshold students. The four threshold students were just in each proficiency level, from just Early Intermediate to just Above Proficient. To engage in these discussions, participants used the PLDs and their knowledge of students.

In each group, participants discussed the expectations for each proficiency level and the differences between them. To focus participants on the lines of demarcation between

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

the proficiency levels, participants were asked to discuss the English language skills that separated students in one proficiency level from those in another. For example, participants were asked to discuss the English language skills that separated the highest performing Early Intermediate students from the lowest performing Intermediate students.

Participants recorded their expectations for students at the thresholds of each proficiency level on electronic whiteboards. Participants were encouraged to review these descriptions frequently throughout the workshop and to consider the threshold students when they placed their bookmarks.

By the end of this discussion, participants had thoroughly considered the LAS Links proficiency level descriptors and threshold students and had worked toward a shared understanding of the types of skills that the threshold student for each proficiency level should have.

Presentation of Benchmarks

To evaluate the existing cut scores, participants were shown the 2005 cut scores as benchmarks in their item maps. Benchmarks refer to any policy-based information that is presented to participants that help participants make their cut-score recommendations. The use of benchmarks at standards validation is well established (Phillips, 2012), especially in the Bookmark Procedure (Lewis, Mitzel, Mercado, & Schulz, 2012; Ferrara, Lewis, & D'Brot, 2021). In this context, the use of the existing cut scores as benchmarks can be thought of as a way of gathering and evaluating validity evidence associated with the 2005 cut scores (McClarty et al., 2013). Thoughtful use of benchmarks can bring policy- and content-based information together in a meaningful way. Participants were shown benchmarks based on the existing cut scores before Round 1 of a two-round process.

Study of the OIBs and Item Maps

Participants examined the OIBs in terms of what each item measured. Participants were instructed to take notes on the item maps about the English language skills required to answer the items correctly. For multi-point items, participants considered each score point separately (e.g., considered the English language skills needed to earn two points out of three).

Secondary Training on Bookmark Placement

On the morning of January 18 (and in a refresher session on January 24), Mr. Mercado provided the participants with additional training on bookmark placement. Participants were reminded how bookmarks are associated with cut scores, how to evaluate the existing cut scores by examining the benchmarks, and how to make cut score recommendations using the OIB.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Following training, participants were tested on their understanding of the Bookmark Procedure with a short quiz, termed a mid-process evaluation. Afterwards, participants were provided the correct answers for the mid-process evaluation, as well as explanations of those answers.

Round 1

Participants then began to consider the benchmarks and whether they were placed appropriately in the OIB. To evaluate each benchmark, participants were asked to consider the English language skills measured by the items (and score points) appearing before the bookmark in the OIB, and participants considered whether those skills best reflected those expected of the threshold student. If there was good correspondence between the English language skills measured by the items before the benchmark and those expected of the threshold student, participants were instructed to retain that position in the OIB as their bookmark. Otherwise, participants were instructed to move their bookmark forward or backward in the OIB, one page at a time, until there was better correspondence.

On their own, participants then made their Round 1 bookmark placements. To do so, they referred to their OIB, item map, threshold student descriptions, and PLDs. Participants recorded their bookmark placements electronically.

Presentation of Round 1 Recommendations

Participants were then presented with a summary of their Round 1 recommendations. Specifically, participants were shown the median bookmark placed for each cut score. Participants were also shown a histogram of the group's Round 1 bookmark placements.

Round 2

Participants then discussed the rationales behind their Round 1 bookmarks. Participants were instructed to engage in a discussion based on English language skills and to focus on items around their Round 1 bookmarks. Participants referred to their OIB, item map, threshold student descriptions, and the PLDs throughout the discussions.

Following this discussion, participants made their Round 2 bookmark placements on their own. Participants were reminded that they were free to retain their Round 1 bookmarks or to change any of them; but in either case, participants would need to have content-based rationales for their decisions.

Presentation of Round 2 Recommendations

Participants were then presented with a summary of their Round 2 recommendations, including the median bookmarks and a histogram of the bookmarks placed in Round 2. Participants were told that the median Round 2 bookmarks would comprise the committee's recommendation.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Repeat the Process

After completing the process for Reading, participants repeated it for Writing, Listening, and Speaking.

For Writing and Speaking, participants were shown the scoring guides (rubrics) for CR items. For Writing, Listening, and Speaking, participants experienced the audio stimuli through headphones, and they had access to written transcripts for longer listening stimuli.

Above Proficient Cut Score Review for Kindergarten Writing

Due to limitations in the availability of highly difficult content on the kindergarten Writing test, two additional modifications to the Bookmark procedure were applied for the purpose of evaluating the Proficient and Above Proficient cut scores.

When the participants began their work, the existing cut scores for Proficient and Above Proficient were on the same item in the OIB. That item was also the last item in the OIB. Therefore, participants could only reasonably begin the process by reviewing the Proficient cut score. However, after participants had made their recommendation that the final Proficient cut score should be lower than the original based on a strict application of the Bookmark procedures described herein, there was an opportunity to consider whether the Above Proficient cut score should also be adjusted downward.

Accordingly, DRC facilitators led the participants in the kindergarten group through a discussion of the Above Proficient expectations, threshold students, and the items in the OIB beyond the adjusted Proficient cut score. Participants were then asked for their recommendations about adjusting the Above Proficient cut score, both informally during discussion and formally through a survey response.

Review of Recommendations

After making their cut score recommendations, participants were presented with the cut score recommendations for all four domains. Participants were reminded that they could consider adjustments to the cut scores, if needed, to promote better articulation (consistency) of the cut scores across grades. Specifically, participants were told that representatives from each group would meet on January 27 to review the recommendations and, as needed, recommend adjustments to them.

Workshop Evaluation

All participants were thanked for their time and effort during the standards validation. To conclude the workshop each week, participants were asked to complete a post-workshop evaluation. Selected results are presented later in this section.

Results

The standard setting was conducted according to the plans created by DRC prior to the workshop. The results of the workshop are presented in this section.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Participants' Recommendations after Round 2

Table 44 shows participants' recommendations from Round 2 of the Bookmark Procedure. The committee's cut score recommendation is defined as the median cut score recommendation from each group of participants. The cut score recommendations are shown on the scale-score metric.

Table 53. Median Round 2 Cut Score Recommendations, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	439	426	369	347
	Intermediate	461	443	385	412
	Proficient	492	470	421	437
	Above Prof.	502	490	448	443
Grade 2	Early Int.	443	442	435	425
	Intermediate	473	463	472	475
	Proficient	504	492	498	504
	Above Prof.	535	536	534	544
Grade 4	Early Int.	451	457	468	434
	Intermediate	475	484	504	498
	Proficient	510	525	535	533
	Above Prof.	559	581	588	584
Grade 7	Early Int.	462	463	502	447
	Intermediate	475	492	530	498
	Proficient	513	528	560	548
	Above Prof.	560	579	608	594
Grade 11	Early Int.	470	480	509	455
	Intermediate	479	516	547	501
	Proficient	515	545.5	583	550
	Above Prof.	541	587	634	585

Participants' Recommendations after the Review Session

During the January 27th meeting, participants reviewed all recommended cut scores for each domain and overall grade spans. As a whole, the group noted the articulation of recommendations over grade spans within domains, considered the amount of error

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

associated with each recommendation, reviewed test items, and discussed the overall reasonableness of the round two recommendations.

Participants included the following people:

- Rachel Lease, general education teacher, Connecticut, grade K
- Kristin Gonzalez, general education teacher, Connecticut, grade 2
- Chalise Ross, ESOL coordinator K-12, Connecticut, grade 4
- Michelle Johnston, administrator of a K-12 district, Mississippi, grade 7
- Rebekah Daly Smith, consultant, Minnesota, grade 11

The group indicated particular concern about the existing cut scores for Proficient and Above Proficient in kindergarten Writing. The group felt strongly that the existing cut scores were too high. This determination was based on a number of factors.

First, they pointed to the round 1 and round 2 results which showed a large change in the recommended cut scores. With the knowledge that the Bookmark method was implemented with fidelity, the group further discussed the many changes that have occurred in learning and assessment of very young English language learners including advancing technologies, traits of young English language learners, changes in English language instruction and classroom environments, and overall impact. Very few kindergarten students have been able to reach the Proficient and Above Proficient levels based on the historical cut scores.

Collectively, the group advised that the round 2 recommended changes were substantial and necessary.

Table 45 shows the cut score recommendations following the review session. These cut scores, as expressed on the scale metric, are considered the final recommendations of the standards validation committee.

Table 54. Cut Score Recommendations from the Review Session, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	439	426	369	347
	Intermediate	461	443	385	412
	Proficient	492	470	421	437
	Above Prof.	534	490	448	443

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Cut Score	Speaking	Listening	Reading	Writing
Grade 2	Early Int.	443	442	435	425
	Intermediate	473	463	472	475
	Proficient	504	492	498	504
	Above Prof.	556	536	534	544
Grade 4	Early Int.	451	457	468	434
	Intermediate	475	484	504	498
	Proficient	510	525	535	533
	Above Prof.	559	581	588	584
Grade 7	Early Int.	462	463	502	447
	Intermediate	475	492	530	498
	Proficient	513	528	560	548
	Above Prof.	560	597	608	594
Grade 11	Early Int.	469	480	509	455
	Intermediate	479	516	547	501
	Proficient	515	544	583	550
	Above Prof.	560	625	634	585

Statistical Error Values Associated with the Recommendations

Two independent sources of statistical error can be associated with the cut score recommendations: error associated with the variance in participants' cut score recommendations during the Bookmark Procedure (also known as the standard error of the cut score or SEcut) and error associated with the level of measurement precision of the test instrument itself (also known as the conditional standard error of measurement or CSEM). These two sources of error can be combined to create a single value, SEcombined.

Standard Error of the Cut Score (SEcut)

Even in Round 2, the cut score recommendations varied among participants. Table 46 shows the standard error associated with the Round 2 recommendations; all are expressed on the scale metric.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 55. Standard Error of Round 2 Cut Score Recommendations, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	0.00	0.00	1.25	4.57
	Intermediate	0.73	0.27	0.65	0.00
	Proficient	0.12	0.73	0.16	0.57
	Above Prof.	4.00	4.06	1.04	0.00
Grade 2	Early Int.	0.00	0.82	0.00	0.00
	Intermediate	0.96	0.12	0.19	0.42
	Proficient	1.66	0.94	1.00	0.45
	Above Prof.	2.38	1.00	2.36	0.00
Grade 4	Early Int.	0.72	0.44	0.67	1.78
	Intermediate	0.67	1.45	0.29	1.73
	Proficient	0.83	1.89	0.18	0.17
	Above Prof.	0.00	3.78	0.60	2.04
Grade 7	Early Int.	0.84	2.13	0.38	1.75
	Intermediate	1.23	1.24	0.00	1.00
	Proficient	0.00	0.85	0.35	1.64
	Above Prof.	0.83	3.32	0.50	1.28
Grade 11	Early Int.	0.77	0.00	1.73	0.00
	Intermediate	0.60	0.00	0.68	0.50
	Proficient	0.00	4.26	2.25	0.67
	Above Prof.	3.85	10.72	5.00	2.40

Rationale for Using ±2 Conditional Standard Errors of Measurement

The CSEM quantifies the random variability in test scores in actual scale score units, making them useful for test score users. CSEMs allow users to make statements regarding the precision of individual test scores. Specifically, they help place ‘reasonable limits’ (Gulliksen, 1950) around observed scores through the construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, X, and adding and subtracting a multiplicative factor of the CSEM (e.g., ±2 CSEM).

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Students with a given true score will have observed scores that fall between ±1 CSEM about two-thirds of the time. For example, if a student were tested many times, the student's scores would likely be similar (but not identical) each time, and the observed scores would be expected to fall within a range of ±1 CSEM about two-thirds of the time. For ±2 CSEM confidence intervals, the student's observed score is expected to fall within this range approximately 95 percent of the time. As a reminder, critical values of ±1.96 (rounded to ±2) are the 95% confidence interval for z tests (i.e., alpha = .05.) The best practices for evaluating score differences in educational measurement are further discussed in the NCES Statistical Standards (see Standard 5.1.3, NCES, 2024).

When we adjust a cut score within ±2 CSEMs, we are moving it within a range where we have a 95% level of confidence that (a) the cut score difference lays within the "reasonable limits" associated with the level of statistical precision on the test scale; and (b) the cut score difference falls within the range of expected performance over multiple theoretical administrations for a student at a given cut score.

Table 47 shows the CSEM values associated with the committee's cut score recommendations, as expressed on the test scale.

Table 56. CSEM Associated with Cut Score Recommendations by Grade and Domain (in scale score units)

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	8	12	14	21
	Intermediate	8	14	12	10
	Proficient	12	21	11	13
	Above Prof.	56	19	17	15
Grade 2	Early Int.	8	15	23	15
	Intermediate	7	14	12	9
	Proficient	10	16	12	11
	Above Prof.	67	73	24	24
Grade 4	Early Int.	10	24	23	17
	Intermediate	9	20	16	15
	Proficient	8	19	15	15
	Above Prof.	14	38	20	18

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Cut Score	Speaking	Listening	Reading	Writing
Grade 7	Early Int.	9	23	20	22
	Intermediate	9	20	14	16
	Proficient	9	22	14	16
	Above Prof.	13	44	21	16
Grade 11	Early Int.	7	26	13	20
	Intermediate	6	21	13	16
	Proficient	8	22	18	17
	Above Prof.	19	40	35	21

Combined Standard Error (SEcombined)

These two independent sources of error can be combined into a single value, SEcombined. Table 48 shows the SEcombined values associated with the committee's cut score recommendations as expressed on the test scale.

Table 57. Combined Standard Error of Cut Score Recommendations, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	8.00	12.00	14.06	21.49
	Intermediate	8.03	14.00	12.02	10.00
	Proficient	12.00	21.01	11.00	13.01
	Above Prof.	56.14	19.43	17.03	15.00
Grade 2	Early Int.	8.00	15.02	23.00	15.00
	Intermediate	7.07	14.00	12.00	9.01
	Proficient	10.14	16.03	12.04	11.01
	Above Prof.	67.04	73.01	24.12	24.00
Grade 4	Early Int.	10.03	24.00	23.01	17.09
	Intermediate	9.02	20.05	16.00	15.10
	Proficient	8.04	19.09	15.00	15.00
	Above Prof.	14.00	38.19	20.01	18.12

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Cut Score	Speaking	Listening	Reading	Writing
Grade 7	Early Int.	9.04	23.10	20.00	22.07
	Intermediate	9.08	20.04	14.00	16.03
	Proficient	9.00	22.02	14.00	16.08
	Above Prof.	13.03	44.13	21.01	16.05
Grade 11	Early Int.	7.04	26.00	13.11	20.00
	Intermediate	6.03	21.00	13.02	16.01
	Proficient	8.00	22.41	18.14	17.01
	Above Prof.	19.39	41.41	35.36	21.14

Comparison of Existing and Recommended Cut Scores

Before the standards validation, DRC noted that it was unlikely that participants at the standards validation would recommend cut scores exactly equal to the existing LAS Links cut scores. Several factors led to this expectation, including the fact that there were different and smaller numbers of participants than were involved in the original 2005 standard setting, that standards validation participants were examining different test items than used at the original standard setting, and that the mode of the workshop (virtual) was different than the mode of the original standard setting (in-person).

Because of these factors, DRC noted in advance that if participants recommended cut scores within a range of ±2 SEcombined of the existing cut scores, then this would be interpreted as evidence that the existing cut scores were still valid for continued use. If participants recommended cut scores outside this range, then DRC would investigate those cut scores further.

Table 49 presents the differences between the committee's recommended cut scores and the existing cut scores (these differences are expressed on the test scales). Table 50 presents these differences as multiples of the cut scores' respective SEcombined values.

Table 58. Differences between Recommended Cut Scores and Existing Cut Scores, in Terms of Scale Score Points, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	9	0	14	0
	Intermediate	0	-1	4	-5
	Proficient	-4	0	0	-51
	Above Prof.	-16	-30	-27	-73

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Cut Score	Speaking	Listening	Reading	Writing
Grade 2	Early Int.	0	0	0	0
	Intermediate	0	1	0	0
	Proficient	-5	0	-1	0
	Above Prof.	-1	0	-13	0
Grade 4	Early Int.	2	0	0	0
	Intermediate	0	0	0	0
	Proficient	0	0	0	0
	Above Prof.	0	0	0	0
Grade 7	Early Int.	11	0	0	0
	Intermediate	-2	0	0	0
	Proficient	0	-5	0	0
	Above Prof.	0	9	0	2
Grade 11	Early Int.	17	0	0	5
	Intermediate	0	0	0	0
	Proficient	-2	-11	0	0
	Above Prof.	0	-5	0	-10

Table 59. Differences between Recommended Cut Scores and Existing Cut Scores, as a Multiple of SEcombined, by Grade and Domain

Grade	Cut Score	Speaking	Listening	Reading	Writing
Kindergarten	Early Int.	1.13	0.00	1.00	0.00
	Intermediate	0.00	-0.07	0.33	-0.50
	Proficient	-0.33	0.00	0.00	-3.92
	Above Prof.	-0.29	-1.54	-1.59	-4.87
Grade 2	Early Int.	0.00	0.00	0.00	0.00
	Intermediate	0.00	0.07	0.00	0.00
	Proficient	-0.49	0.00	-0.08	0.00
	Above Prof.	-0.01	0.00	-0.54	0.00

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Cut Score	Speaking	Listening	Reading	Writing
Grade 4	Early Int.	0.20	0.00	0.00	0.00
	Intermediate	0.00	0.00	0.00	0.00
	Proficient	0.00	0.00	0.00	0.00
	Above Prof.	0.00	0.00	0.00	0.00
Grade 7	Early Int.	1.22	0.00	0.00	0.00
	Intermediate	-0.22	0.00	0.00	0.00
	Proficient	0.00	-0.23	0.00	0.00
	Above Prof.	0.00	0.20	0.00	0.12
Grade 11	Early Int.	2.41	0.00	0.00	0.25
	Intermediate	0.00	0.00	0.00	0.00
	Proficient	-0.25	-0.49	0.00	0.00
	Above Prof.	0.00	-0.12	0.00	-0.47

With several notable exceptions, participants' recommended cut scores were highly consistent with the existing cut scores. Of the 80 cut scores actively considered by participants, the committee recommended the same cut scores as currently exist 52 times, scores within a range of ±0.5 SEcombined of the current scores 72 times, scores within a range of ±1.0 SEcombined of the current scores 73 times, and scores within a range of ±2.0 SEcombined of the current scores 77 times.

Examination of the Kindergarten Cut Scores

As shown in Table 50, the greatest deviations from the existing cut scores were among the recommendations for kindergarten. Of the 16 cut scores actively considered, six differed by at least ±1.0 SEcombined and two differed by more than ±2.0 SEcombined. Based on these recommendations, DRC investigated the cut scores further.

During the January 27th meeting, the group also indicated particular concern about the existing cut scores for Proficient and Above Proficient in kindergarten Writing. The group felt strongly that the existing cut scores were too high. DRC heard the group's concern and reasonings, particularly the many changes that have recently occurred in the learning and assessment of very young English language learners including:

- advancing technologies,
- new traits of young English language learners,
- changes in English language instruction and classroom environments, and

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- overall impact in which very few kindergarten students have been able to reach the Proficient and Above Proficient levels based on the historical cut scores.

The group's content-based judgments, along with these additional considerations, were treated as appropriate and compelling reasons to implement adjustments to the kindergarten Writing cut score for Proficient and Above Proficient.

Consideration of the Grade 1 Cut Scores

With the changes made to the Kindergarten cut scores as described above, DRC then considered the grade 1 cut scores. Although standards validation did not directly consider these cut scores, DRC acknowledged the following:

- Students in kindergarten and grade 1 take substantially similar test forms.
- The changes to language demands affecting Kindergarteners likely also affect grade 1 students.
- The changes to the test items affecting Kindergarteners also affect grade 1 students.
- The PLDs for kindergarten and grade 1 are similar, but also somewhat different.

The existing sets of LAS Links cut scores for kindergarten and grade 1 are very similar: the cut scores for students in these two grades are similar, suggesting the English language skills expected of students in these grades is similar. However, the PLDs for these two grades have subtle but important differences.

To reflect the changes to the kindergarten cut scores and reflect the voice of the standards validation committee, DRC interpolated cut scores for grade 1 by taking the simple average of the cut scores for kindergarten and grade 1.

The grade 1 cut scores in Table 41 preserve the same scale-score differences between the kindergarten and grade 1 cut scores found in the existing cut scores and were reviewed and validated by a group of 4 DRC test development experts, using the same process as described for kindergarten.

Table 51 summarizes the adjustments that will be made to the LAS Links cuts scores for Forms E and F in kindergarten and grade 1 Writing and the associated composite scores.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 60. Cut Score Adjustments for LAS Links Forms E/F in Kindergarten and Grade 1

Grade	Cut Score	Writing	Overall	Literacy	Productive
Kindergarten	Early Intermediate	347	389	351	388
	Intermediate	417	425	399	439
	Proficient	**437*	**456*	**429*	**466*
	Above Proficient	**452*	**499*	**463*	**501*
Grade 1	Early Intermediate	355	394	357	393
	Intermediate	435	433	410	448
	Proficient	**470*	**466*	**446*	**483*
	Above Proficient	**498*	**512*	**488*	**524*

Note. Asterisks indicate which cut scores were adjusted.

Examination of the Grade 11 Speaking Cut Scores

Participants in the grade 11 group recommended a substantially higher cut score for Early Intermediate than had been previously implemented. Based on this recommendation, DRC investigated the cut score further.

Due to a number of process factors, DRC had low confidence in the recommended adjustment. These factors included the following:

- absenteeism that led to one participant being moved from the grade 7 group to grade 11 to compensate and resulting in 5 participants making the final recommendation
- too few items with difficulty in the Early Intermediate ranges

As a result, DRC sought to achieve a difficult balance between the level of confidence that the recommended changes represented true departures from the original cut scores against the risk of disruption to current and longitudinal test score interpretations for the varied purposes that LAS Links serves. In this case, for the reasons listed, confidence in the recommended change to the Early Intermediate cut score was too low to warrant a system wide disruption in interpretation and use of proficiency level designations at the high school level.

<u>After the Standard Setting</u>

After the standard setting, DRC reviewed participants' cut score recommendations. DRC noted that the standards validation had been conducted according to industry best practices and that the cut score recommendations were collected in a defensible

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

manner. As the publisher of LAS Links, DRC has a responsibility to evaluate participants' cut score recommendations and to implement cut scores that best reflect the English-language skills expected of students in each proficiency level. Therefore, the conclusions drawn from panelist recommendations were as follows: (a) change the kindergarten cut score, (b) change the grade 1 cut scores, and (c) preserve the remaining cuts.

Table 52 shows the final cut scores for LAS Links Forms E and F with adjusted cut scores indicated with an asterisk.

Table 61. Final Cut Scores for LAS Links Forms E and F

Domain/Composite	Grade	Early Intermediate	Intermediate	Proficient	Above Proficient
Reading	K	355	381	421	475
	1	360	385	423	479
	2	435	472	499	547
	3	436	474	504	549
	4	468	504	535	588
	5	470	505	536	590
	6	501	529	559	608
	7	502	530	560	608
	8	502	532	561	608
	9	507	545	581	632
	10	508	546	582	633
	11	509	547	583	634
	12	510	548	584	635
Listening	K	426	444	470	520
	1	432	450	476	521
	2	442	462	492	536
	3	447	468	504	546
	4	457	484	525	581
	5	458	490	528	584
	6	462	489	532	586
	7	463	492	533	588
	8	467	498	535	590
	9	471	509	546	625
	10	475	511	550	627
	11	480	516	555	630
	12	481	517	560	633

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Domain/Composite	Grade	Early Intermediate	Intermediate	Proficient	Above Proficient
Speaking	K	430	461	496	550
	1	432	462	496	551
	2	443	473	509	557
	3	443	474	509	558
	4	449	475	510	559
	5	449	475	511	559
	6	451	476	512	560
	7	451	477	513	560
	8	451	477	514	560
	9	452	478	515	560
	10	452	478	516	560
	11	452	479	517	560
	12	452	479	518	560
Writing	K	347	417	*437	*452
	1	355	435	*470	*498
	2	425	475	504	544
	3	428	484	529	560
	4	434	498	533	584
	5	435	499	538	585
	6	447	498	548	591
	7	447	498	548	592
	8	448	499	548	593
	9	449	500	549	594
	10	449	500	549	594
	11	450	501	550	595
	12	451	502	550	596
Overall	K	389	425	*456	*499
	1	394	433	*466	*512
	2	436	470	501	546
	3	438	475	511	553
	4	452	490	525	578
	5	453	492	528	579
	6	465	498	537	586
	7	465	499	538	587
	8	467	501	539	587
	9	469	508	547	602
	10	471	508	549	603
	11	472	510	551	604
	12	473	511	553	606

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Domain/Composite	Grade	Early Intermediate	Intermediate	Proficient	Above Proficient
Oral	K	430	461	487	526
	1	432	463	490	530
	2	443	470	495	540
	3	444	471	505	548
	4	450	478	514	575
	5	452	485	516	580
	6	455	481	518	575
	7	460	485	521	580
	8	465	492	525	582
	9	465	490	525	561
	10	468	495	527	566
	11	471	497	530	567
	12	472	500	531	569
Productive	K	388	439	*466	*501
	1	393	448	*483	*524
	2	434	474	506	550
	3	435	479	519	559
	4	441	486	521	571
	5	442	487	524	572
	6	449	487	530	575
	7	449	487	530	576
	8	449	488	531	576
	9	450	489	532	577
	10	450	489	532	577
	11	451	490	533	577
	12	451	490	534	578
Literacy	K	351	399	*429	*463
	1	357	410	*446	*488
	2	430	473	501	545
	3	432	479	516	554
	4	451	501	534	586
	5	452	502	537	587
	6	474	513	553	599
	7	474	514	554	600
	8	475	515	554	600
	9	478	522	565	613
	10	478	523	565	613
	11	479	524	566	614
	12	480	525	567	615

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Domain/Composite	Grade	Early Intermediate	Intermediate	Proficient	Above Proficient
Comprehension	K	386	412	449	467
	1	390	416	452	486
	2	448	473	495	531
	3	452	482	500	533
	4	485	506	526	563
	5	491	509	531	573
	6	499	514	540	574
	7	500	517	546	576
	8	501	519	553	579
	9	512	534	567	597
	10	514	536	569	606
	11	515	540	575	608
	12	515	540	575	610

Note. Asterisks indicate which cut scores were adjusted.

Evidence of Procedural Validity

The standard setting was conducted using a diverse, well-trained committee and was perceived as valid by participants. This section supports these claims.

Committee Training

During the standard setting workshop, it was clear to the facilitators that participants understood how to make judgments as part of the standard setting methodology (e.g., Bookmark ratings).

To confirm participants' knowledge of the methodology, participants were given a short quiz each week, termed a mid-process evaluation, after training. Participants answered items 1–5 on the mid-process evaluation correctly most of the time. This indicates that, on the whole, participants were well prepared to make judgments and that the training was effective. Results for each week can be found in Table 53.

Table 62. Participants Answering Each Item Correctly on the Weeks One and Two Training Quiz

Week	Item 1	Item 2	Item 3	Item 4	Item 5
Week 1	89%	100%	53%	100%	84%
Week 2	95%	100%	47%	100%	90%

Most participants struggled with question #3 which asked participants to identify which performance level a student would be in if they had a particular number of skills identified in the OIB. Participants saw the results of their training quizzes and correct

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

answers immediately. DRC retrained the committee following these results, and participants voiced their understanding.

The mid-process evaluation also asked participants whether they felt the goals of the standard setting were made clear and whether they felt ready to proceed, or if they had additional questions to be answered before they made their bookmark placements. These questions were answered before the committee.

Participants' Perceived Validity of the Workshop

Participants indicated their perceived validity of the workshop and their recommendations as part of the post-workshop evaluation. Hambleton (2001) noted that evaluations are important evidence for establishing the validity of performance levels.

Generally, participants were satisfied with their recommendations and with the workshop as a whole. Tables 54 and 55 show participants' level of satisfaction with their recommendations each week. Particularly, participants understood the connection between the threshold students and their cut score recommendations, and participants generally agreed that the final recommendations reflected the work of the standard setting committee.

Table 63. Participants' Agreement with Various Statements on the Post-Workshop Evaluation Regarding Their Satisfaction with the Process and the Final Recommendations, Week One

Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Agree + Strongly Agree
The proficiency level descriptors (PLDs) were useful during the process.	0%	0%	30%	70%	**100%**
My opinions were valued by my group.	0%	0%	27%	73%	**100%**
The descriptions of the threshold students were useful during the process.	0%	0%	30%	70%	**100%**
The facilitator provided clear instructions.	0%	0%	22%	78%	**100%**
I believe this process will yield defensible cut scores.	0%	2%	28%	70%	**98%**

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table 64. Participants' Agreement with Various Statements on the Post-Workshop Evaluation Regarding Their Satisfaction with the Process and the Final Recommendations, Week Two

Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Agree + Strongly Agree
The proficiency level descriptors (PLDs) were useful during the process.	0%	0%	26%	74%	**100%**
My opinions were valued by my group.	0%	0%	24%	76%	**100%**
The descriptions of the threshold students were useful during the process.	0%	0%	18%	82%	**100%**
The facilitator provided clear instructions.	0%	0%	24%	76%	**100%**
I believe this process will yield defensible cut scores.	0%	3%	29%	68%	**97%**

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 6: Validity Evidence Related to the Relationship of Scores with Other Variables

This chapter provides evidence that the relationship between LAS Links domain and composite scale scores suggests an appropriate relationship, along with recommendations for how educational authorities that use LAS Links can further demonstrate the criterion-related validity of LAS Links scores in their own unique contexts.

Where tests are intended to measure similar constructs, stronger correlations among their test scores might be expected; and where tests are intended to measure different constructs, weaker correlations might be expected. The Standards (2014) refer to this as convergent and discriminant validity, respectively.

The correlations among LAS Links domain and composite scores are provided in Table 56. Results show that the highest correlations tend to be between the overall score and each domain score (see Overall column correlations in Table 56 relative to the domain and remaining composite scores). This is desirable as the overall score is most often used for decisions about student access to English language supports, so evidence that the overall scores relates well to each related component represent evidence in support of convergent validity.

Conversely, it is expected that individual domain score relationships may be relatively weak between domains that measure components that are arguably more different from each other. For example, the relationship between reading and speaking scores tends to be quite low as might be expected, demonstrating reasonable discriminant validity.

Table 65. Domain and Composite Scale Score Correlations

Grade	Variable	Overall	Listening	Speaking	Reading	Writing	Comp.	Literacy	Oral	Prod.
K	**Overall**	1	-	-	-	-	-	-	-	-
	Listening	0.7	1	-	-	-	-	-	-	-
	Speaking	0.7	0.53	1	-	-	-	-	-	-
	Reading	0.77	0.42	0.38	1	-	-	-	-	-
	Writing	0.82	0.44	0.48	0.52	1	-	-	-	-
	Comp	0.88	0.75	0.5	0.89	0.55	1	-	-	-
	Literacy	0.92	0.48	0.49	0.84	0.88	0.82	1	-	-
	Oral	0.8	0.87	0.85	0.45	0.52	0.72	0.55	1	-
	Prod.	0.9	0.53	0.74	0.53	0.92	0.61	0.84	0.73	1

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Variable	Overall	Listening	Speaking	Reading	Writing	Comp.	Literacy	Oral	Prod.
1	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.77	1	-	-	-	-	-	-	-
	Speaking	0.78	0.58	1	-	-	-	-	-	-
	Reading	0.89	0.58	0.6	1	-	-	-	-	-
	Writing	0.92	0.6	0.65	0.8	1	-	-	-	-
	Comp	0.94	0.85	0.66	0.9	0.78	1	-	-	-
	Literacy	0.96	0.62	0.66	0.93	0.96	0.88	1	-	-
	Oral	0.87	0.9	0.85	0.66	0.69	0.86	0.71	1	-
	Prod.	0.95	0.64	0.83	0.79	0.96	0.8	0.93	0.81	1
2	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.76	1	-	-	-	-	-	-	-
	Speaking	0.73	0.61	1	-	-	-	-	-	-
	Reading	0.86	0.55	0.51	1	-	-	-	-	-
	Writing	0.89	0.6	0.64	0.66	1	-	-	-	-
	Comp	0.93	0.79	0.6	0.93	0.71	1	-	-	-
	Literacy	0.97	0.62	0.62	0.9	0.9	0.9	1	-	-
	Oral	0.83	0.93	0.84	0.59	0.68	0.79	0.69	1	-
	Prod.	0.92	0.66	0.8	0.67	0.96	0.73	0.89	0.79	1
3	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.81	1	-	-	-	-	-	-	-
	Speaking	0.75	0.61	1	-	-	-	-	-	-
	Reading	0.91	0.65	0.58	1	-	-	-	-	-
	Writing	0.88	0.62	0.62	0.75	1	-	-	-	-
	Comp	0.96	0.85	0.64	0.94	0.76	1	-	-	-
	Literacy	0.96	0.68	0.64	0.94	0.91	0.92	1	-	-
	Oral	0.87	0.94	0.83	0.68	0.67	0.85	0.73	1	-
	Prod.	0.92	0.66	0.82	0.76	0.95	0.78	0.9	0.79	1
4	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.83	1	-	-	-	-	-	-	-
	Speaking	0.8	0.61	1	-	-	-	-	-	-
	Reading	0.9	0.66	0.64	1	-	-	-	-	-
	Writing	0.91	0.64	0.71	0.77	1	-	-	-	-
	Comp	0.95	0.89	0.68	0.92	0.77	1	-	-	-
	Literacy	0.96	0.69	0.72	0.93	0.94	0.89	1	-	-
	Oral	0.91	0.94	0.82	0.72	0.73	0.9	0.77	1	-
	Prod.	0.93	0.67	0.85	0.78	0.97	0.79	0.93	0.81	1
5	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.84	1	-	-	-	-	-	-	-
	Speaking	0.79	0.6	1	-	-	-	-	-	-
	Reading	0.9	0.67	0.62	1	-	-	-	-	-
	Writing	0.89	0.62	0.7	0.76	1	-	-	-	-
	Comp	0.95	0.9	0.66	0.91	0.75	1	-	-	-
	Literacy	0.96	0.69	0.71	0.93	0.94	0.89	1	-	-
	Oral	0.91	0.94	0.81	0.72	0.71	0.9	0.76	1	-
	Prod.	0.92	0.65	0.85	0.76	0.96	0.77	0.92	0.8	1

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Variable	Overall	Listening	Speaking	Reading	Writing	Comp.	Literacy	Oral	Prod.
6	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.83	1	-	-	-	-	-	-	-
	Speaking	0.79	0.6	1	-	-	-	-	-	-
	Reading	0.85	0.62	0.57	1	-	-	-	-	-
	Writing	0.89	0.63	0.69	0.69	1	-	-	-	-
	Comp	0.94	0.89	0.64	0.89	0.73	1	-	-	-
	Literacy	0.95	0.68	0.69	0.9	0.93	0.88	1	-	-
	Oral	0.91	0.94	0.81	0.66	0.71	0.88	0.75	1	-
	Prod.	0.92	0.66	0.85	0.69	0.96	0.75	0.91	0.8	1
7	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.87	1	-	-	-	-	-	-	-
	Speaking	0.81	0.64	1	-	-	-	-	-	-
	Reading	0.88	0.69	0.62	1	-	-	-	-	-
	Writing	0.9	0.69	0.7	0.72	1	-	-	-	-
	Comp	0.95	0.91	0.68	0.92	0.76	1	-	-	-
	Literacy	0.96	0.74	0.71	0.91	0.93	0.9	1	-	-
	Oral	0.93	0.94	0.85	0.72	0.75	0.9	0.79	1	-
	Prod.	0.93	0.72	0.87	0.74	0.96	0.78	0.91	0.84	1
8	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.86	1	-	-	-	-	-	-	-
	Speaking	0.81	0.63	1	-	-	-	-	-	-
	Reading	0.88	0.68	0.63	1	-	-	-	-	-
	Writing	0.9	0.68	0.72	0.73	1	-	-	-	-
	Comp	0.95	0.91	0.68	0.91	0.76	1	-	-	-
	Literacy	0.96	0.73	0.72	0.92	0.93	0.9	1	-	-
	Oral	0.93	0.94	0.83	0.72	0.76	0.9	0.79	1	-
	Prod.	0.93	0.7	0.88	0.73	0.96	0.78	0.91	0.84	1
9	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.82	1	-	-	-	-	-	-	-
	Speaking	0.84	0.62	1	-	-	-	-	-	-
	Reading	0.87	0.63	0.66	1	-	-	-	-	-
	Writing	0.92	0.65	0.76	0.76	1	-	-	-	-
	Comp	0.93	0.9	0.69	0.89	0.76	1	-	-	-
	Literacy	0.96	0.67	0.76	0.91	0.95	0.87	1	-	-
	Oral	0.92	0.92	0.85	0.7	0.75	0.9	0.77	1	-
	Prod.	0.95	0.67	0.89	0.76	0.96	0.78	0.93	0.84	1
10	Overall	1	-	-	-	-	-	-	-	-
	Listening	0.85	1	-	-	-	-	-	-	-
	Speaking	0.82	0.63	1	-	-	-	-	-	-
	Reading	0.89	0.68	0.67	1	-	-	-	-	-
	Writing	0.91	0.68	0.73	0.78	1	-	-	-	-
	Comp	0.95	0.91	0.7	0.91	0.79	1	-	-	-
	Literacy	0.96	0.72	0.74	0.93	0.95	0.89	1	-	-
	Oral	0.93	0.94	0.83	0.73	0.76	0.91	0.79	1	-
	Prod.	0.94	0.7	0.87	0.78	0.96	0.8	0.93	0.84	1

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade	Variable	Overall	Listening	Speaking	Reading	Writing	Comp.	Literacy	Oral	Prod.
11	**Overall**	1	-	-	-	-	-	-	-	-
	Listening	0.85	1	-	-	-	-	-	-	-
	Speaking	0.8	0.62	1	-	-	-	-	-	-
	Reading	0.9	0.69	0.66	1	-	-	-	-	-
	Writing	0.91	0.66	0.7	0.78	1	-	-	-	-
	Comp	0.95	0.92	0.68	0.91	0.77	1	-	-	-
	Literacy	0.96	0.71	0.72	0.93	0.95	0.88	1	-	-
	Oral	0.91	0.95	0.81	0.73	0.72	0.91	0.77	1	-
	Prod.	0.93	0.68	0.84	0.78	0.96	0.79	0.93	0.81	1
12	**Overall**	1	-	-	-	-	-	-	-	-
	Listening	0.85	1	-	-	-	-	-	-	-
	Speaking	0.8	0.61	1	-	-	-	-	-	-
	Reading	0.91	0.71	0.67	1	-	-	-	-	-
	Writing	0.9	0.65	0.69	0.78	1	-	-	-	-
	Comp	0.95	0.92	0.69	0.92	0.77	1	-	-	-
	Literacy	0.96	0.71	0.72	0.93	0.94	0.89	1	-	-
	Oral	0.92	0.94	0.81	0.76	0.72	0.92	0.78	1	-

Educational authorities may use other data sources to which they have access to further explore the relationship of LAS Links scores with other similar and different measures in their own context. For example, correlations between LAS Links scale scores and state summative assessments in ELA, used here as a measure of learning outcomes, would provide strong evidence of criterion-related validity of LAS Links scores. In other words, if such correlations are strong, this would provide additional evidence that LAS Links scores are appropriately related to learning goals for the students.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Chapter 7: Consequential Validity and Fairness

This chapter discusses the consequential validity of LAS Links test scores when used as intended, including special attention to the relationship between consequences and the fairness of test score uses. In support of consequential validity, this chapter discusses evidence that DIF over student groups is not present, that bias reviews have not detected bias or unfairness in item content, that test score reliabilities by student groups are reasonably constant, and that universal design and accommodations support student access to test content.

Consequential Validity

Messick (1989) begins a discussion of validity with a definition: “Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of the inferences and actions based on test scores or other modes of assessment” (p.13). In the same paragraph, he provides an important elaboration of this definition by writing the following: “Broadly speaking, then, validity is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use.” The use of the word inductive is informative. An inductive argument is one in which the premises are true, but the conclusions may still be false. Reasons are provided that support a possible, or even likely, truth, but that truth is not proven in any absolute sense. The use of this word provides insight into the complex interplay between evidence and theory, and ultimately into how the consequences of test score use might be supported (or not) by both.

Evidence is often thought to be concrete. It can exist without question, but in this framework of validity, the existence of evidence does not constitute a complete validity argument, and the more qualitative aspects of validation of uses are explicitly included. Accordingly, Messick further elaborates what he means by validation by discussing both the interpretive and action inferences that are made, explaining that:

> To validate an interpretive inference is to ascertain the degree to which multiple lines of evidence are consonant with the inference, while establishing that alternative inferences are less well supported. To validate an action inference requires validation not only of score meaning but also of value implications and action outcomes, especially appraisals of the relevance and utility of the test scores for particular applied purposes and of the social consequences of using the scores for applied decision making (p. 13).

These two sentences communicate the essence of Messick’s unified validity concept and its applications to test validation. Validation of the interpretive inference alone is insufficient without validation of the action inferences, and validation of action inferences cannot occur without validation of the interpretive inferences.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

The Standards define validity as "…the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests" (p.11). This definition draws on Messick (1989). Although it does not include explicit reference to the "integrated evaluative judgments" that are central in Messick's characterization of validity, these principles are operationalized to some extent in the inclusion of five sources of validity evidence in the validation process, namely evidence based on 1) test content, 2) response processes, 3) internal structure, 4) relation to other variables, and 5) testing consequences.

This chapter provides evidence in support of the intended action outcomes (aka consequences) of the uses of LAS Links test scores and where consideration of fairness principles is treated as foundationally related to consequential validity.

Fairness

The *Standards* emphasize fairness as a "…fundamental issue in protecting test takers and test users in all aspects of testing," further stating that, "Fairness is a fundamental validity issue…" (p. 49). This description suggests that validity and fairness are matters requiring attention at every point in the testing process, from specification through score use.

To communicate the complex nature of fairness in testing, the *Standards* lay out four views of fairness. They are (a) "Fairness in Treatment During the Testing Process," (b) "Fairness as a Lack of Measurement Bias," (c) "Fairness in Access to the Constructs as Measured," and (d) "Fairness as Validity of Individual Test Interpretations for Their Intended Uses" (pp. 51-53). The first two views focus on standardization and psychometric quality features of testing—factors that attend to the equal treatment of examinees in measuring performance relative to the intended construct.

The second two views on fairness also have equity components, but it is here where the fairness and equity discussions become more complex because they relate to many conditions beyond the test and test scores, including what is happening in the classroom, in homes, and at policy and judicial levels, namely, to the consequences of testing. This makes the notion of fairness far more difficult to characterize in terms of a single value such as equity or even by a competing set of values (for example, equity versus merit). It also implies a limit to what can be demonstrated through test data analysis alone.

For convenience and clarity of presentation amidst this complexity, then, the consequential validity of LAS Links is discussed within the validity and fairness frameworks provided in the Standards. Specifically, evidence is presented that the consequences of LAS Links use are valid and fair with respect to access to test content, measurement invariance, and how these two important sources of validity evidence are linked to testing purpose and LAS Links construct definitions.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Validity and Fairness

Evidence of validity and fairness with respect to testing purpose begins with a thoughtful, explicit, and clear description of the rationale for a test's purpose that articulates the individual and social values explicitly considered and prioritized (i.e., the substantive components of fairness; Camilli, 2013; Kane 2006; Linn, 1984; Messick, 1989). A focus on articulating the values that underly a test's rationale can provide a level of clarity of purpose to support coherent and effective communication of testing purpose. This can be particularly useful toward defending test score uses where there is misunderstanding, or competing values are held by stakeholders.

Testing Purpose Construct Definition

The underlying rationale for LAS Links is based on the value that is placed on supporting students in their acquisition of English language proficiency such that they are able to succeed in English language instructional contexts. Chapter 1 details the intended uses of LAS Links scores in measuring English language proficiency, specifically to

- identify students who require specific English language instruction and support programs,
- plan instructional programs,
- evaluate students' English language proficiency growth,
- determine student readiness to transition out of English language instruction support programs, and
- include student scale scores and proficiency levels in public reporting and accountability systems.

Chapters 2–6 provide evidence of the validity of LAS Links scores relative to these defined uses. This evidence is discussed using the framework provided in the Standards, which calls for evidence in support of the validity of

- Test content: Chapter 2 provides detailed descriptions of the construct, its rationale, development procedures, and adherence to universal design.
- Response processes: Chapter 3 provides detailed descriptions of field testing and scoring processes and procedures, score quality, and the relationship of these elements to understanding the appropriateness of student response processes.
- Internal test structures: Chapter 4 provides detailed analysis results that demonstrate the validity of the internal test structures, including test reliabilities, item-level statistics, measurement error, decision consistency, and factor structures.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

- Relationship with other variables: Chapters 5 and 6 detail the scores provided, their specific interpretations and methods employed to determine those interpretations, and evidence that the relationship of LAS Links domain scores with one another, and with scores on a similar measure, is appropriate.
- Consequences of testing: The remaining sections of this chapter provide evidence from LAS Links accessibility features, results of bias reviews, and from test data analysis that the consequences of LAS Links testing are fair and valid—that scores do not depend on any construct-irrelevant student traits.

Accessibility

LAS Links was developed under the principles of universal design as discussed in detail in Chapter 2. Additionally, universal tools, designated supports, and accommodations are available for students who require further supports to fully access the assessment content. Guidance on accommodations and their use is provided to users in an accommodations supplement for LAS Links assessments (DRC, 2023) and summarized here.

Universal Tools

Universal tools are online tools that are available to all students and are “on for all,” meaning they are available by default. For LAS Links, the following universal tools are available to all students:

- Pointer
- Cross-Off
- Highlighter
- Sticky Notes
- Magnifier
- Line Guide
- Mark for Review/Flag (Reading and Writing)
- Color Choices
- Color Contrast
- Reverse Contrast
- Masking

Designated Supports

Designated supports are locally approved supports for students for whom a need has been identified by the student’s educator or team. Schools and districts may choose to document such supports during student data collection but are not required to do so.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Accommodations

Accommodations provide changes to the format of a test (e.g., large print or Braille), the administration procedure, the student response method (e.g., communication device), or the setting (e.g., small group, special lighting).

Designated school professionals, typically the student's IEP/SSP team or other instructional team, must determine appropriate accommodations on an individual basis. There should be a clearly articulated purpose for using any accommodation based on the student's current level of functioning, learning characteristics, and individual needs. School, district, or state policy should dictate the specific determination and implementation of accommodations.

Print-Based Administration

The following assessment accommodations are available for print-based administration. Students are not limited to this list, and any accommodations not included in these options can be recorded in the student data in the Other category.

- Braille
- Large Print
- Manipulating Test Materials
- Read Aloud
- Scribe
- Sign Language
- Other

Online Administration

The following assessment accommodations are available for online administration. Students are not limited to this list, and any accommodations not included in these options can be recorded as Other.

- Manipulating Test Materials
- Read Aloud
- Scribe
- Sign Language
- Text To Speech (TTS) for Reading and Writing Subtests
- Other

Measurement Invariance

An important source of evidence for fairness and consequential validity is derived from analyses that evaluates measurement properties across different student groups, often referred to as measurement invariance. When scores are interpreted as intended and supported by the validity evidence described throughout this report, it is important to

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

provide evidence of measurement invariance that demonstrates that those interpretations are well-supported across student groups. To that end, DIF is used to evaluate content at the item level, and reliabilities by student groups provide information about invariance at the test level.

Differential Item Functioning

DIF analyses were performed on all field-tested items on gender. It is important to note that DIF analyses using racial and ethnic groups is particularly challenging to use in an English Language proficiency assessment because it requires setting reference and focal groups within some context that that one group has an advantage relative to another group (Ferne & Rupp, 2007). However, for an English Language assessment, technically all participants, regardless of race or ethnicity, are at a disadvantage. Furthermore, the ability to create matched groups across race or ethnic groups for DIF analyses requires collection of other variables in addition to achievement to ensure that the DIF statistics truly convey information about potential item bias. For example, to create matched groups to support DIF analysis it would be important to know the language spoken at home. Insufficient detail was available about the examinees who participated in the standardization study to perform meaningful DIF analysis with respect to race or ethnic groups.

The DIF studies for Las Links included a systematic item analysis to determine whether examinees with the same underlying level of ability had the same probability of getting the item correct. The Mantel-Haenszel (M-H) procedure (Mantel & Haenszel, 1959) was applied in the DIF analyses. The M-H procedure has been widely used in DIF studies. In this procedure, the focal and reference groups are matched on ability using a test score interval as a proxy.

Based on the DIF statistics, an item can be classified into one of three categories: A, B, or C. These categories stand for negligible, intermediate, and large DIF, respectively. The classification rules that were used in the evaluation are listed below. These rules align with those used in the National Assessment of Educational Progress (NAEP) to determine DIF (U.S. Department of Education, Office of Educational Research and Improvement, & National Center for Education Statistics, 2001). Delta statistics for MC items were also considered with the criteria of |Delta|<1 applied for Category A.

- Category A. Either Mantel's chi-square is not significantly different from zero ($p \geq 0.05$), or if the absolute value of the effect size is less than or equal to 0.17.
- Category B. Mantel's chi-square is significant, and the absolute value of the effect size is over 0.17 and less than or equal to 0.25.
- Category C. Mantel's chi-square is significant, and the absolute value of the effect size is over 0.25.

Typically, items flagged with Category B or Category C are examined to determine whether item performance differences between identifiable subgroups of the population

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

were due to extraneous or construct-irrelevant information, making the items unfairly difficult. The Data Review section of Chapter 2 discusses DIF analysis results and item reviews in detail. Very few items were flagged for DIF during item analyses, and item selections for the final LAS Links 3rd Edition forms minimized inclusion of such items, even where no source of bias was identified during the data review process.

Reliability

Table 57 provides reliability and SEM for each domain and grade span tested by student groups based on gender, ethnicity, and accommodations. The reader may refer to Chapter 4 for a detailed discussion of analysis procedures and for the overall test reliability and SEM information.

Reliabilities by student groups based on a binary gender indication, ethnicity, and student accommodations show that reliabilities for students who use accommodations tend to be lower than for students who do not, however, reliabilities for male and female and Hispanic and non-Hispanic students are very similar.

Table 66. Reliability by Student Group, Domain, and Grade Span

Student Group	Domain	Grade	N	Alpha	SEM
Hispanic	**Listening**	K	4434	0.74	1.99
		1	4251	0.81	1.71
		2-3	8798	0.78	1.82
		4-5	7757	0.78	1.96
		6-8	8803	0.79	2.13
		9-12	8538	0.74	2.20
	Reading	K	4431	0.70	2.35
		1	4249	0.87	2.26
		2-3	8790	0.86	2.44
		4-5	7753	0.85	2.42
		6-8	8796	0.83	2.49
		9-12	8537	0.88	2.35
	Speaking	K	4426	0.85	1.93
		1	4246	0.88	2.24
		2-3	8788	0.91	2.06
		4-5	7751	0.89	2.15
		6-8	8784	0.91	2.21
		9-12	8494	0.92	2.21
	Writing	K	4429	0.68	1.96
		1	4246	0.88	2.52
		2-3	8785	0.89	2.75
		4-5	7756	0.90	2.42
		6-8	8789	0.88	2.31
		9-12	8520	0.88	2.45

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Student Group	Domain	Grade	N	Alpha	SEM
Not Hispanic	Listening	K	2006	0.77	1.92
		1	1925	0.79	1.58
		2-3	3730	0.79	1.72
		4-5	2896	0.77	1.91
		6-8	2261	0.79	2.02
		9-12	2080	0.74	2.18
	Reading	K	2004	0.81	2.26
		1	1925	0.89	2.00
		2-3	3728	0.89	2.29
		4-5	2900	0.86	2.41
		6-8	2258	0.85	2.46
		9-12	2078	0.88	2.38
	Speaking	K	1998	0.81	2.01
		1	1921	0.86	2.03
		2-3	3725	0.89	1.92
		4-5	2894	0.87	2.12
		6-8	2256	0.90	2.14
		9-12	2068	0.89	2.26
	Writing	K	1998	0.74	1.95
		1	1924	0.87	2.67
		2-3	3723	0.89	2.67
		4-5	2898	0.88	2.49
		6-8	2255	0.87	2.37
		9-12	2074	0.87	2.51
Male	Listening	K	3440	0.75	2.00
		1	3294	0.81	1.71
		2-3	6633	0.79	1.83
		4-5	5854	0.79	1.96
		6-8	6153	0.79	2.12
		9-12	6055	0.74	2.22
	Reading	K	3438	0.76	2.33
		1	3293	0.89	2.18
		2-3	6626	0.87	2.45
		4-5	5855	0.85	2.47
		6-8	6149	0.84	2.46
		9-12	6050	0.88	2.40
	Speaking	K	3433	0.85	1.92
		1	3290	0.88	2.19
		2-3	6620	0.90	2.11
		4-5	5849	0.89	2.13
		6-8	6143	0.91	2.19
		9-12	6020	0.92	2.20

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Student Group	Domain	Grade	N	Alpha	SEM
	Writing	K	3431	0.73	1.94
		1	3292	0.89	2.53
		2-3	6616	0.89	2.80
		4-5	5850	0.90	2.40
		6-8	6142	0.88	2.32
		9-12	6041	0.88	2.45
Female	Listening	K	3090	0.75	1.95
		1	2941	0.80	1.64
		2-3	5980	0.78	1.79
		4-5	4877	0.76	1.98
		6-8	5001	0.80	2.07
		9-12	4649	0.74	2.21
	Reading	K	3087	0.75	2.31
		1	2940	0.88	2.17
		2-3	5977	0.88	2.38
		4-5	4876	0.85	2.47
		6-8	4995	0.84	2.45
		9-12	4651	0.87	2.41
	Speaking	K	3081	0.84	1.98
		1	2936	0.88	2.16
		2-3	5978	0.91	1.99
		4-5	4874	0.89	2.13
		6-8	4987	0.91	2.18
		9-12	4628	0.91	2.27
	Writing	K	3086	0.71	1.94
		1	2937	0.88	2.65
		2-3	5977	0.89	2.77
		4-5	4882	0.89	2.53
		6-8	4992	0.87	2.38
		9-12	4639	0.87	2.55
Accommodations	Listening	K	64	0.66	2.08
		1	94	0.77	1.89
		2-3	453	0.68	1.99
		4-5	739	0.68	2.07
		6-8	990	0.72	2.16
		9-12	365	0.62	2.22
	Reading	K	64	0.76	2.35
		1	93	0.86	2.39
		2-3	453	0.78	2.56
		4-5	743	0.77	2.48
		6-8	987	0.80	2.47
		9-12	361	0.84	2.44

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Student Group	Domain	Grade	N	Alpha	SEM
	Speaking	K	63	0.82	1.88
		1	93	0.88	2.16
		2-3	451	0.88	2.09
		4-5	739	0.85	2.04
		6-8	983	0.86	2.18
		9-12	362	0.87	2.26
	Writing	K	63	0.74	1.78
		1	94	0.89	2.52
		2-3	452	0.87	2.75
		4-5	742	0.89	2.31
		6-8	985	0.86	2.24
		9-12	364	0.83	2.48
No Accommodations	**Listening**	K	6467	0.75	1.98
		1	6142	0.81	1.66
		2-3	12161	0.79	1.79
		4-5	9992	0.78	1.95
		6-8	10165	0.80	2.08
		9-12	10342	0.75	2.18
	Reading	K	6462	0.76	2.30
		1	6141	0.88	2.23
		2-3	12151	0.88	2.38
		4-5	9988	0.86	2.40
		6-8	10158	0.84	2.47
		9-12	10343	0.88	2.38
	Speaking	K	6452	0.84	1.99
		1	6134	0.88	2.18
		2-3	12148	0.91	2.00
		4-5	9984	0.89	2.15
		6-8	10148	0.91	2.22
		9-12	10289	0.91	2.33
	Writing	K	6455	0.72	1.95
		1	6136	0.88	2.65
		2-3	12142	0.89	2.78
		4-5	9990	0.89	2.51
		6-8	10150	0.88	2.32
		9-12	10319	0.88	2.48

Summary and Recommendations

The preceding describes several important sources of consequential validity evidence that are derived from test development and data analysis processes and procedures. Collectively this evidence offers assurances that the intended uses for LAS Links 3rd Edition test scores are appropriate in that students are afforded comprehensive access to the test content through universal design and test accommodations and that the risk

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

of content bias and construct irrelevant variance is mitigated in substantive ways as demonstrated by the results of DIF analysis and score consistency and accuracy analysis by student groups.

It is important to note, however, that the consequences of test score use extend well beyond the procedures used to develop LAS Links and the student group level analysis that are possible given the test data alone. The Standards acknowledge that

> A full consideration of the topic would explore the multiple functions of testing in relation to its many goals, including the broad goal of achieving equality of opportunity in our society. It would consider the technical properties of tests, the ways in which test results are reported and used, the factors that affect the validity of score interpretations, and the consequences of test use. A comprehensive analysis of fairness in testing also would examine the regulations, statutes, and case law that govern test use and the remedies for harmful testing practices. The Standards cannot hope to deal adequately with all of these broad issues, some of which have occasioned sharp disagreement among testing specialists and others interested in testing (p. 49).

Therefore, LAS Links 3rd Edition test score users are encouraged to routinely examine the context in which they are using scores and the consequences for stakeholders. Where scores are used to make judgments about students, care should be taken to use multiple measures to the extent practical. Educational authorities should also monitor test score use for any unintended consequences or misuse of test scores that may arise. Finally, although disparate outcomes over student groups are not considered to be inherently unfair (AERA et al, 2014), consistent disparate impact is cause for heightened scrutiny to ensure that both interpretation and use of LAS Links scores are consonant with the intended uses described herein and for which this body of evidence is offered.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

References

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 Technical Report*. National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/pdf/main1998/2001509.pdf

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

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on academic English: Implications for K–12 English language learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.

Asparouhov, T. & Muthén, B. (2010). Simple second order chi-square correction (Technical Note). Retrieved from: http://www.statmodel.com/download/WLSMV_new_chi21.pdf.

Bailey, A.L., & Huang, B. H. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing, 28*(3), 343–365. https://doi.org/10.1177/0265532211404187

Beck, I.L., McKeown, M.G., Sinatra, G.M., & Loxterman, J.A. (1991). Revising social studies text from a text-processing perspectiva: Evidence of improved comprensibility. *Reading Research Quarterly, 26*l(3), 251–275. https://doi.org/10.2307/747763

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51.

Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice, 17*(1), 5–9. https://doi.org/10.1111/j.1745-3992.1998.tb00615.x

Brennan, R. (2004). BB-CLASS (Version 1.0). [Computer software]. Center for Advanced Studies in Measurement & Assessment. https://education.uiowa.edu/casma/computer-programs

Brown, G. (1995). Dimensions of difficulty in listening comprehension. In D. Mendelshon & J. Rubin, (Eds.), *A guide for the teaching of second language listening* (pp. 59–73). Dominie Press.

Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation, 19* (2-3), 104-120. https://doi.org/10.1080/13803611.2013.767602

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Council of Chief State School Officers. (2012). *Framework for English Language Proficiency Development standards corresponding to the Common Core State Standards and the Next Generation Science Standards*.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth Group/Thompson Learning.

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

Cronbach, L.J., & Shavelson R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391–418. https://doi.org/10.1177/0013164404266386

CTB/McGraw-Hill LLC. (2006). *LAS Links First Edition, Forms A and B Technical Manual* (Report).

CTB/McGraw-Hill LLC. (2013). *LAS Links Second Edition, Forms C and D Technical Manual* (Report).

Data Recognition Corporation (2023). *Accommodations Supplement for LAS Links Assessments*.

Data Recognition Corporation (2023). *Test Administration Manual for LAS Links Assessments*.

Egan, K.L., Schneider, M.C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 79–106). Routledge.

Ferne, T & Rupp, A. A. (2007) A Synthesis of 15 Years of Research on DIF in Language Testing: Methodological Advances, Challenges, and Recommendations, Language Assessment Quarterly, 4:2, 113-148, DOI: 10.1080/15434300701375923

Ferrara, S., Lewis, D., & D'Brot, J. (2021). Setting benchmarked performance standards: A content focused, judgmental approach, procedures, and some empirical results. *Journal of Applied Testing Technology, 22*(1), 52–73.

Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice, 24*(3), 21–28. https://doi.org/10.1111/j.1745-3992.2005.00016.x

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Gee, J. P. (2008). What is academic language? In A. S. Rosebery & B. Warren (Eds.), *Teaching science to English language learners: Building on students' strengths* (pp. 57–70). National Science Teachers Association.

Gibbons, P. (1998). Classroom talk and the learning of new registers in a second language. *Language and Education, 12*(2), 99–118. https://doi.org/10.1080/09500789808666742

Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in a content-based classroom. *TESOL Quarterly, 37*(2), 247–273. https://doi.org/10.2307/3588504

Gulliksen, H. (1950). *Theory of mental tests*. John Wiley and Sons. https://doi.org/10.1037/13240-000

Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89–116). Lawrence Erlbaum.

Hambleton, R.K, & Novick, M.R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*l(3), 159–170.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practices, 10*(2), 33–41. https://doi.org/10.1111/j.1745-3992.1991.tb00195.x

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 60*(1), pp. 1—55. https://doi.org/10.1080/10705519909540118

Huynh, H. (1976). On the reliability of decisions in domain referenced testing. Journal of *Educational Measurement, 13*l(4), 253–264. https://doi.org/10.1111/j.1745-3984.1976.tb00016.x

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp.17—64). American Council on Education.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practices*. Springer.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring* [Symposium]. Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ, United States.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225–253). Routledge.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Linn, R. L. (1984). Selection bias: Multiple meanings. *Journal of Educational Measurement, 21*(1), 33-47. https://doi.org/10.1111/j.1745-3984.1984.tb00219.x

Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement 32*l(2), 179–197. https://doi.org/10.1111/j.1745-3984.1995.tb00462.x

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

MacCallum, R.C. (2003). 2001 presidential address: Working with imperfect models. *Multivariate Behavioral Research, 38*(1), pp. 113-139. https://doi.org/10.1207/S15327906MBR3801_5

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719–748. https://doi.org/10.3102/0013189X12470855

McClarty, K.L., Way, W.D., Porter, A.C., Beimers, J.N., & Miles, J.A. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher 42*(2), 78–88. https://doi.org/10.3102/0013189X12470855

Messick, S. (1989). Validity. In R. L. Brennan (Ed.), *Educational Measurement* (3rd ed., pp.3–104). American Council on Education.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176. https://doi.org/10.1177/014662169201600206

National Center for Education Statistics (2024). *NCES Statistical Standards*. https://nces.ed.gov/statprog/2002/std5_1.asp.

National Governors Association Center for Best Practices [NGA Center], & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. https://www.thecorestandards.org/wp-content/uploads/ELA_Standards1.pdf

Phillips, G.W. (2012). The benchmark method of standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 232–346). Routledge.

Satorra A. & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Sage.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Scarcella, R. (2003). Academic English: A conceptual framework (Report No. 2003-1). Irvine, CA: The University of California Linguistic Minority Research Institute.

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistic perspective*. Lawrence Erlbaum.

Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*(2), 201–210.

Teachers of English to Speakers of Other Languages. (2006). *PreK–12 English language proficiency standards*.

Thompson, S.J., Johnstone, C.J. & Thurlow, M.L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/Synth44.pdf

van Lier, L., & Walqui, A. (2012, January 13—14). *How teachers and educators can most usefully and deliberately consider language* [Paper presentation]. Understanding Language Conference, Stanford, CA, United States.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Appendix A Scale Score Descriptive Statistics

Table A.1. Forms E Speaking Scale Score Descriptive Statistics

Grades	Mean	SD	Median
K	451	46.43	463
1	469	41.63	476
2-3	488	30.03	493
4-5	493	36.54	499
6-8	496	41.31	504
9-12	483	40.59	493

Table A.2. Forms E Listening Scale Score Descriptive Statistics

Grades	Mean	SD	Median
K	428	41.43	433
1	450	40.17	455
2-3	466	39.99	468
4-5	494	55.08	503
6-8	497	55.09	500
9-12	502	59.00	509

Table A.3. Forms E Reading Scale Score Descriptive Statistics

Grades	Mean	SD	Median
K	349	57.54	363
1	393	57.97	395
2-3	450	65.30	458
4-5	497	57.94	504
6-8	509	57.25	517
9-12	513	54.57	518

Table A.4. Forms E Writing Scale Score Descriptive Statistics

Grades	Mean	SD	Median
K	315	70.94	316
1	369	77.76	374
2-3	449	70.05	464
4-5	487	73.88	503
6-8	501	67.68	513
9-12	475	73.05	486

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Appendix B Proficiency Level Definitions and Proficiency Level Descriptors

Table B.1. LAS Links Proficiency Level Definitions

Proficiency Level	Proficiency Description
5 Above Proficient	Level 5 students communicate effectively in English, with few if any errors, across a wide range of grade-level-appropriate language demands in social, school, and academic contexts. The students command a high degree of productive and receptive control of lexical, syntactic, phonological, and discourse features when addressing new or familiar topics. Level 5 students apply their language mastery to critically evaluate and synthesize written and oral information and to formulate hypotheses. Their facility with language allows them to analyze information, draw sophisticated inferences, and explain their reasoning. They skillfully organize information for presentations and can express subtle nuances of meaning. They apply literary techniques such as identifying author tone and point of view and can tailor language to a particular purpose and audience.
4 Proficient	Level 4 students communicate effectively in English, but with some errors, across a range of grade-level-appropriate language demands in social, school, and academic contexts. The students exhibit productive and receptive control of lexical, syntactic, phonological, and discourse features when addressing new or familiar topics. Level 4 students interpret, analyze, and evaluate written and oral information, basing their responses on implicit and explicit context clues and information from personal and academic experiences. They adequately express themselves and organize their responses in logical and sequenced order. They distinguish nuances of meaning and incorporate idiomatic expressions and academic vocabulary.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	Proficiency Description
3 Intermediate	Level 3 students communicate in English across a range of grade-level-appropriate language demands in social, school, and academic contexts. However, errors interfere with their communication and comprehension. Repetition and clarification are often needed. The students exhibit a limited range of productive and receptive control of lexical, syntactic, phonological, and discourse features when addressing new or familiar topics. Level 3 students use limited vocabulary when defining concepts across and within academic disciplines. They can compare, contrast, summarize, and relate text to graphic organizers. They decode words, apply grammar conventions, and use context clues to identify word meanings. They identify proper and improper use of basic grammar. Although their language is generally coherent, it lacks significant elaboration or detail.
2 Early Intermediate	Level 2 students are developing the ability to communicate in English in social, school, and academic contexts. Errors frequently impede basic communication and comprehension. Their receptive and productive control of lexical, syntactic, phonological, and discourse features of English is emerging. Early Intermediate students have minimal vocabulary and grammar skills. They identify, describe, and discuss simple pictorial or text prompts. Students interpret language related to familiar social, school, and academic topics. They draw simple inferences and make simple comparisons. They restate rather than create original expressions. Restricted vocabulary and rudimentary grammar limit their expression and comprehension.
1 Beginning	Level 1 students are starting to develop receptive and productive uses of English in social, school, and academic contexts. Their comprehension may be demonstrated nonverbally or through their native language rather than in English.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table B.2. Proficiency Level Descriptors, Kindergarten

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Kindergarten	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically use basic vocabulary and simple phrases to name or describe common objects and express opinions or preferences in social and academic situations. They narrate a story related to a sequence of pictures about school-related activities using basic vocabulary. Restricted vocabulary and developing grammar limit expression. Errors frequently impede communication.	Early Intermediate students typically follow some simple oral directions using knowledge of everyday tasks and basic academic vocabulary. They identify common shapes, letters, numbers, and familiar locations. They identify details in simple oral stories. Their restricted vocabulary and developing grammar limit comprehension. Errors frequently impede communication and comprehension.	Early Intermediate students typically identify capital and lowercase letters in isolation and identify beginning, middle, and ending sounds. They identify main ideas and details in simple text, match text to pictures, and apply letter-sound relationships. Their restricted vocabulary and developing grammar limit comprehension. Errors frequently impede comprehension.	Early Intermediate students copy simple words and sentences that describe pictures or respond to other prompts. Errors frequently impede communication.
3 Intermediate	Intermediate students typically use appropriate words and phrases when conducting transactions, making requests, and asking for clarification in social and academic settings. They narrate a story related to a sequence of pictures about school-related activities using mostly accurate, although limited, vocabulary. They provide mostly clear information although errors interfere with communication.	Intermediate students typically follow simple oral directions and identify locations. They identify main ideas and make some inferences in simple oral stories. Errors interfere with communication and comprehension.	Intermediate students typically decode words with short vowel sounds, match text to pictures, and recall details and main ideas in short passages. Students make simple inferences and recognize words that relate to spatial relationships. Errors interfere with comprehension.	Intermediate students typically write one or more words to describe a picture or respond to other prompts. Students are beginning to recognize correct sentence format. Errors interfere with communication.
4 Proficient	Proficient students typically produce simple and accurate sentences when making requests and asking for clarifications. They use appropriate words and phrases to label and describe the purpose of less common objects. They narrate a story related to a sequence of pictures about school-related activities using accurate vocabulary. Minor errors do not interfere with communication.	Proficient students typically follow oral directions to distinguish the location of an object in relation to another object, recall details in an oral story, and make inferences. They identify main ideas in more complex stories.	Proficient students typically identify rhyming words, match words to definitions or descriptions, make inferences, recall events from short passages, and read simple sentences independently. Errors do not interfere with comprehension.	Proficient students typically use correct basic grammar, capitalize the beginning of a sentence, and use correct ending punctuation in declarative, interrogative, and imperative sentences. They identify standard sentence structure and generate descriptive and explanatory sentences. Errors do not interfere with communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Kindergarten	Speaking	Listening	Reading	Writing
5 Above Proficient	Above Proficient students typically produce simple sentences and use correct grammar when making requests, asking for clarification, and describing situations. They narrate a story with extensive and accurate vocabulary and grammar appropriate to their age.	Above Proficient students typically recall details and sequence of events, and determine main ideas in oral stories that have advanced vocabulary.	Above Proficient students typically use context clues to determine meanings of words and recall subtle details. They identify sequence in short passages and recognize words that relate to spatial relationships.	Above Proficient students typically write a complete sentence to describe a picture or respond to other prompts. They form regular plural nouns and possessive pronouns, and choose correct sentence-ending punctuation. Communication is clear and complete, although content may contain minor errors.

Table B.3. Proficiency Level Descriptors, Grade 1

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Grade 1	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically use basic vocabulary and simple phrases to name or describe common objects and express opinions or preferences in social and academic situations. They narrate a story related to a sequence of pictures about school-related activities using basic vocabulary. Restricted vocabulary and developing grammar limit expression. Errors frequently impede communication.	Early Intermediate students typically follow some simple oral directions using knowledge of everyday tasks and basic academic vocabulary. They identify common shapes, letters, numbers, and familiar locations. They identify details in simple oral stories. Their restricted vocabulary and developing grammar limit comprehension. Errors frequently impede communication and comprehension.	Early Intermediate students typically identify capital and lowercase letters in isolation, identify beginning, middle, and ending sounds, and recall main ideas and important details in simple text. They apply letter-sound relationships. Their restricted vocabulary and developing grammar limit comprehension. Errors frequently impede comprehension.	Early Intermediate students typically copy simple sentences and write one or more words to describe or explain a picture. They select grammatically correct sentences from a set of choices. Their restricted vocabulary and developing grammar limit expression. Errors frequently impede communication.
3 Intermediate	Intermediate students typically use appropriate words and phrases when conducting transactions, making requests, and asking for clarification in social and academic settings. They narrate a story related to a sequence of pictures about school-related activities using mostly accurate, although limited, vocabulary. They provide mostly clear information although errors interfere with communication.	Intermediate students typically follow simple oral directions and identify locations. They identify main ideas and draw simple inferences in simple oral stories. Errors interfere with communication and comprehension.	Intermediate students typically decode basic words and match text to pictures. Students make simple inferences and recognize words related to spatial relationships. Errors interfere with comprehension.	Intermediate students typically write words, phrases, or sentences that attempt to describe or explain a picture. They are beginning to recognize sentences illustrating correct grammar, proper subject/verb agreement, and correct pluralization and capitalization. They have limited range of vocabulary knowledge. Errors interfere with communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grade 1	Speaking	Listening	Reading	Writing
4 Proficient	Proficient students typically produce simple and accurate sentences when making requests and asking for clarifications. They use appropriate words and phrases to label and describe the purpose of less common objects. They narrate a story related to a sequence of pictures about school-related activities using accurate vocabulary. Minor errors do not interfere with communication.	Proficient students typically follow oral directions to distinguish the location of an object in relation to another object, recall details in an oral story, and draw inferences. They identify main ideas in more complex stories.	Proficient students typically identify rhyming words, match basic text to pictures, make inferences, recall details and main ideas in short passages, and read simple sentences independently. Errors do not interfere with comprehension.	Proficient students typically use correct basic grammar, capitalize the beginning of a sentence, and use correct ending punctuation in declarative, interrogative, and imperative sentences. They identify standard sentence structure and generate descriptive and explanatory sentences. Errors do not interfere with communication.
5 Above Proficient	Above Proficient students typically produce simple sentences and use correct grammar when making requests and conducting transactions in the classroom or describing familiar social situations or a process. They narrate a story with extensive and accurate vocabulary and grammar appropriate to their age.	Above Proficient students typically recall details and the sequence of events, and determine main ideas in oral stories that have advanced vocabulary.	Above Proficient students use context clues to determine meanings of words, recall subtle details, and determine sequence in short passages. They use interpretation and inference to comprehend a story. Students recognize words that relate to spatial relationships.	Above Proficient students typically write a complete sentence to describe a picture or respond to other prompts. They form regular plural nouns and possessive pronouns, and choose correct sentence-ending punctuation. Communication is clear and complete, although content may contain minor errors.

Table B.4. Proficiency Level Descriptors, Grades 2-3

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Grades 2-3	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically use basic vocabulary and grammar, and simple phrases or sentences to make requests or comparisons, ask questions, express opinions or preferences, or describe a sequence of pictures about familiar events and situations. Errors frequently impede communication.	Early Intermediate students typically follow simple oral directions and identify high-frequency vocabulary. They identify a few details and draw simple inferences in oral stories. Errors frequently impede communication and comprehension.	Early Intermediate students typically understand word meanings and synonyms, possess basic knowledge of morphemes and syllables, identify one-syllable words, recognize simple rhyming words, and make simple inferences. Errors frequently impede comprehension.	Early Intermediate students typically describe, explain, or express ideas in sentences. They make simple comparisons. Students demonstrate basic vocabulary knowledge and grammar skills such as use of auxiliary verbs, verb tenses, and conjunctions. Errors frequently impede communication.
3 Intermediate	Intermediate students typically use appropriate words and phrases when expressing a preference, asking questions, providing information and explanations, naming common objects, and describing common functions. They produce mostly accurate sentences when narrating simple stories about familiar events and situations. Errors interfere with communication.	Intermediate students typically understand a limited range of vocabulary. They recall details, identify main ideas, and draw inferences in more complex oral stories. Errors interfere with communication and comprehension.	Intermediate students typically match words to definitions or descriptions, interpret words and basic phrases, and apply knowledge of morphemes and syllables. They recall stated details and main ideas, make inferences, and determine characters' feelings. Errors interfere with comprehension.	Intermediate students typically respond to various prompts or pictures using multiple sentences. Students make simple predictions and express some opinions in response to pictures. Meaning is somewhat clear although vocabulary may be limited. They identify appropriate verb forms and articles based on contextual clues. Errors interfere with communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grades 2-3	Speaking	Listening	Reading	Writing
4 Proficient	Proficient students typically produce complete sentences with few grammatical and vocabulary errors when describing situations, explaining their reasoning, or narrating a story. They use broad vocabulary to accurately express opinions or preferences and ask appropriate questions. Minor errors do not interfere with communication.	Proficient students typically understand academic vocabulary and follow some complex directions. They recall subtle details, determine main ideas, and identify speaker purpose.	Proficient students typically identify synonyms of social and academic vocabulary and interpret words and phrases. They use context clues to determine meaning, recall implicit details and main ideas, draw complex inferences, identify literary features, and transfer concepts to new situations. Errors do not interfere with comprehension.	Proficient students typically make predictions and express opinions in response to pictures using complete sentences. They use correct auxiliary verb forms and verb tenses and correctly use writing conventions such as capitalization and punctuation. They organize and write responses in logical and sequential order. Errors do not interfere with communication.
5 Above Proficient	Above Proficient students typically produce sentences with sophisticated vocabulary and correct grammar when providing information, describing situations, or explaining their reasoning.	Above Proficient students typically recall details and sequence of events, and determine main ideas in oral stories that have advanced vocabulary.	Above Proficient students typically identify two-syllable words and rhyming words written with digraphs, use common multiple-meaning words, and recognize synonyms. They determine story sequence and details of fictional and academic texts, make generalizations, and use self-monitoring techniques to check for understanding.	Above Proficient students typically write fluently to a variety of pictures, prompts, or purposes with precise vocabulary and ease of expression. They use correct verb tenses and subject/verb agreement, appropriate articles and punctuation. Responses contain few digressions or repetitions. Communication is clear and complete, though it may contain minor errors.

Table B.5. Proficiency Level Descriptors, Grades 4-5

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Grades 4-5	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically use basic vocabulary and grammar and simple sentences to identify common objects and describe their function, provide basic information, make requests, ask questions, and express opinions or preferences. They construct a narrative from a sequence of pictures about familiar events and school-related activities and compare and contrast information found in texts and graphic organizers using basic vocabulary. Errors frequently impede communication.	Early Intermediate students typically follow some simple oral directions and understand common vocabulary and idiomatic expressions. They identify details. Errors frequently impede communication and comprehension.	Early Intermediate students typically interpret basic words and phrases and identify some main ideas and details in simple text. Errors frequently impede comprehension.	Early Intermediate students typically write sentences using basic vocabulary and grammar to describe and discuss text, interpret graphic organizers, and compare and contrast information. Errors in organization, grammar, word choice, and mechanics frequently impede communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grades 4-5	Speaking	Listening	Reading	Writing
3 Intermediate	Intermediate students typically use appropriate words and phrases and complete sentences when making requests, expressing opinions or preferences, providing information, and describing locations. They construct a narrative from a sequence of pictures and compare and contrast information found in texts and graphic organizers using mostly accurate, although limited, vocabulary. Errors interfere with communication.	Intermediate students typically follow oral directions and interpret both basic vocabulary and idiomatic expressions. They identify some main ideas and make simple inferences from passages and understand details within graphic organizers. Errors interfere with communication and comprehension.	Intermediate students typically use knowledge of high-frequency affixes to determine word meanings. They recall main ideas and stated details in text, and interpret simple words and phrases. Errors interfere with comprehension.	Intermediate students typically respond appropriately to various verbal prompts or graphic organizers by using complete sentences that exhibit correct basic grammar. Meaning is somewhat clear, although vocabulary may be limited. They demonstrate a grasp of pronouns, prepositions, auxiliary verbs and verb tenses. Errors in organization, grammar, word choice, and mechanics interfere with communication.
4 Proficient	Proficient students typically produce complete sentences when providing information, asking questions, explaining a process, expressing an opinion, and narrating a story. They organize responses in logical and sequential order. They accurately identify and compare and contrast features of less common objects. Minor errors do not interfere with communication.	Proficient students typically follow multistep directions using academic vocabulary, recall details, identify main ideas, and determine sequence of steps in classroom discussions and lessons. They draw inferences from more complex oral stories and interpret tables and other graphic organizers.	Proficient students typically use knowledge of more advanced affixes to determine word meanings. They identify synonyms, use context clues to determine word meanings, and interpret slightly complex words and phrases. They read for specific information in graphic organizers, infer information, and draw conclusions. Errors do not interfere with comprehension.	Proficient students typically write complete sentences with mostly accurate vocabulary and grammar that demonstrates appropriate use of punctuation, prepositional phrases, and other conventions. They summarize passages; interpret, compare, and contrast information from graphic organizers and from implicit and explicit context clues; and organize and write responses to open-ended questions in logical and sequential order. Errors do not interfere with communication.
5 Above Proficient	Above Proficient students typically produce sentences with sophisticated vocabulary and correct grammar when providing information, describing situations, asking questions, expressing opinions and subtle nuances of meanings, and explaining processes and their reasoning. They create a detailed and structured narrative.	Above Proficient students typically follow directions that use verb phrases and determine key information to summarize a task. They recall subtle details, identify main ideas and speaker purpose, and draw sophisticated inferences from classroom discussions and lessons.	Above Proficient students typically identify synonyms and antonyms of less familiar words and interpret complex words and phrases. They use prediction, determine story sequence, and use self-monitoring techniques to check for understanding.	Above Proficient students typically write fluently in response to a variety of prompts and purposes. They skillfully organize, interpret, summarize, and evaluate information from texts and graphic organizers. Communication is clear and complete, though it may contain minor errors.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table B.6. Proficiency Level Descriptors, Grades 6-8

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Grades 6-8	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically produce simple sentences using basic vocabulary and grammar when describing social situations, giving instructions, and identifying locations. They construct a narrative from a sequence of pictures about familiar events and school-related activities and compare and contrast information found in texts and graphic organizers. Errors frequently impede communication.	Early Intermediate students typically follow simple oral directions and understand common vocabulary and idiomatic expressions. They identify details. Errors frequently impede communication and comprehension.	Early Intermediate students typically follow simple oral directions and understand common vocabulary and idiomatic expressions. They identify some details. Errors frequently impede comprehension.	Early Intermediate students typically write complete sentences using basic vocabulary and grammar to describe, explain, or compare verbal or graphic prompts. They respond to simple open-ended questions and summarize simple passages. Errors in organization, grammar, word choice, and mechanics frequently impede communication.
3 Intermediate	Intermediate students typically use appropriate words and phrases and complete sentences when expressing opinions, providing information, conducting transactions, or describing common functions. They describe common social situations and narrate simple stories. Grammatical or vocabulary errors interfere with communication, but the intended meaning is somewhat clear.	Intermediate students typically follow multistep directions that use academic vocabulary. They recall details from class discussions or short oral stories and identify the main purpose of conversation. They interpret graphic organizers and extrapolate conclusions from discussions. Errors interfere with communication and comprehension.	Intermediate students identify synonyms of familiar social and academic vocabulary and interpret common idioms using context clues. They distinguish main ideas from supporting details and draw inferences from clues in text. Errors interfere with comprehension.	Intermediate students typically write complete sentences to describe, explain, or compare or contrast verbal or graphic prompts. They write responses to open-ended questions and summarize passages. They use sentence-ending punctuation, pronouns, prepositional phrases, auxiliary verbs, and verb tenses. Responses have limited range of vocabulary. Errors in organization, grammar, word choice, and mechanics interfere with communication.
4 Proficient	Proficient students typically produce complete sentences to express opinions, provide information, conduct transactions, make a request, explain processes, give instructions, and describe social situations. They produce generally fluent narratives with some hesitations or self-corrections that do not obscure meaning. They organize responses in logical and sequential order and incorporate idiomatic expressions. Speech is coherent and clear but lacks elaboration or detail.	Proficient students typically follow complex multistep directions. They determine main ideas, infer directions, draw simple conclusions and predict logical outcomes in oral stories. They understand metaphorical language and uncommon idiomatic expressions, and recognize technical academic vocabulary.	Proficient students interpret idioms and determine synonyms of grade-level words. They recall stated and implicit details in a variety of genres, identify specific information in graphic organizers, and determine main ideas in fiction and academic texts. They analyze the structure of texts and identify literary techniques. Errors do not interfere with comprehension.	Proficient students typically write logically-sequenced responses that incorporate idiomatic expressions and convey original thought in response to open-ended prompts. They accurately interpret pictures or graphical information. They use correct verb tense and agreement, subordinating conjunctions, capitalization, punctuation, and adjective and adverb placement. Errors do not interfere with communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grades 6-8	Speaking	Listening	Reading	Writing
5 Above Proficient	Above Proficient students typically produce sentences with sophisticated vocabulary and correct grammar and subtle nuances of meaning, when expressing opinions, providing information, making requests, identifying and describing objects, and explaining processes and their reasoning. They produce detailed narratives of complex structure and skillfully organize information for presentations.	Above Proficient students typically follow complex instructions, recall subtle details, determine and evaluate key information to summarize a task, and make sophisticated inferences and predictions from classroom discussions or lengthy oral stories. They understand increasingly abstract idiomatic expressions, locate new information in a wider context, and distinguish relevant from extraneous information.	Above Proficient students typically identify synonyms and antonyms, interpret less familiar idioms, apply word definitions, and restate meanings in variant language. They prioritize main and supporting details, and read closely to make logical inferences. They use prediction to read fluently and to identify author's purpose and literary techniques.	Above Proficient students typically craft original responses to prompts, fluently conveying sequenced logical exposition. Students respond to open-ended questions requiring them to extrapolate from information indicated in prompts, interpret and synthesize complex information from graphic organizers, draw sophisticated inferences, explain reasoning, and express and support opinions. Minor errors are possible, but generally negligible.

Table B.7. Proficiency Level Descriptors, Grades 9-12

Please note that the performance level descriptors represent a progression of skills and abilities. Skills and abilities specified in lower-performance levels are likely demonstrated by students in the higher-performance levels and may not be noted in the higher-level descriptors for a grade or grade range.

Grades 9-12	Speaking	Listening	Reading	Writing
1 Beginning	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.	Beginning students are starting to develop receptive and productive skills in English.
2 Early Intermediate	Early Intermediate students typically produce simple sentences using basic vocabulary and grammar when interpreting language related to social, school, and academic contexts, explaining personal preferences or describing a sequence of pictures about familiar events and social situations. Minimal vocabulary and grammar knowledge and errors frequently impede communication.	Early Intermediate students typically follow multistep directions. They identify main ideas and draw simple inferences and conclusions. Errors frequently impede communication and comprehension.	Early intermediate students recall simple information from text, identify main ideas and supporting details, and make simple inferences. They identify common idiomatic expressions and paraphrase passages. Errors frequently impede comprehension.	Early Intermediate students typically write complete sentences using basic vocabulary and grammar to express ideas. They compare and summarize information found in texts or graphic organizers. They demonstrate a basic knowledge of auxiliary verbs, pronouns, and conjunctions. Errors in organization, grammar, word choice, and mechanics frequently impede communication.
3 Intermediate	Intermediate students typically use appropriate words and phrases and complete sentences when providing information, expressing preferences, conducting transactions, and describing personal experiences. They describe social situations, give instructions, and narrate a simple story. Intended meaning is mostly clear, but sometimes requires comprehension-check questions. They are capable of communicating some nuances of meaning. Grammatical or vocabulary errors interfere with communication, but the intended meaning is somewhat clear.	Intermediate students typically interpret simple academic vocabulary and idiomatic expressions. They extrapolate logical outcomes, place new information in a broader context, and recall details from classroom discussions or oral stories. Errors interfere with communication and comprehension.	Intermediate students typically use knowledge of high-frequency affixes and context clues to determine word meanings and identify synonyms of high-frequency social and academic vocabulary. From a simple narrative, they recall stated and implicit details, distinguish main ideas, compare and contrast information, draw conclusions, and make some inferences. Errors interfere with comprehension.	Intermediate students typically use correct basic grammar and begin to demonstrate use of conjunctions in compound sentences. They summarize texts and analyze information in graphic organizers. Meaning is somewhat clear, although vocabulary may be limited. Errors interfere with communication.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Grades 9-12	Speaking	Listening	Reading	Writing
4 Proficient	Proficient students typically use complete sentences to express opinions, explain processes, conduct transactions, and describe personal experiences. They use accurate vocabulary and grammar to describe the purpose of less common objects and fluently narrate stories with creative detail. They organize responses in logical and sequential order and incorporate idiomatic expressions. They convey subtle distinctions through rich, specific, and varied vocabulary.	Proficient students typically interpret idiomatic expressions and complex academic vocabulary and concepts. They distinguish essential details and nuances of meaning, synthesize answers from fragmentary information, and determine key information to summarize a task from complex narratives and discussions.	Proficient students typically draw complex conclusions from lengthy passages and distinguish nuances of meanings. They interpret alternate expressions of ideas, analyze the organization of passages, and identify theme, tone, and author purpose. Errors do not interfere with comprehension.	Proficient students typically write fluently, using complete sentences with accurate vocabulary to interpret texts and graphical information, while distinguishing nuances of meaning. They incorporate idiomatic expressions and produce responses to open-ended questions and write summaries and comparisons that correctly use verb forms, capitalization, punctuation, and advanced grammar. Responses exhibit minor errors in grammar and content organization that do not interfere with communication.
5 Above Proficient	Above Proficient students typically produce complex sentences with sophisticated and precise vocabulary and correct grammar. They convey detailed academic content and expressive nuances of meaning and skillfully organize information for presentations.	Above Proficient students typically interpret more complex grammar and academic vocabulary to follow complex instructions. They use context clues to interpret new vocabulary and draw conclusions about characters in oral stories. They distinguish subtleties of tone and point of view, recall extensive details, grasp abstract and uncommon idiomatic expressions, and analyze the structure of oral passages.	Above Proficient students recognize uncommon synonyms, subtle gradations of meanings using context clues, and unfamiliar idioms. They use prediction to read fluently, make inferences from challenging texts, synthesize text, recognize literary techniques, and use self-monitoring techniques to check for understanding.	Above Proficient students typically write using precise, sophisticated, and varied vocabulary. They demonstrate fluent and varied expression, express subtle nuances of meaning, and expand responses to prompts using related background knowledge. Minor errors are possible, but generally negligible.

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Appendix C Scoring Tables

Table C.1. Form E Kindergarten Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	300	123
	1	406	17
	2	418	12
	3	425	11
Level 2	4	431	10
	5	436	9
	6	440	9
	7	445	9
	8	449	9
	9	453	9
	10	458	9
Level 3	11	462	9
	12	467	9
	13	473	10
	14	478	10
	15	485	11
	16	493	13
Level 4	17	505	17
	18	526	24
Level 5	19	580	61

Table C.2. Form E Kindergarten Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	300	115
	1	300	115
	2	300	115
	3	300	115
	4	300	115
	5	343	72
	6	382	33
	7	395	20
	8	403	15
	9	410	13
	10	416	12
	11	421	12
Level 2	12	427	12
	13	432	12
	14	438	12
Level 3	15	445	13
	16	454	15
	17	465	18
Level 4	18	481	22
	19	500	21
Level 5	20	530	39

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.3. Form E Kindergarten Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	240	136
	1	240	136
	2	240	136
	3	240	136
	4	240	136
	5	240	136
	6	240	136
	7	240	136
	8	309	67
	9	333	43
	10	348	30
Level 2	11	359	21
	12	367	16
	13	373	14
	14	379	12
Level 3	15	384	12
	16	390	11
	17	395	12
	18	401	12
	19	408	13
	20	416	15
Level 4	21	426	17
	22	439	21
	23	462	32
Level 5	24	550	120

Table C.4. Form E Kindergarten Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	200	91
	1	200	91
	2	200	91
	3	200	91
	4	240	51
	5	268	39
	6	289	34
	7	306	31
	8	321	29
	9	336	27
Level 2	10	350	26
	11	364	24
	12	377	22
	13	389	19
	14	401	18
	15	413	18
Level 3	16	427	21
Level 4	17	448	28
Level 5	18	630	210

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.5. Form E Grade 1 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	300	120
	1	403	17
	2	414	12
	3	421	10
	4	427	9
	5	431	8
Level 2	6	434	8
	7	438	7
	8	441	7
	9	444	7
	10	447	7
	11	450	7
	12	453	7
	13	457	8
	14	460	8
Level 3	15	464	8
	16	467	8
	17	471	8
	18	475	9
	19	480	9
	20	485	10
	21	491	11
Level 4	22	498	12
	23	509	16
	24	529	23
Level 5	25	580	59

Table C.6. Form E Grade 1 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	300	115
	1	300	115
	2	300	115
	3	300	115
	4	300	115
	5	343	72
	6	382	33
	7	395	20
	8	403	15
	9	410	13
	10	416	12
	11	421	12
	12	427	12
Level 2	13	432	12
	14	438	12
	15	445	13
Level 3	16	454	15
	17	465	18
Level 4	18	481	22
	19	500	21
Level 5	20	530	39

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.7. Form E Grade 1 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	240	139
	1	240	139
	2	240	139
	3	240	139
	4	240	139
	5	240	139
	6	240	139
	7	240	139
	8	240	139
	9	287	92
	10	323	56
	11	341	39
	12	354	27
Level 2	13	363	20
	14	370	16
	15	375	14
	16	380	12
Level 3	17	385	12
	18	390	11
	19	394	11
	20	399	11
	21	404	11
	22	409	11
	23	414	11
	24	419	11
Level 4	25	425	11
	26	431	12
	27	439	14
	28	450	18
	29	470	28
Level 5	30	550	108

Table C.8. Form E Grade 1 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	200	91
	1	200	91
	2	200	91
	3	200	91
	4	239	52
	5	267	38
	6	287	33
	7	304	29
	8	318	27
	9	330	24
	10	342	22
	11	352	20

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 2	12	361	18
	13	369	16
	14	377	14
	15	383	13
	16	389	12
	17	394	11
	18	399	11
	19	404	11
	20	408	10
	21	413	10
	22	417	10
	23	422	11
	24	427	11
	25	432	12
Level 3	26	438	13
	27	446	15
	28	456	18
Level 4	29	474	26
Level 5	30	630	182

Table C.9. Form E Grade 2 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	350	85
	1	420	15
	2	430	11
	3	436	9
	4	441	8
Level 2	5	444	7
	6	447	7
	7	450	6
	8	453	6
	9	455	6
	10	457	6
	11	459	6
	12	462	6
	13	464	6
	14	466	6
	15	468	6
	16	470	6
Level 3	17	473	6
	18	476	6
	19	478	7
	20	481	7
	21	485	7
	22	489	8
	23	493	8
	24	498	9
	25	503	9

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 4	26	509	10
	27	516	11
	28	525	12
	29	535	13
	30	550	18
Level 5	31	600	68

Table C.10. Form E Grade 2 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	310	130
	1	310	130
	2	310	130
	3	310	130
	4	310	130
	5	381	59
	6	409	31
	7	422	21
	8	431	18
	9	440	17
Level 2	10	448	16
	11	456	16
Level 3	12	464	15
	13	471	14
	14	479	14
	15	486	14
Level 4	16	495	14
	17	504	15
	18	517	19
Level 5	19	543	32
	20	560	41

Table C.11. Form E Grade 2 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	300	156
	1	300	156
	2	300	156
	3	300	156
	4	300	156
	5	300	156
	6	300	156
	7	300	156
	8	300	156
	9	396	60
	10	418	38
	11	430	27

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 2	12	439	21
	13	447	17
	14	453	15
	15	458	14
	16	463	13
	17	468	12
Level 3	18	473	12
	19	478	11
	20	482	11
	21	487	11
	22	492	11
	23	497	11
Level 4	24	502	12
	25	508	12
	26	515	13
	27	523	14
	28	534	17
Level 5	29	551	24
	30	610	82

Table C.12. Form E Grade 2 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	270	144
	1	270	144
	2	270	144
	3	270	144
	4	369	45
	5	389	27
	6	401	20
	7	410	17
	8	417	15
	9	423	14
Level 2	10	429	13
	11	434	13
	12	439	12
	13	443	12
	14	448	12
	15	452	11
	16	457	11
	17	461	11
	18	465	11
	19	469	11
	20	474	11
Level 3	21	478	11
	22	483	11
	23	487	11
	24	492	12
	25	498	12
	26	503	13

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 4	27	510	13
	28	517	14
	29	526	16
	30	538	19
Level 5	31	556	25
	32	640	109

Table C.13. Form E Grade 3 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	350	85
	1	420	15
	2	430	11
	3	436	9
	4	441	8
Level 2	5	444	7
	6	447	7
	7	450	6
	8	453	6
	9	455	6
	10	457	6
	11	459	6
	12	462	6
	13	464	6
	14	466	6
	15	468	6
	16	470	6
	17	473	6
Level 3	18	476	6
	19	478	7
	20	481	7
	21	485	7
	22	489	8
	23	493	8
	24	498	9
	25	503	9
Level 4	26	509	10
	27	516	11
	28	525	12
	29	535	13
	30	550	18
Level 5	31	600	68

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.14. Form E Grade 3 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	310	130
	1	310	130
	2	310	130
	3	310	130
	4	310	130
	5	381	59
	6	409	31
	7	422	21
	8	431	18
	9	440	17
Level 2	10	448	16
	11	456	16
	12	464	15
Level 3	13	471	14
	14	479	14
	15	486	14
	16	495	14
Level 4	17	504	15
	18	517	19
	19	543	32
Level 5	20	560	41

Table C.15. Form E Grade 3 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	300	156
	1	300	156
	2	300	156
	3	300	156
	4	300	156
	5	300	156
	6	300	156
	7	300	156
	8	300	156
	9	396	60
	10	418	38
	11	430	27
Level 2	12	439	21
	13	447	17
	14	453	15
	15	458	14
	16	463	13
	17	468	12
	18	473	12
Level 3	19	478	11
	20	482	11
	21	487	11
	22	492	11
	23	497	11
	24	502	12

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 4	25	508	12
	26	515	13
	27	523	14
	28	534	17
Level 5	29	551	24
	30	610	82

Table C.16. Form E Grade 3 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	270	144
	1	270	144
	2	270	144
	3	270	144
	4	369	45
	5	389	27
	6	401	20
	7	410	17
	8	417	15
	9	423	14
Level 2	10	429	13
	11	434	13
	12	439	12
	13	443	12
	14	448	12
	15	452	11
	16	457	11
	17	461	11
	18	465	11
	19	469	11
	20	474	11
	21	478	11
	22	483	11
Level 3	23	487	11
	24	492	12
	25	498	12
	26	503	13
	27	510	13
	28	517	14
	29	526	16
Level 4	30	538	19
	31	556	25
Level 5	32	640	109

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.17. Form E Grade 4 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	360	75
	1	415	20
	2	426	15
	3	433	13
	4	440	11
	5	445	10
Level 2	6	449	10
	7	453	9
	8	457	9
	9	461	9
	10	465	8
	11	468	8
	12	472	8
Level 3	13	476	9
	14	480	9
	15	485	9
	16	489	9
	17	494	9
	18	499	9
	19	503	8
	20	507	8
Level 4	21	512	8
	22	516	8
	23	520	8
	24	524	8
	25	529	8
	26	534	9
	27	540	0
	28	546	1
	29	556	4
Level 5	30	572	21
	31	635	84

Table C.18. Form E Grade 4 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	350	126
	1	350	126
	2	350	126
	3	350	126
	4	350	126
	5	356	120
	6	420	56
	7	441	35
	8	455	27
Level 2	9	467	23
	10	477	21

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	11	486	19
	12	495	18
	13	503	18
	14	512	18
	15	521	18
Level 4	16	531	19
	17	543	21
	18	559	26
Level 5	19	587	38
	20	640	79

Table C.19. Form E Grade 4 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	360	136
	1	360	136
	2	360	136
	3	360	136
	4	360	136
	5	360	136
	6	360	136
	7	418	78
	8	451	45
	9	466	30
Level 2	10	477	23
	11	485	20
	12	493	17
	13	499	16
Level 3	14	505	15
	15	511	15
	16	517	15
	17	522	14
	18	528	14
	19	534	14
Level 4	20	540	14
	21	546	14
	22	552	15
	23	559	15
	24	567	16
	25	575	16
	26	584	17
Level 5	27	595	18
	28	609	21
	29	630	29
	30	680	71

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.20. Form E Grade 4 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	290	134
	1	290	134
	2	290	134
	3	369	55
	4	395	33
	5	411	25
	6	422	21
	7	432	19
Level 2	8	440	17
	9	448	16
	10	455	16
	11	462	15
	12	468	15
	13	474	15
	14	481	15
	15	487	15
	16	493	15
Level 3	17	500	15
	18	506	15
	19	512	15
	20	519	15
	21	526	15
	22	532	15
Level 4	23	539	15
	24	546	15
	25	554	16
	26	562	16
	27	570	17
	28	581	19
Level 5	29	593	21
	30	610	26
	31	638	36
	32	680	63

Table C.21. Form E Grade 5 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	360	75
	1	415	20
	2	426	15
	3	433	13
	4	440	11
	5	445	10
Level 2	6	449	10
	7	453	9
	8	457	9
	9	461	9
	10	465	8
	11	468	8
	12	472	8

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	13	476	9
	14	480	9
	15	485	9
	16	489	9
	17	494	9
	18	499	9
	19	503	8
	20	507	8
Level 4	21	512	8
	22	516	8
	23	520	8
	24	524	8
	25	529	8
	26	534	9
	27	540	10
	28	546	11
	29	556	14
Level 5	30	572	21
	31	635	84

Table C.22. Form E Grade 5 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	350	126
	1	350	126
	2	350	126
	3	350	126
	4	350	126
	5	356	120
	6	420	56
	7	441	35
	8	455	27
Level 2	9	467	23
	10	477	21
	11	486	19
Level 3	12	495	18
	13	503	18
	14	512	18
	15	521	18
Level 4	16	531	19
	17	543	21
	18	559	26
Level 5	19	587	38
	20	640	79

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.23. Form E Grade 5 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	360	136
	1	360	136
	2	360	136
	3	360	136
	4	360	136
	5	360	136
	6	360	136
	7	418	78
	8	451	45
	9	466	30
Level 2	10	477	23
	11	485	20
	12	493	17
	13	499	16
Level 3	14	505	15
	15	511	15
	16	517	15
	17	522	14
	18	528	14
	19	534	14
Level 4	20	540	14
	21	546	14
	22	552	15
	23	559	15
	24	567	16
	25	575	16
	26	584	17
Level 5	27	595	18
	28	609	21
	29	630	29
	30	680	71

Table C.24. Form E Grade 5 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	290	134
	1	290	134
	2	290	134
	3	369	55
	4	395	33
	5	411	25
	6	422	21
	7	432	19

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 2	8	440	17
	9	448	16
	10	455	16
	11	462	15
	12	468	15
	13	474	15
	14	481	15
	15	487	15
	16	493	15
Level 3	17	500	15
	18	506	15
	19	512	15
	20	519	15
	21	526	15
	22	532	15
Level 4	23	539	15
	24	546	15
	25	554	16
	26	562	16
	27	570	17
	28	581	19
Level 5	29	593	21
	30	610	26
	31	638	36
	32	680	63

Table C.25. Form E Grade 6 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	365	82
	1	426	21
	2	438	15
	3	445	13
Level 2	4	451	11
	5	456	10
	6	460	9
	7	464	9
	8	467	9
	9	471	8
	10	474	8
Level 3	11	477	8
	12	480	8
	13	484	9
	14	487	9
	15	491	9
	16	495	9
	17	499	9
	18	504	9
	19	508	9

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 4	20	513	10
	21	519	10
	22	524	10
	23	530	10
	24	536	10
	25	542	10
	26	548	10
	27	554	10
Level 5	28	562	12
	29	572	15
	30	589	22
	31	645	77

Table C.26. Form E Grade 6 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	360	112
	1	360	112
	2	360	112
	3	360	112
	4	360	112
	5	360	112
	6	382	90
	7	422	50
	8	441	33
	9	454	26
Level 2	10	465	23
	11	474	21
	12	483	20
Level 3	13	492	19
	14	500	19
	15	509	20
	16	519	20
	17	529	21
Level 4	18	540	22
	19	554	24
	20	569	26
Level 5	21	589	30
	22	620	41
	23	680	84

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.27. Form E Grade 6 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	380	135
	1	380	135
	2	380	135
	3	380	135
	4	380	135
	5	380	135
	6	380	135
	7	452	63
	8	477	38
	9	490	27
	10	500	21
Level 2	11	508	18
	12	514	16
	13	520	15
	14	526	14
Level 3	15	531	13
	16	536	13
	17	541	13
	18	546	13
	19	551	13
	20	556	13
Level 4	21	562	13
	22	567	13
	23	574	14
	24	580	14
	25	588	15
	26	596	16
	27	606	17
Level 5	28	619	21
	29	640	29
	30	690	72

Table C.28. Form E Grade 6 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	300	148
	1	300	148
	2	321	127
	3	395	53
	4	420	33
	5	437	25
Level 2	6	449	21
	7	459	19
	8	467	18
	9	475	17
	10	482	16
	11	489	16
	12	496	16

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	13	503	16
	14	509	16
	15	516	16
	16	523	16
	17	531	16
	18	538	16
	19	546	16
Level 4	20	554	17
	21	561	17
	22	570	17
	23	578	17
	24	587	17
Level 5	25	596	18
	26	606	18
	27	616	19
	28	628	20
	29	641	22
	30	658	26
	31	685	36
	32	710	51

Table C.29. Form E Grade 7 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	365	82
	1	426	21
	2	438	15
	3	445	13
Level 2	4	451	11
	5	456	10
	6	460	9
	7	464	9
	8	467	9
	9	471	8
	10	474	8
Level 3	11	477	8
	12	480	8
	13	484	9
	14	487	9
	15	491	9
	16	495	9
	17	499	9
	18	504	9
	19	508	9
Level 4	20	513	10
	21	519	10
	22	524	10
	23	530	10
	24	536	10
	25	542	10
	26	548	10
	27	554	10

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 5	28	562	12
	29	572	15
	30	589	22
	31	645	77

Table C.30. Form E Grade 7 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	360	112
	1	360	112
	2	360	112
	3	360	112
	4	360	112
	5	360	112
	6	382	90
	7	422	50
	8	441	33
	9	454	26
Level 2	10	465	23
	11	474	21
	12	483	20
Level 3	13	492	19
	14	500	19
	15	509	20
	16	519	20
	17	529	21
Level 4	18	540	22
	19	554	24
	20	569	26
Level 5	21	589	30
	22	620	41
	23	680	84

Table C.31. Form E Grade 7 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	380	135
	1	380	135
	2	380	135
	3	380	135
	4	380	135
	5	380	135
	6	380	135
	7	452	63
	8	477	38
	9	490	27
	10	500	21
Level 2	11	508	18
	12	514	16
	13	520	15
	14	526	14

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	15	531	13
	16	536	13
	17	541	13
	18	546	13
	19	551	13
	20	556	13
Level 4	21	562	13
	22	567	13
	23	574	14
	24	580	14
	25	588	15
	26	596	16
	27	606	17
Level 5	28	619	21
	29	640	29
	30	690	72

Table C.32. Form E Grade 7 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	300	148
	1	300	148
	2	321	127
	3	395	53
	4	420	33
	5	437	25
Level 2	6	449	21
	7	459	19
	8	467	18
	9	475	17
	10	482	16
	11	489	16
	12	496	16
Level 3	13	503	16
	14	509	16
	15	516	16
	16	523	16
	17	531	16
	18	538	16
	19	546	16
Level 4	20	554	17
	21	561	17
	22	570	17
	23	578	17
	24	587	17

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 5	25	596	18
	26	606	18
	27	616	19
	28	628	20
	29	641	22
	30	658	26
	31	685	36
	32	710	51

Table C.33. Form E Grade 8 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	365	82
	1	426	21
	2	438	15
	3	445	13
Level 2	4	451	11
	5	456	10
	6	460	9
	7	464	9
	8	467	9
	9	471	8
	10	474	8
Level 3	11	477	8
	12	480	8
	13	484	9
	14	487	9
	15	491	9
	16	495	9
	17	499	9
	18	504	9
	19	508	9
	20	513	10
Level 4	21	519	10
	22	524	10
	23	530	10
	24	536	10
	25	542	10
	26	548	10
	27	554	10
Level 5	28	562	12
	29	572	15
	30	589	22
	31	645	77

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.34. Form E Grade 8 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	360	112
	1	360	112
	2	360	112
	3	360	112
	4	360	112
	5	360	112
	6	382	90
	7	422	50
	8	441	33
	9	454	26
	10	465	23
Level 2	11	474	21
	12	483	20
	13	492	19
Level 3	14	500	19
	15	509	20
	16	519	20
	17	529	21
Level 4	18	540	22
	19	554	24
	20	569	26
	21	589	30
Level 5	22	620	41
	23	680	84

Table C.35. Form E Grade 8 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	380	135
	1	380	135
	2	380	135
	3	380	135
	4	380	135
	5	380	135
	6	380	135
	7	452	63
	8	477	38
	9	490	27
	10	500	21
Level 2	11	508	18
	12	514	16
	13	520	15
	14	526	14
	15	531	13
Level 3	16	536	13
	17	541	13
	18	546	13
	19	551	13
	20	556	13

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 4	21	562	13
	22	567	13
	23	574	14
	24	580	14
	25	588	15
	26	596	16
	27	606	17
Level 5	28	619	21
	29	640	29
	30	690	72

Table C.36. Form E Grade 8 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	300	148
	1	300	148
	2	321	127
	3	395	53
	4	420	33
	5	437	25
Level 2	6	449	21
	7	459	19
	8	467	18
	9	475	17
	10	482	16
	11	489	16
	12	496	16
Level 3	13	503	16
	14	509	16
	15	516	16
	16	523	16
	17	531	16
	18	538	16
	19	546	16
Level 4	20	554	17
	21	561	17
	22	570	17
	23	578	17
	24	587	17
Level 5	25	596	18
	26	606	18
	27	616	19
	28	628	20
	29	641	22
	30	658	26
	31	685	36
	32	710	51

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.37. Form E Grade 9 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	370	85
	1	442	13
	2	451	10
Level 2	3	456	8
	4	460	7
	5	464	7
	6	467	6
	7	470	6
	8	472	6
	9	475	6
	10	477	6
Level 3	11	480	6
	12	483	6
	13	485	6
	14	488	6
	15	491	6
	16	494	6
	17	497	6
	18	500	7
	19	504	7
	20	507	7
	21	511	7
Level 4	22	515	7
	23	519	7
	24	523	7
	25	528	8
	26	533	8
	27	539	9
	28	546	10
	29	556	12
Level 5	30	576	20
	31	650	94

Table C.38. Form E Grade 9 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	370	135
	1	370	135
	2	370	135
	3	370	135
	4	370	135
	5	370	135
	6	377	128
	7	438	67
	8	463	42
Level 2	9	479	32
	10	492	26
	11	502	23

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	12	512	21
	13	521	20
	14	530	20
	15	539	20
Level 4	16	548	20
	17	558	21
	18	569	22
	19	582	24
	20	597	27
	21	619	34
Level 5	22	656	52
	23	730	126

Table C.39. Form E Grade 9 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	390	120
	1	390	120
	2	390	120
	3	390	120
	4	390	120
	5	390	120
	6	390	120
	7	434	76
	8	464	46
	9	480	30
	10	490	22
	11	498	18
	12	504	15
Level 2	13	510	14
	14	515	13
	15	520	12
	16	524	12
	17	529	11
	18	533	11
	19	538	12
	20	543	12
Level 3	21	548	12
	22	553	13
	23	559	13
	24	566	14
	25	574	16
Level 4	26	583	17
	27	594	19
	28	609	23
Level 5	29	632	31
	30	715	114

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.40. Form E Grade 9 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	310	126
	1	310	126
	2	310	126
	3	372	64
	4	402	37
	5	420	28
	6	433	23
	7	444	21
Level 2	8	453	19
	9	461	18
	10	469	17
	11	476	17
	12	483	16
	13	490	16
	14	496	16
Level 3	15	503	16
	16	509	16
	17	516	16
	18	522	16
	19	529	16
	20	536	16
	21	543	17
Level 4	22	551	17
	23	559	18
	24	568	19
	25	578	20
	26	589	21
Level 5	27	602	23
	28	616	25
	29	634	28
	30	657	34
	31	695	50
	32	720	68

Table C.41. Form E Grade 10 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	370	85
	1	442	13
	2	451	10
Level 2	3	456	8
	4	460	7
	5	464	7
	6	467	6
	7	470	6
	8	472	6
	9	475	6
	10	477	6

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	11	480	6
	12	483	6
	13	485	6
	14	488	6
	15	491	6
	16	494	6
	17	497	6
	18	500	7
	19	504	7
	20	507	7
	21	511	7
	22	515	7
Level 4	23	519	7
	24	523	7
	25	528	8
	26	533	8
	27	539	9
	28	546	10
	29	556	12
Level 5	30	576	20
	31	650	94

Table C.42. Form E Grade 10 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	370	135
	1	370	135
	2	370	135
	3	370	135
	4	370	135
	5	370	135
	6	377	128
	7	438	67
	8	463	42
Level 2	9	479	32
	10	492	26
	11	502	23
Level 3	12	512	21
	13	521	20
	14	530	20
	15	539	20
	16	548	20
Level 4	17	558	21
	18	569	22
	19	582	24
	20	597	27
	21	619	34
Level 5	22	656	52
	23	730	126

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.43. Form E Grade 10 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	390	120
	1	390	120
	2	390	120
	3	390	120
	4	390	120
	5	390	120
	6	390	120
	7	434	76
	8	464	46
	9	480	30
	10	490	22
	11	498	18
	12	504	15
Level 2	13	510	14
	14	515	13
	15	520	12
	16	524	12
	17	529	11
	18	533	11
	19	538	12
	20	543	12
Level 3	21	548	12
	22	553	13
	23	559	13
	24	566	14
	25	574	16
Level 4	26	583	17
	27	594	19
	28	609	23
	29	632	31
Level 5	30	715	114

Table C.44. Form E Grade 10 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	310	126
	1	310	126
	2	310	126
	3	372	64
	4	402	37
	5	420	28
	6	433	23
	7	444	21
Level 2	8	453	19
	9	461	18
	10	469	17
	11	476	17
	12	483	16
	13	490	16
	14	496	16

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	15	503	16
	16	509	16
	17	516	16
	18	522	16
	19	529	16
	20	536	16
	21	543	17
Level 4	22	551	17
	23	559	18
	24	568	19
	25	578	20
	26	589	21
Level 5	27	602	23
	28	616	25
	29	634	28
	30	657	34
	31	695	50
	32	720	68

Table C.45. Form E Grade 11 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	370	85
	1	442	13
	2	451	10
Level 2	3	456	8
	4	460	7
	5	464	7
	6	467	6
	7	470	6
	8	472	6
	9	475	6
	10	477	6
Level 3	11	480	6
	12	483	6
	13	485	6
	14	488	6
	15	491	6
	16	494	6
	17	497	6
	18	500	7
	19	504	7
	20	507	7
	21	511	7
	22	515	7
Level 4	23	519	7
	24	523	7
	25	528	8
	26	533	8
	27	539	9
	28	546	10
	29	556	12

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 5	30	576	20
	31	650	94

Table C.46. Form E Grade 11 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	370	135
	1	370	135
	2	370	135
	3	370	135
	4	370	135
	5	370	135
	6	377	128
	7	438	67
	8	463	42
	9	479	32
Level 2	10	492	26
	11	502	23
	12	512	21
Level 3	13	521	20
	14	530	20
	15	539	20
	16	548	20
Level 4	17	558	21
	18	569	22
	19	582	24
	20	597	27
	21	619	34
Level 5	22	656	52
	23	730	126

Table C.47. Form E Grade 11 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	390	120
	1	390	120
	2	390	120
	3	390	120
	4	390	120
	5	390	120
	6	390	120
	7	434	76
	8	464	46
	9	480	30
	10	490	22
	11	498	18
	12	504	15

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 2	13	510	14
	14	515	13
	15	520	12
	16	524	12
	17	529	11
	18	533	11
	19	538	12
	20	543	12
Level 3	21	548	12
	22	553	13
	23	559	13
	24	566	14
	25	574	16
Level 4	26	583	17
	27	594	19
	28	609	23
	29	632	31
Level 5	30	715	114

Table C.48. Form E Grade 11 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	310	126
	1	310	126
	2	310	126
	3	372	64
	4	402	37
	5	420	28
	6	433	23
	7	444	21
Level 2	8	453	19
	9	461	18
	10	469	17
	11	476	17
	12	483	16
	13	490	16
	14	496	16
Level 3	15	503	16
	16	509	16
	17	516	16
	18	522	16
	19	529	16
	20	536	16
	21	543	17
Level 4	22	551	17
	23	559	18
	24	568	19
	25	578	20
	26	589	21

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 5	27	602	23
	28	616	25
	29	634	28
	30	657	34
	31	695	50
	32	720	68

Table C.49. Form E Grade 12 Scoring Table: Speaking

Proficiency Level	RS	SS	SEM
Level 1	0	370	85
	1	442	13
	2	451	10
Level 2	3	456	8
	4	460	7
	5	464	7
	6	467	6
	7	470	6
	8	472	6
	9	475	6
	10	477	6
Level 3	11	480	6
	12	483	6
	13	485	6
	14	488	6
	15	491	6
	16	494	6
	17	497	6
	18	500	7
	19	504	7
	20	507	7
	21	511	7
	22	515	7
Level 4	23	519	7
	24	523	7
	25	528	8
	26	533	8
	27	539	9
	28	546	10
	29	556	12
Level 5	30	576	20
	31	650	94

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table C.50. Form E Grade 12 Scoring Table: Listening

Proficiency Level	RS	SS	SEM
Level 1	0	370	135
	1	370	135
	2	370	135
	3	370	135
	4	370	135
	5	370	135
	6	377	128
	7	438	67
	8	463	42
	9	479	32
Level 2	10	492	26
	11	502	23
	12	512	21
Level 3	13	521	20
	14	530	20
	15	539	20
	16	548	20
	17	558	21
Level 4	18	569	22
	19	582	24
	20	597	27
	21	619	34
Level 5	22	656	52
	23	730	126

Table C.51. Form E Grade 12 Scoring Table: Reading

Proficiency Level	RS	SS	SEM
Level 1	0	390	120
	1	390	120
	2	390	120
	3	390	120
	4	390	120
	5	390	120
	6	390	120
	7	434	76
	8	464	46
	9	480	30
	10	490	22
	11	498	18
	12	504	15
Level 2	13	510	14
	14	515	13
	15	520	12
	16	524	12
	17	529	11
	18	533	11
	19	538	12
	20	543	12

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Proficiency Level	RS	SS	SEM
Level 3	21	548	12
	22	553	13
	23	559	13
	24	566	14
	25	574	16
	26	583	17
Level 4	27	594	19
	28	609	23
	29	632	31
Level 5	30	715	114

Table C.52. Form E Grade 12 Scoring Table: Writing

Proficiency Level	RS	SS	SEM
Level 1	0	310	126
	1	310	126
	2	310	126
	3	372	64
	4	402	37
	5	420	28
	6	433	23
	7	444	21
Level 2	8	453	19
	9	461	18
	10	469	17
	11	476	17
	12	483	16
	13	490	16
	14	496	16
Level 3	15	503	16
	16	509	16
	17	516	16
	18	522	16
	19	529	16
	20	536	16
	21	543	17
Level 4	22	551	17
	23	559	18
	24	568	19
	25	578	20
	26	589	21
Level 5	27	602	23
	28	616	25
	29	634	28
	30	657	34
	31	695	50
	32	720	68

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Appendix D Item Difficulty

Table D.1. Form E Speaking Item Difficulty

Item	Grade K	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
1	0.82	0.93	0.93	0.96	0.97	0.94
2	0.86	0.92	0.62	0.99	0.94	0.98
3	-	-	0.98	0.91	-	0.88
4	0.94	0.97	0.71	0.89	0.69	0.86
5	0.93	0.97	0.84	0.89	0.66	0.30
6	0.82	0.94	0.74	0.61	0.58	0.60
7	0.79	0.87	0.71	0.80	0.43	0.31
8	0.65	0.81	0.88	0.88	0.82	0.89
9	0.55	0.71	0.91	0.91	0.85	0.93
10	0.59	0.74	0.87	0.58	0.66	0.78
11	0.57	0.74	0.83	0.62	0.67	0.78
12	0.59	0.76	0.78	0.68	0.65	0.78
13	0.66	0.81	0.78	0.61	0.78	0.86
14	-	0.82	0.84	0.71	0.66	0.80
15	-	0.77	0.85	0.73	0.68	0.80
16	-	0.76	0.82	0.76	0.73	0.79
17	-	0.68	0.79	0.85	0.77	0.86
18	0.50	0.64	0.73	0.73	0.65	0.77

Table D.2. Form E Listening Item Difficulty

Item	Grade K	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
1	0.80	0.95	0.47	0.90	0.83	1
2	0.67	0.81	0.70	0.75	0.68	2
3	0.59	0.69	0.65	0.76	0.49	3
4	0.72	0.89	0.93	0.78	0.55	4
5	0.62	0.91	0.71	0.91	0.40	5
6	0.89	0.95	0.82	0.81	0.72	6
7	0.75	0.76	0.78	0.92	0.58	7
8	0.79	0.91	0.81	0.58	0.45	8
9	0.88	0.87	0.58	0.77	0.64	9
10	0.80	0.53	0.68	0.84	0.68	10
11	0.74	0.75	0.83	0.60	0.52	11
12	0.79	0.92	0.65	0.70	0.74	12
13	0.72	0.87	0.66	0.55	0.84	13
14	0.70	0.75	0.49	0.47	0.72	14
15	0.73	0.73	0.64	0.49	0.84	15
16	0.63	0.43	0.81	0.59	0.72	16
17	0.83	0.36	0.75	0.77	0.73	17

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Item	Grade K	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
18	0.60	0.74	0.42	0.47	0.58	18
19	0.63	0.91	0.51	0.44	0.65	19
20	0.37	0.70	0.47	0.49	0.44	20
21	-	-	-	0.58	0.19	21
22	-	-	-	0.81	0.85	22
23	-	-	-	0.74	0.58	23

Table D.3. Form E Reading Item Difficulty

Item	Grade K	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
1	0.47	0.56	0.68	0.77	0.72	0.79
2	0.87	0.93	0.66	0.81	0.78	0.69
3	0.76	0.82	0.83	0.67	0.52	0.49
4	0.68	0.82	0.56	0.44	0.52	0.41
5	0.69	0.80	0.66	0.58	0.62	0.77
6	0.63	0.81	0.75	0.61	0.48	0.56
7	0.61	0.80	0.76	0.50	0.72	0.51
8	0.66	0.83	0.65	0.47	0.40	0.69
9	0.57	0.78	0.58	0.62	0.57	0.41
10	0.51	0.72	0.76	0.77	0.69	0.47
11	0.56	0.75	0.77	0.51	0.52	0.71
12	0.55	0.75	0.74	0.56	0.82	0.77
13	0.77	0.85	0.48	0.68	0.88	0.70
14	0.59	0.75	0.74	0.56	0.73	0.57
15	0.57	0.73	0.69	0.53	0.29	0.51
16	0.65	0.77	0.73	0.55	0.42	0.44
17	0.53	0.69	0.59	0.34	0.48	0.62
18	0.75	0.86	0.41	0.52	0.53	0.50
19	0.41	0.58	0.65	0.68	0.61	0.42
20	0.20	0.31	0.58	0.49	0.46	0.53
21	0.27	0.43	0.60	0.56	0.53	0.52
22	0.41	0.55	0.47	0.77	0.44	0.46
23	-	0.59	0.40	0.50	0.41	0.72
24	-	0.54	0.42	0.37	0.65	0.82
25	-	0.45	0.48	0.47	0.53	0.48
26	-	0.43	0.60	0.64	0.66	0.46
27	0.55	0.72	0.62	0.67	0.26	0.54
28	0.53	0.67	0.66	0.30	0.46	0.49
29	0.45	0.61	0.58	0.24	0.32	0.77
30	0.44	0.59	0.66	0.41	0.61	0.77

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table D.4. Form E Writing Item Difficulty

Item	Grade K	Grade 1	Grades 2-3	Grades 4-5	Grades 6-8	Grades 9-12
1	0.92	0.98	0.67	0.88	0.74	0.88
2	0.92	0.99	0.57	0.78	0.79	0.74
3	0.67	0.84	0.61	0.80	0.73	0.67
4	0.79	0.94	0.55	0.73	0.65	0.42
5	0.63	0.83	0.57	0.91	0.63	0.71
6	0.45	0.83	0.57	0.79	0.67	0.91
7	0.31	0.65	0.72	0.67	0.59	0.87
8	0.44	0.80	0.54	0.88	0.63	0.50
9	0.47	0.86	0.41	0.90	0.59	0.49
10	0.22	0.58	0.63	0.59	0.55	0.43
11	-	0.62	0.79	0.53	0.55	0.47
12	-	0.67	0.78	0.63	0.38	0.65
13	-	0.63	0.75	0.48	0.41	0.54
14	-	0.56	0.77	0.52	0.46	0.62
15	0.46	0.60	0.67	0.52	0.50	0.61
16	0.48	0.56	0.60	0.47	0.39	0.54
17	0.54	0.81	0.70	0.65	0.65	0.76
18	0.69	0.87	-	-	-	-
19	0.56	0.82	-	-	-	-
20	0.47	0.68	-	-	-	-

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Appendix E Inter-rater Reliability

Table E.1. Speaking Grade K Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	96	4	0
05	1	94	6	0
06	1	97	3	0
08	3	86	13	1
09	3	96	4	0
13	3	89	11	0
14	3	89	11	0
15	4	90	10	0

Table E.2. Form E Writing Grade K Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
73	1	100	0	0
74	1	100	0	0
78	3	97	3	0
79	1	99	1	0
80	1	99	1	0
82	3	95	5	0

Table E.3 Form E Speaking Grades 1 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	100	0	0
05	1	99	1	0
06	1	99	1	0
08	3	93	7	0
09	3	93	7	0
13	3	95	5	0
14	3	91	9	0
15	3	91	9	0
16	3	94	6	0
17	4	89	11	0

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table E.4. Form E Writing Grades 1 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
81	1	100	0	0
82	1	97	3	0
85	3	98	2	0
86	1	99	1	0
87	1	99	1	0
90	3	92	8	0
92	3	94	6	0
93	3	91	9	0
94	3	96	4	0
95	3	92	8	0

Table E.5. Form E Speaking Grades 2–3 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	98	2	0
05	1	98	2	0
06	1	97	3	0
09	3	93	7	0
10	3	93	7	0
13	3	91	9	0
14	3	89	11	0
15	3	92	8	0
16	3	89	11	0
17	3	94	6	0
18	3	92	8	0
19	4	85	15	0

Table E.6 Form E Writing Grades 2–3 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
94	3	97	3	0
95	3	93	7	0
96	3	92	8	0
97	3	95	5	0

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
98	3	95	5	0
99	3	94	6	0
102	4	95	5	0

Table E.7. Form E Speaking Grades 4–5 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	99	1	0
05	1	99	1	0
06	1	99	1	0
09	3	93	7	0
10	3	95	5	0
13	3	93	7	0
14	3	91	8	1
15	3	88	12	0
16	3	92	8	0
17	3	91	9	0
18	3	88	12	0
19	4	85	15	0

Table E.8. Form E Writing Grades 4–5 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
92	3	90	10	0
93	3	91	9	0
94	3	91	9	0
95	3	92	8	0
96	3	93	7	0
97	3	88	12	0
101	4	86	14	0

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Table E.9. Form E Speaking Grades 6–8 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	97	3	0
05	1	94	6	0
06	1	99	1	0
09	3	84	16	0
10	3	84	16	0
13	3	91	9	0
14	3	89	11	0
15	3	90	10	0
16	3	87	13	0
17	3	91	9	0
18	3	86	14	0
19	4	87	13	0

Table E.10. Form E Writing Grades 6–8 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
95	3	94	6	0
96	3	94	6	0
97	3	95	5	0
98	3	97	3	0
99	3	96	4	0
100	3	92	8	0
104	4	90	10	0

Table E.11. Form E Speaking Grades 9–12 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
02	1	97	3	0
05	1	98	2	0
06	1	95	5	0
09	3	95	5	0
10	3	90	10	0
13	3	94	6	0
14	3	90	10	0

Copyright © 2026 by Data Recognition Corporation. All rights reserved.

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
15	3	94	6	0
16	3	95	5	0
17	3	95	5	0
18	3	96	4	0
19	4	94	6	0

Table E.12. Form E Writing Grades 9–12 Inter-rater Reliability

Item	Maximum Score	Percentage Absolute Difference: Perfect	Percentage Absolute Difference: Adjacent	Percentage Absolute Difference: Discrepant
95	3	93	7	0
96	3	94	6	0
97	3	93	7	0
98	3	93	7	0
99	3	94	6	0
100	3	90	10	0
104	4	87	13	0

Copyright © 2026 by Data Recognition Corporation. All rights reserved.