RESEARCH BASE

Burst Reading Assessment (Now mCLASS Intervention) Technical Manual

Amplify.

Table of Contents

Introduction

Overview of Burst®:Reading Assessment

Amplify's Burst:Reading Assessment is a collection of measures administered using mobile technology and designed to help educators ascertain student reading ability in grades K–6. These measures were designed for use within the Burst:Reading Intervention alongside the Dynamic Indicators of Basic Early Literacy Skills – Next Edition (DIBELS Next®; Good, Kaminski, Dewey, Walin, Powell-Smith, & Latimer, 2013) to screen students for intervention need and to guide intervention grouping and placement. The Burst:Reading assessments target three reading skills that empirically relate to future literacy outcomes, as highlighted by the National Reading Panel: Alphabetic Principle, Vocabulary, and Comprehension (National Institute of Child Health and Human Development, 2000).

Collectively, the Burst:Reading assessments measure these essential literacy skills using three research-based measures. The measures serve as brief, targeted, screening and progress- monitoring assessments to help teachers understand their students' instructional needs in content areas that often go unmeasured.

The Burst:Reading assessments include these measures:

- Decode, a measure of the alphabetic principle;
- Vocabulary, a measure of vocabulary breadth;
- Comprehension Skills, a measure of reading comprehension.

The Burst:Reading assessments were originally developed in 2009 and were subsequently revised in 2013. The following sections summarize the rationale for developing the measures and how the measures provide teachers with targeted information on student skill deficits and instructional needs.

Burst Decode

Theory

Beginning readers learning to decode are typically taught the 26 most common letter-sound correspondences of English (e.g., **a** is pronounced /ae/ as in **cat**, **b** is pronounced /b/ as in **ball**) and how to blend these sounds into words; with this knowledge, students can sound out regular words in which each letter represents its most common sound. Given that English orthography is not fully transparent (i.e., the relationship between sounds and letters is many-to-one rather than one-

to-one), students must extend their knowledge to include alternate pronunciations of letters (e.g., c can be pronounced /k/ or /s/, depending on its position in a word) and combinations of letters, such as combinations that form single sounds (e.g., ph is pronounced /f/), syllables (e.g., bet as in better and alphabet), and morphemes, which are the smallest parts of words that convey meaning (e.g., the superlative affix –est, as in greatest; Carlisle, 2003). Ehri (1998) describes how students develop these skills during different phases of sight word development. In the partial alphabetic phase, students recognize words based on some, but not all, of their letters (e.g., they recognize the word spoon based on the s and n), while in the full alphabetic phase, students can pronounce each of the letters in a word to decode it. This type of reading is laborious and prohibitively slow for the purpose of reading comprehension (Samuels & Flor, 1997); in the consolidated alphabetic phase, readers begin to recognize multi-letter units in words (e.g., rimes such as ost in lost and cost) that allow them to decode words more quickly. With practice, entire words are consolidated as sight words, leading to the sight word phase, which is characterized by mature and fluent reading that frees up processing resources for comprehension (Samuels, 2002).

Beginning readers do not neatly fall into one phase or another: at any given moment in their reading development, they may recognize some words fully, other words partially with the help of context, and they may need to sound out yet other words letter-by-letter (Chall, 1983; Barron, 1986; Perfetti, Bell, & Delaney, 1988). Even mature readers may revert to alphabetic reading when presented with an unfamiliar word (e.g., sesquipedalian). Thus, the phases of alphabetic reading may also be thought of as skills or strategies for word reading that both beginning and advanced readers can use and should be explicitly taught. The object of the **Decode (DEC)** measure is to gauge the degree to which students acquired different skills that both pave the way for the eventual consolidation of sight words and facilitate on-the-spot decoding when necessary.

Design

Four DEC submeasures were developed for the Burst:Reading assessments in grades 1–6[1]: Regular Words, Letter Combinations, Advanced Phonics, and Irregular Words. Each of the four submeasures targets different word reading skills developed in the full alphabetic, consolidated alphabetic, and sight word phases of reading that contribute to fluent reading (Carnine, Silbert, Kame'enui, & Tarver, 2004). Each submeasure consists of a printed list of eight words the student reads aloud during an independent timed reading, and the administrator marks words as correct,

1 Of the three Burst:Reading assessments, only DEC is available for use with students in grades 4–6. Empirical study results are reported for students in grades K–3, as the DEC measures were designed with these grades in mind; however, the DEC measures are recommended for use with students in grades 4–6 who are severely behind in word reading skills.

incorrect, or no response if the student is unable to respond to a word within 5 seconds. The maximum score on each DEC submeasure is 8.

Paris (2005) made the distinction between constrained skills that can be mastered within a relatively short period of time and unconstrained skills that can never be fully mastered. While vocabulary and reading comprehension are unconstrained skills that continue to develop over a lifetime, decoding is a constrained skill that students are expected to master in early elementary school; thus, the DEC submeasures are grade-independent and identical for grades 1–6. Four benchmark forms were developed for each DEC submeasure to allow assessment at three benchmark periods with different, equivalent material; a fourth alternate form is available to accommodate the potential need for test re- administration due to unforeseen circumstances. In addition, 11 more forms were developed for each DEC submeasure to allow for biweekly progress monitoring within the Burst:Reading Intervention. All benchmark and progress monitoring forms for each DEC submeasure are of equivalent difficulty so educators can track student growth in decoding across a full school year.

Regular Words

The Regular Words (RW) submeasure tests a student's ability to make use of regular, one-to-one letter-sound correspondences in decoding real, monosyllabic words. Regular words are defined as "any word in which each letter represents its respective, most common sound" (Carnine, et al., 2004, p. 46), so all words on RW were required to be composed exclusively of letters that make their most common (highest frequency) sound. Words with initial or final consonant blends assess students' ability to read regular words beyond those containing simple VC and CVC patterns; however, words were constrained to contain either an initial or a final consonant blend so educators could clearly identify specific RW skills that need further instruction. Monosyllabic words allow students to display their letter-sound knowledge without the additional challenge of reading lengthier words. Low- frequency words, with ≤ 4 appearances per million words in texts for grades 1–3 (Zeno, Ivens, Duvvuri, & Millard, 1995), increase the likelihood that a student will rely on decoding strategies (i.e., full or consolidated alphabetic skills) to read them rather than memory (i.e., sight word reading). In summary, words were included on the RW submeasure based on the following criteria (Zeno, et al., 1995):

- Words contain only letters, or graphemes, that make their most common sound.
- Words contain either an initial or a final consonant blend.
- Words are monosyllabic.
- Words have frequency values ≤ 4 for grades 1–3.

We randomly distributed words that fit these criteria among each of the 15 Benchmark and Progress Monitoring forms to ensure each form contained a balance of four words with initial consonant blends and four words with final consonant

blends. The compiled lists were then randomized to create finalized forms of eight words each.

Letter Combinations

The Letter Combinations (LC) submeasure tests a student's ability to decode monosyllabic words containing basic letter combinations, including vowel and consonant digraphs, diphthongs, VCe words, and r- and l-controlled vowels. A letter combination is defined as "a group of consecutive letters that represents a particular sound(s) in the majority of words in which it appears" (Carnine, et al., 2004, p. 151). In order for students to succeed at reading words with letter combinations, they must distinguish between single letters that represent single sounds and groups of letters that represent either single sounds or strings of sounds, which they begin to do in the consolidated alphabetic phase. Letter combinations commonly taught after single letter-sound correspondences include digraphs, diphthongs, VCe combinations, and r- and l-controlled vowels. Digraphs are two- letter combinations that represent a single sound, such as **sh** in **ship** or **oo** in **cook** (unlike the blends in RW, which are two-letter combinations that represent two sounds). Diphthongs are combinations of two vowels that form one gliding sound within a single syllable, such as **ou** in **loud.** VCe words contain the pattern vowel-consonant-e; the final **e** renders the first vowel long. For example, the **i** in the word **fin** is short, but adding an **e** at the end renders the vowel long, producing the word **fine**. R- and L-controlled vowels appear before the letter **r** or **l**; the presence of either of these two subsequent consonants alters the pronunciation of the preceding vowel. For example, compare the vowel sounds in the words **art** and **at**, or those in the words **old** and **odd**; the vowel sounds differ slightly when followed by **r** or **l.**

Each word on the LC submeasure contains only one of these letter combinations; to facilitate identification of specific LC skills in need of further instruction, all other letters in LC words make their most common sound. As in RW, LC words are monosyllabic, allowing students to display their letter- combination knowledge without the additional challenge of reading lengthier words, and low- frequency words, with 6 appearances per million words in texts for grades 1–3 (Zeno, et al., 1995), were chosen to encourage the use of decoding phonics skills and decrease the possibility of sight- word reading. In summary, words were included on the LC submeasure based on the following criteria:

- Words contain one of the following basic letter combinations: vowel or consonant digraph, diphthong, VCe, R- or L-controlled vowel.
- Besides the target letter combination, all other letters in a word make their most common sound.
- Words are monosyllabic.
- Words have frequency values ≤ 6 for grades 1–3 (Zeno, et al., 1995).

We randomly distributed words that fit these criteria among each of the 15 Benchmark and Progress Monitoring forms to ensure each form contained two words with initial consonant digraphs, one word with a final consonant digraph, three words with vowel combinations (digraphs or diphthongs), one word with VCe, and one word with an R- or L-controlled vowel. The compiled lists were then randomized to create finalized forms of eight words each.

Advanced Phonics

The Advanced Phonics (AP) submeasure tests a student's ability to decode mono-, di-, and multisyllabic words that can be broken down into parts, including morphemes and phonograms. Breaking the words on the AP submeasure into orthographic and/or semantic parts can facilitate decoding and comprehension, which is a useful strategy for reading longer words. Words with prefixes, suffixes, inflections, contractions, and compound words contain two or more morphemes. For instance, the word **walked** contains the morphemes **walk** and **ed**; the first morpheme carries the primary semantic content and the second morpheme is a past-tense marker. Recognition of individual morphemes within morphologically complex words can facilitate word reading; some linguists (e.g., Bybee, 1995) theorize that mature readers store high-frequency morphologically complex words as wholes, while they decipher low-frequency morphologically complex words from their component morphemes. Because the letter-sound correspondences in multi-morpheme words can be complex (e.g., in **walked** and **started**, the former **–ed** is pronounced as /t/ while the latter is pronounced as /əd/), breaking these words down into parts can help students read them. Likewise, reading words containing the complex spelling patterns found in phonograms (e.g., **–ood** in **hood**) is facilitated by breaking them down into recognizable chunks rather than single letters. The phonograms that appear in the AP submeasure can be pronounced in at least two ways. For instance, **–ood** can be pronounced as in **food** or as in **hood**: **food** is orthographically regular, while **hood** is an exception to decoding rules. Words with phonograms on the AP submeasure always contain the exception spelling, thus requiring students to recognize the string of letters as a group rather than decoding letter-by-letter. Students are most likely to succeed on the AP submeasure by using consolidated alphabetic skills.

Each word on the AP submeasure contains only one advanced phonics feature to facilitate identification of specific AP skills in need of further instruction. Because students developing advanced phonics skills encounter multisyllabic words in text, and because words containing multiple parts frequently contain multiple syllables, bi- and trisyllabic words appear on the AP submeasure.

Compounds and suffixed words contain two to three syllables, prefixed words contain two syllables, and contracted words contain one or two syllables. Words containing phonograms on the AP submeasure were constrained to monosyllables in order to focus on the exception spelling without additional challenges. Finally, with the exception of contracted words and words containing phonograms, AP words were

constrained to the low-frequency band of ≥ 1 and 10 appearances per million words in texts for grades 1–3 (Zeno, et al., 1995) to encourage the use of advanced phonics decoding skills and decrease the possibility of sight-word reading. Contractions and words with phonograms were not constrained to this range because they tend to be higher frequency and challenging to read in their own right: contractions present a string of consonants that do not normally appear together (e.g., wouldn't), while the phonogram words selected for the assessments contain exception spellings. In summary, words were included on the AP submeasure based on the following criteria (Zeno, et al., 1995):

- Words contain one of the following advanced phonics features: prefix, suffix or final inflection, compound, contraction, phonogram.
- Words contain 1–3 syllables: compound and suffixed words contain 2–3 syllables; prefixed words contain 2 syllables; contracted words contain 1–2 syllables; words with phonograms are monosyllabic.
- Words have frequency values ≥ 1 and ≤ 10 for grades 1–3, except for contracted words and words containing phonograms.

We randomly distributed words that fit these criteria among each of the 15 Benchmark and Progress Monitoring forms to ensure each form contained one disyllabic compound word, one trisyllabic compound word, one disyllabic suffixed word, one trisyllabic suffixed word, one prefixed word, one contraction, and two words containing phonograms. The compiled lists were then randomized to create finalized forms of eight words each.

Irregular Words

The Irregular Words (IW) submeasure tests holistic word recognition skills for high-frequency irregular words, also known as sight words. Irregular words are defined as "words that a student cannot read by applying the letter-sound correspondence knowledge that the student has learned in the reading program" (Carnine, et al., 2004, p. 98). Not all English words can be decoded using letter-sound correspondence rules (Baron & Strawson, 1976) because the long history of English orthography has frozen the spelling of some words in time despite naturally occurring shifts in their pronunciation (Wolman, 2008). For instance, when the word the precedes a word with an initial consonant, it is pronounced with a short or lax /ə/; this pronunciation does not follow regular phonics rules and cannot be analogized, so it must be learned holistically (NICHHD, 2000). Many words that appear with high frequency in text are irregularly spelled words: Johns and Lenski (1997) report that 13 words account for approximately 25 percent of all the words in school-level texts; of these, nine are irregularly spelled. Thus, it is crucial that students master a set of high-frequency irregularly spelled words early in reading development. Automatic recognition of irregularly spelled words builds novice readers' confidence and fluency in reading by freeing up their cognitive resources to focus on words that can be decoded and have not yet been internalized holistically. Words on the IW submeasure were selected from a list of 220 high-frequency words found in children's literature

(Dolch, 1948) that is widely used with beginning readers due to the high utility of automatically recognizing these words. The list includes mono-, di-, and trisyllabic words, and the words are categorized by grade level (pre- primer, primer, first, second, and third grades). Some words on the Dolch List of Basic Sight Vocabulary are completely regular (e.g., **him**), while others are partly irregular (e.g., **could**) or completely irregular (e.g., **the**). To encourage sight-word recognition rather than decoding, the AP submeasure included only partly or completely irregular words. In summary, words were included on the IW submeasure based on the following criteria:

- Words appear on the Dolch List of Basic Sight Vocabulary.
- Words are irregular.

We randomly distributed words that fit these criteria among each of the 15 Benchmark and Progress Monitoring forms to ensure each form contained two words from the pre-primer and primer lists, four words from the grade 1 list, and two words from the lists for grades 2 and 3. Grade 1 words were emphasized because sight word recognition is a critical skill for grade 1. The compiled list for each form was then randomized to create the finalized version.

Burst Vocabulary

Theory

According to the Simple View of Reading, reading comprehension is the product of decoding skill and language comprehension (Gough & Tunmer, 1986; Catts & Weismer, 2006). Unpacking the surface code of a text is just one part of the process: readers must also attend to vocabulary and draw upon semantic knowledge to construct meaning. Oral vocabulary size in preschool and kindergarten predicts reading comprehension and reading achievement in grades 1–3 (Muter, Hulme, Snowling, & Stevenson, 2004; Roth, Speece, & Cooper, 2002; Hart & Risley, 2003), highlighting the importance of early language development in later reading development. Distressingly, studies show that students from economically disadvantaged families, as well as English language learners, are more likely to experience oral vocabulary deficiencies that may severely limit reading development (Hart & Risley, 2003; Nation, 2001). Hart and Risley (2003) revealed a serious discrepancy in vocabulary size aligned with the poverty gap in—children from low, middle, and high socioeconomic status (SES) families. They found that the oral vocabulary used by both parents and children of low SES families is much narrower in range than that found in higher SES families, and that by age three, children from low SES families are exposed to 30 million fewer words than children from high SES families. Word poverty in socioeconomically disadvantaged students is a result of what Stanovich calls the "Matthew effect," based on the Gospel adage that the rich get richer as the poor get poorer: "The very children who are reading well and who

have good vocabularies will read more, learn more word meanings, and hence read even better. Children with inadequate vocabularies—who read slowly and without enjoyment—read less, and as a result have slower development of vocabulary knowledge, which inhibits further growth in reading ability" (1986, p. 381). A reader who enters school without a strong oral vocabulary base can only extract a partial semantic representation from text, with vague, incomplete comprehension. This reader will not progress to more difficult texts and, consequently, will fall behind vocabulary-rich peers.

The crucial nature of early vocabulary development motivated the creation of a receptive vocabulary screener for the Burst:Reading assessments. This screener, the **Vocabulary (VOC)** measure, identifies students who exhibit major deficiencies in their basic oral vocabularies that, if left unresolved, may potentially disrupt the overall development of reading comprehension.

Design

The VOC measure assesses the breadth of students' basic receptive vocabulary in kindergarten through grade 3. VOC is a picture-matching assessment in which a student is shown a set of images and asked to point to the image that exemplifies a word the test administrator read aloud. The picture-matching format isolates a student's receptive vocabulary knowledge from both reading and speaking skills, which may be confounded with vocabulary knowledge in the case of either a word-matching or an expressive vocabulary task. Choosing the correct image among distractors indicates a student's familiarity with the word's meaning.

Four Benchmark forms were developed for each grade level in kindergarten through grade 3 to allow assessment at three benchmark periods with different, equivalent material; a fourth form is available to accommodate the potential need for test re-administration due to unforeseen circumstances. In addition, six more forms were developed for each grade to allow for monthly progress monitoring within the Burst:Reading Intervention; progress monitoring for VOC is recommended on a monthly rather than a biweekly basis because vocabulary is an unconstrained skill that develops at a slower pace than a constrained skill like decoding (Paris, 2005). Each Benchmark and Progress Monitoring form consists of five pages with three words, three target images, and three distractor images on each page, for a total of 15 words and 30 images per form. The test administrator reads each word aloud and asks the student to match the word to its exemplifying image. Student responses are scored as correct, incorrect, or no response if the student does not provide an answer within 5 seconds. The maximum score on VOC is 15.

Grade-specific word frequency values represent the number of times a word appears per million words of grade-level text; for instance, a word with a frequency value of 80 in grade 2 appears 80 times per million words in texts targeted towards grade

2 students. Because the VOC measure is intended to screen students for basic rather than academic vocabulary, high-frequency, grade- appropriate words were selected (Zeno, et al., 1995). The frequency bands chosen target Tier One words, which Beck, McKeown, and Kucan define as "the most basic words—**warm**, **dog**, **tired**, **run**, **talk**, **party**, **look**, **swim**, and so on. These are words that typically appear in oral conversations, so children are exposed to them at a high frequency from a very early age. This high exposure means that children become familiar with this set of words readily, and so these Tier One words rarely require instructional attention to their meanings in school" (2013, p. 9). A student who does not know the meaning of Tier One words is unlikely to comprehend beginning reader texts, so additional vocabulary instruction is necessary. Zeno, et al. (1995), provided examples of words found in different frequency bands, and the frequency bands most representative of Tier One words for each grade were chosen for the VOC measure. Grade-specific word frequency was not provided for kindergarten, so higher frequency grade 1 words were chosen for the kindergarten forms. In summary, words were included on the VOC measure based on the following criteria:

- Words fall into the following frequency ranges for each grade:
  - Kindergarten: 100–1000 for grade 1
  - Grade 1: 30–100 for grade 1
  - Grade 2: 30–100 for grade 2
  - Grade 3: 30–100 for grade 3
- Words are nouns, verbs, and adjectives with concrete and context-independent meanings that can be clearly illustrated.
- Words do not have homographs.

The picture-matching format of the VOC measure requires words that can easily and clearly translate into images. We chose nouns, verbs, and adjectives with concrete meanings for the measure, while adverbs, function words, and abstract nouns, verbs, and adjectives were excluded to avoid ambiguous or abstract images. In addition, because words are read aloud to students, homographs (words with the same spelling but different pronunciations and meanings, e.g., **tear**) were excluded from the VOC measure to avoid confusion about the word's pronunciation.

We designed simple, clear target and distracter images for the vocabulary assessment to avoid ambiguity. In addition, culturally specific references (e.g., illustrating the word **early** with a rooster or **promise** with a pinky swear) were avoided to ensure an unbiased assessment for students with a variety of cultural backgrounds. Written words and punctuation were excluded from images to ensure that no reading is required to succeed on the assessment. Other symbols such as arrows were excluded to avoid producing images that illustrated a word using contrast (e.g., illustrating the word **short** with an arrow pointing to a child standing next to an adult) or a part/whole relationship (e.g., illustrating the word **shoulder**

by pointing to the body part on a person), as such images could cause confusion. Images with people in them include a balance of genders and ethnicities. The style and content of distracter images look purposefully similar to target images so they do not look like obviously incorrect answers. In summary, we designed the accompanying images on the VOC measure based on the following criteria:

- Images are simple, black-and-white line drawings with as few elements as necessary to depict the word's meaning.
- Images are interpretable using general, culturally independent knowledge.
- Images exclude written words, punctuation, arrows, or any other symbols.
- Images include a representative balance of genders and ethnicities.

Words, target images, and distracter images were assembled into forms to ensure each form contained nine nouns, three verbs, and three adjectives. Nouns are more heavily represented in the VOC assessment because they are more prevalent in both speech and writing (Hudson, 1994), and they tend to work best as images. Words, target images, and distracter images were arranged to ensure each page includes only one possible correct answer with no ambiguity between target and distracter images. Target images were never reused as distracter images.

Burst Comprehension Skills

Theory

The Rand Reading Study Group (RRSG) defined reading comprehension as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (Snow, 2002, xiii). Beyond fluently reading the words in a text and extracting meaning, the reader must comprehend the text's linguistic structures (syntax) as well as its words and phrases (semantics) and must situate this information within her own knowledge base, constructing meaning. Meaning is constructed at multiple levels, starting from understanding the propositions explicitly stated in the text, moving to a deeper understanding of the text's connection with other sources of knowledge (Kintsch & Van Dijk, 1978). Thus, reading comprehension is an iterative process of constructing a mental representation (also known as a schema) of a text and revising that mental representation as more text information is extracted. A reader's final mental representation of the text may or may not be similar to the mental representation of the text's author or of other readers.

Because these representations exist within the mind, defining and measuring what constitutes comprehension of a text is a complex task.

A useful distinction in measuring reading comprehension has been made between the literal level of comprehension and the inferential level (Carnine, et al., 2004). Literal comprehension is based on information explicitly stated in the text, while inferential comprehension is based on implicit assumptions made while reading and relies on the reader's knowledge to fill informational gaps in the text. The Comprehension Skills (CS) measure screens for difficulties in reading comprehension at both the literal and the inferential levels. Additional theoretical background about literal and inferential comprehension is provided in the following sections describing each component of the CS measure.

Design

The CS assessment consists of fiction and nonfiction texts and accompanying comprehension assessments for grades 1–32. The assessment is administered individually and begins with the administrator selecting either a fiction or nonfiction text. The student is then instructed to "read to yourself," implicitly encouraging the student to read silently so that cognitive resources can be maximally focused on interpreting meaning, although out-loud reading is permitted. After the student finishes reading, the text is removed and the student is asked to retell (for fiction texts) or recall (for nonfiction texts) its content for the administrator. The text is then returned to the student, and the administrator asks the student a series of five literal questions that refer to what is explicitly stated in the text. The student is allowed to refer to the text so that literal comprehension skill can be assessed in isolation from individual differences in memory. Following the literal questions, the examiner asks the student a series of five inferential questions and again allows the student to refer back to the text when answering.

Eight Benchmark forms (four fiction and four nonfiction) were developed for each grade level in grades 1–3 to allow assessment at three benchmark periods with different, equivalent material; two alternate forms (one fiction and one nonfiction) are available to accommodate the potential need for test re-administration due to unforeseen circumstances. In addition, four more forms (two fiction, two nonfiction) were developed for grade 1 and six more forms (three fiction and three nonfiction) were developed for both grades 2 and 3 to allow for monthly progress monitoring; fewer grade 1 forms are available because screening for reading comprehension difficulties typically starts later in grade 1 (although a Beginning-of-Year benchmark form is available for testing more advanced grade 1 students). Progress monitoring for CS within the Burst:Reading Intervention is recommended on a monthly rather than a biweekly basis because, like vocabulary, reading comprehension is

2 CS is available in grade 3 for users of the Burst: Reading Assessment only; DIBELS: Next DAZE is used to screen students for intervention need in comprehension and to guide intervention grouping and placement in grade 3 for Burst: Reading Intervention users.

an unconstrained skill that develops at a slower pace than a constrained skill like decoding (Paris, 2005).

Texts

The CS measure consists of fiction and nonfiction reading texts developed for grades 1–3. Because motivation and background knowledge are important factors in reading comprehension (Snow, 2002), we developed texts about familiar topics that interest young students. Fiction texts pertain to everyday topics such as cleaning up or going to the dentist, and nonfiction texts focus on science, social studies, and technical topics typically encountered in early elementary school, such as the first Thanksgiving and the four seasons. Because clear and coherent text structure supports reading comprehension (Pearson & Dole, 1987), the fiction texts were written as narratives that include clear beginning, middle, and end sections, and the nonfiction texts include compare-and-contrast, cause- and-effect, and steps-in-a-sequence formats. Texts were constrained to 140–155 words in length to minimize testing time. Texts were also written to fall within the Spache-Revised (Spache, 1974) grade-level readability ranges listed earlier. The Spache-Revised readability formula determines the grade level of a text according to average sentence length and number of unique, unfamiliar words based on a list of words generally unfamiliar to students below grade 4. The Spache-Revised formula was selected for the CS assessment because it is particularly useful in leveling grades 1–3 texts (Stenner, Burdick, Sanford, & Burdick, 2007; Begeny & Greene, 2014). Academic and domain-specific vocabulary words were included in texts to allow for vocabulary-in-context questions; we checked the grade-appropriateness of these vocabulary words against research-based lists of word difficulty (Biemiller, 2010; Mogilner & Mogilner, 2006; Taylor, Frackenpohl, & White, 1989). In summary, texts were written for the CS measure based on the following criteria:

- Texts pertain to familiar topics that interest young students and require general rather than specific background knowledge to understand.
- Fiction texts are narratives that include beginning, middle, and end sections.
- Nonfiction texts present science, social studies, and technical topics in formats including compare and contrast, cause and effect, and steps in a sequence.
- Texts are 140–155 words long.
- Texts fall within the following Spache-Revised readability ranges for each grade level:
  - Grade 1: 1.8–1.9
  - Grade 2: 2.2–2.5
  - Grade 3: 2.8–3.3
- Texts contain grade-appropriate vocabulary, including academic and domain-specific words.

Retell/Recall

After reading is complete, the text is removed and the student is asked to retell (for fiction texts) or recall (for nonfiction texts) what he or she just read. Oral retell or recall tests a student's literal comprehension of a text and requires the student to assimilate, organize, and reconstruct information gained from the overall text (Morrow, 1988). Retell/recall is assessed orally for two reasons. First, oral, as opposed to written, retell or recall prevents confounding comprehension skills with writing ability Second, because oral-language abilities are an important component of reading comprehension (Storch & Whitehurst, 2002), the retell/recall portion of the assessment gives the examiner information on the student's ability to organize and express thoughts verbally.

Students get explicit instructions before retelling/recalling a text, to clarify how they should respond. In retelling fiction texts, students are asked to retell the beginning, middle, and ending sections of the story. When scoring a student's retell of a fiction text, the order in which the sections are retold is important, as there are causal relationships between the beginning, middle, and ending sections of a story. A student who retells a story out of order may not understand these causal connections or may have trouble reconstructing the sections in a meaningful order. In recalling nonfiction texts, students are asked to provide the main idea and at least two key details from the text. When scoring a student's recall of a nonfiction text, the ability to distinguish between the main idea and key details of the text is important, and provision of the main idea is weighted more heavily than key details. Order is not a factor in recall because not all nonfiction texts in the assessment contain temporally ordered events. Responses are scored as Insufficient (0 points), Basic (2 points), or Good (4 points) based on the following rubrics, which are accompanied by text-specific exemplar responses to aid scoring:

Retell rubric (Fiction texts):

- Good: Student provides three sections in a meaningful order.
- Basic: Student provides two sections in a meaningful order or three sections out of order.
- Insufficient: Student provides two or fewer sections out of order.

Recall rubric (Nonfiction texts):

- Good: Student provides the main idea and two or more key details.
- Basic: Student provides two or more key details or the main idea and one key detail.
- Insufficient: Any other result.

The maximum retell/recall score is 4 points.

Literal Questions

After retelling or recalling is complete, the text is returned to the student to help answer five literal questions. Literal questions target the "5 W's and 1 H": who, what, when, where, why, and how, assessing a student's understanding of basic facts in the text. Answering literal questions requires comprehension at the sentence level, as it is possible for students to use key words and syntactic information to locate the appropriate response to the question within the text (Snow, 2002). Literal questions require explicit comprehension of individual or adjacent propositions but do not require a student to integrate multiple propositions across the text or to supplement explicit information from the text with knowledge from other sources. In fact, we specifically wrote text-dependent questions that students could not answer correctly through background knowledge alone. To help students locate the correct answer, a literal question uses phrasing similar to the sentence the answer is drawn from. In summary, literal questions were written for the CS measure based on the following criteria:

- Questions are who, what, when, where, why and how questions.
- Questions rely on one or two adjacent propositions.
- Questions are text dependent and do not require background knowledge.
- Question syntax mimics the syntax found in the text.

Responses to literal questions are scored dichotomously as follows, and text-specific exemplar responses are provided to aid scoring:

- Zero points are awarded for no response or an incorrect response.
- One point is awarded for a correct response.

The maximum Literal Questions score is 5 points.

Inferential Questions

Following the Literal section, the administrator asks the student five inferential questions; the student may still refer back to the text. Inferential questions target information that the text implicitly contains rather than explicitly states. These questions focus on describing, explaining, comparing, contrasting, predicting, defining words, and connecting text content to background knowledge. Answering inferential questions requires a student determine the text's implicit meaning based on two or more nonadjacent text propositions and possibly prior knowledge (McKoon & Ratcliff, 1992). We specifically wrote text-dependent inferential questions that students could not answer through background knowledge alone. In summary, inferential questions were written for the CS measure based on the following criteria:

- Questions focus on describing, explaining, comparing, contrasting, predicting, defining words, and connecting text content to background knowledge.
- Questions rely on information from multiple text propositions or on information from the text and background knowledge.

- Questions are text dependent, even if additional background knowledge is required.

Responses to inferential questions are scored dichotomously as follows, and text-specific exemplar responses are provided to aid scoring:

- Zero points are awarded for no response or an incorrect response.
- One point is awarded for a correct response.

The maximum Inferential Questions score is 5 points.

Overview of Research on Burst:Reading Assessments

This chapter describes data and analyses from research studies examining the Burst:Reading Benchmark assessments3, including the purpose of each study, how participants were recruited, demographics for the participants, experimental design, and the descriptive statistics calculated for each study.

Study A: Initial Field Study

Purpose: Study A obtained data to conduct item analysis, examine reliability and validity, and establish empirical cut points for the Burst Reading Assessment Benchmark Forms.

Recruitment: We used two different strategies to recruit districts and schools. First, Amplify Account Management and Sales teams were asked to contact existing users of mCLASS®:DIBELS Next® (Dynamic Indicators of Basic Early Literacy Skills – Next Edition; Good, et al., 2013) customers with information about the field study. This approach yielded moderate success, so another strategy was adopted in which existing mCLASS:DIBELS Next customers received an informational flyer about the study via email listserv. Using these recruitment methods, seven schools successfully enrolled in the study. All seven schools participated in the study during the middle of year (MOY) benchmark period. During end of year (EOY), five of these seven schools participated in the study; two schools cited scheduling conflicts as a reason for withdrawal.

Participants: This field study was conducted around the 2013–2014 middle- (December through February) and end-of-year (April through June) benchmark administration periods. In total, 1,116 students in kindergarten through grade 3 from seven schools in seven districts in six states participated in the study. Across both testing periods, 321 kindergarten students, 303 grade 1 students, 267 grade 2 students, and 225 grade 3 students participated in the study. As noted above, there was some school-level attrition between MOY and EOY; in addition, when students transferred out of a school after the MOY testing period, schools were instructed to add new students to the study to replace those who left. A total of 441 students

3 Progress Monitoring forms for the Burst:Reading assessments were not included in the empirical research studies described in this chapter, but they were designed to be equivalent to the Benchmark assessments and were tested informally in a pilot study that was conducted by members of the Amplify research team with students in kindergarten through grade 3 in April 2014. The results of the pilot study were used to refine items on the Progress Monitoring forms as well as general administration and scoring procedures for both Benchmarking and Progress Monitoring with the Burst:Reading assessments.

participated in the study at both MOY and EOY; 536 students participated only in the MOY testing period and 139 students participated only in the EOY testing period.

The Burst:Reading assessments were administered to students by 18 district employees who participated in a half-day training in administering the assessments.

Demographic Information: Participants in this field study included students from across the United States, including the following geographic divisions: Mountain, Middle Atlantic, South Atlantic, and West South Central (US Census Bureau, n.d.). The sample was comprised of 31 percent male, 37 percent female, 32 percent unspecified gender; 23 percent White, 26 percent Black, 5 percent Hispanic, 4 percent American Indian, 0.4 percent Asian, 0.6 percent mixed race, and 41 percent unspecified race. Appendix 1 provides a comparison of the field study sample, the mCLASS national database of Joule, and all public schools in the U.S.

Testers were district employees: 28 percent were substitute teachers, 17 percent were reading specialists or coaches, and 55 percent held another position, including Response-to-Intervention Specialist, Paraeducator or Aide, and Resource Room Specialist. All testers worked with kindergarten through grade 4 students and had an average of 10 years of experience as educators (range: 0–40; SD: 9.26). They were 94 percent female, 6 percent male; 78 percent White, 17 percent Black, and 5 percent unspecified race. In terms of education, 11 percent of testers had an associate's degree or some college credit, 44 percent had a bachelor's degree, 17 percent had a master's degree, 11 percent had a degree beyond master's, and 17 percent had some other type of degree or certification.

Experimental Design: Ideally, field-study results should reflect the full spectrum of student ability, and not just high or low performers. Therefore, the study asked participating educators to select equal proportions of students according to performance level, based on MOY DIBELS Next results. For example, a site that contributes 15 students per grade would select five students from each performance category at MOY: "Well Below Benchmark," "Below Benchmark," and "At or Above Benchmark."

Each student was then administered the Benchmark Burst:Reading assessments. The assessments were administered using paper-and-pencil forms to allow for revisions after the study but before implementing the assessments in software. All instructions and procedures for administering each assessment were identical to those used in the Burst:Reading Assessment software. Students were tested individually in a pull-out setting.

Each participating student was tested with each of the four Benchmark forms for each Burst:Reading Assessment they could take, according to their grade; kindergarteners were tested with VOC only, while students from grades 1–3 were tested with VOC, DEC, and CS. Two forms were administered at MOY, and the other two forms were administered at EOY; the order of form administration was counterbalanced both across and within Benchmark periods to mitigate testing order effects, and schools were randomly assigned to pairs of forms. At each benchmark

period, testing in grades 1–3 was split into two sessions to avoid testing fatigue; total testing time in grades 1–3 was approximately 50 minutes per Benchmark period, while total testing time in kindergarten was approximately 5 minutes per Benchmark period.

Test administrators entered student data they collected themselves into an online data-entry tool for delivery to Amplify after both the MOY and EOY testing periods; testers were permitted to enter additional comments they had about assessment items. After the MOY testing period ended, test administrators completed a survey to provide feedback on the measures, the test administration procedures, and the study procedures. Results of this survey, along with study data, further refined the measures prior to their final implementation in software form.

At EOY, students were tested with an additional external, multicomponent measure of reading, STAR Early Literacy (SEL; Renaissance Learning, 2001), in order to collect data to support validity and cut point analyses. This computer-adaptive assessment tests students in 41 different skill sets in three key domains (Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, and Numbers and Operations) and ten sub-domains: alphabetic principle, concept of a word, visual discrimination, phonemic awareness, phonics, structural analysis, vocabulary, sentence-level comprehension, paragraph-level comprehension, and early numeracy (Renaissance Learning, 2014). SEL is administered on a computer with headphones, and students respond to aurally-presented multiple-choice questions using a mouse or keyboard. SEL takes approximately 10–15 minutes for students to complete. Administration of SEL was overseen either by testers who had administered the Burst:Reading assessments to students or by school information technology personnel.

Renaissance Learning personnel trained SEL administrators prior to administering the assessment.

Descriptive Statistics: Tables 1 through 11 provide descriptive statistics for student performance on each Burst:Reading Assessment Benchmark form, including mean, median, standard deviation (SD), minimum score, maximum score, and number of students tested for each of the forms. Results for DEC are provided for each form both across grades and by grade because the same forms are used across grades; results for each VOC and CS form are provided by grade because the forms for these measures are grade specific.

Decode

Descriptive statistics are provided separately for each of the four DEC Benchmark submeasures. Because the DEC submeasures are grade-independent (i.e., the same four Benchmark forms are used across grades), descriptive statistics are reported both for the entire sample and by grade. In grades 2 and 3, ceiling effects

are observed for many DEC forms; this is an unsurprising result because students learning to read are typically able to demonstrate proficiency in reading words with the spelling patterns targeted by DEC by the end of grade 1. The maximum score a student can achieve on the DEC submeasure is 8.

Regular Words

Descriptive statistics for each RW form are provided for the general sample in Table 1 and for each grade in Table 2.

Table 1. Study A – Descriptive Statistics for DEC RW (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	608	5.85	6	2.28	0	8
All	B	458	5.85	6	2.23	0	8
All	C	491	5.85	6	2.06	0	8
All	D	461	6.15	7	2.18	0	8

Table 2. Study A – Descriptive Statistics for DEC RW (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	207	4.80	5	2.46	0	8
1	B	157	5.29	6	2.47	0	8
1	C	211	5.27	6	2.35	0	8
1	D	163	5.67	7	2.57	0	8
2	A	203	6.16	7	2.02	0	8
2	B	169	6.08	7	2.06	0	8
2	C	144	5.94	6	1.82	1	8
2	D	168	5.98	6	2.06	0	8
3	A	198	6.61	7	1.91	0	8
3	B	132	6.23	7	2.02	0	8
3	C	136	6.65	7	1.48	0	8
3	D	130	6.97	7.5	1.49	0	8

Letter Combinations

Descriptive statistics for each LC form are provided for the general sample in Table 3 and for each grade in Table 4.

Table 3. Study A – Descriptive Statistics for DEC LC (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	605	4.95	5	2.36	0	8
All	B	457	4.56	5	2.57	0	8
All	C	489	5.40	6	2.24	0	8
All	D	462	5.26	6	2.29	0	8

Table 4. Study A – Descriptive Statistics for DEC LC (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	208	3.85	4	2.39	0	8
1	B	155	3.61	3	2.54	0	8
1	C	207	4.54	5	2.45	0	8
1	D	165	4.55	5	2.46	0	8
2	A	200	5.34	5	2.09	0	8
2	B	171	4.85	5	2.49	0	8
2	C	144	5.53	6	1.89	1	8
2	D	168	5.26	6	2.23	0	8
3	A	197	5.71	6	2.14	0	8
3	B	131	5.32	6	2.35	0	8
3	C	138	6.56	7	1.63	0	8
3	D	129	6.18	7	1.77	0	8

Advanced Phonics

Descriptive statistics for each AP form are provided for the general sample in Table 5 and for each grade in Table 6.

Table 5. Study A – Descriptive Statistics for DEC AP (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	605	5.79	7	2.64	0	8
All	B	458	5.41	6	2.49	0	8
All	C	492	5.81	7	2.48	0	8
All	D	461	5.15	6	2.68	0	8

Table 6. Study A – Descriptive Statistics for DEC AP (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	206	3.78	4	2.92	0	8
1	B	157	4.19	5	2.74	0	8
1	C	209	4.29	4	2.72	0	8
1	D	164	3.50	3	2.94	0	8
2	A	200	6.63	7	1.80	0	8
2	B	170	5.75	6	2.07	0	8
2	C	146	6.40	7	1.62	1	8
2	D	168	5.48	6	2.25	0	8
3	A	199	7.03	8	1.64	0	8
3	B	131	6.42	7	2.05	0	8
3	C	137	7.48	8	1.13	0	8
3	D	129	6.81	7	1.37	0	8

Irregular Words

Descriptive statistics for each IW form are provided for the general sample in Table 7 and for each grade in Table 8.

Table 7. Study A – Descriptive Statistics for DEC IW (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	605	6.92	8	1.78	0	8
All	B	461	6.90	8	1.84	0	8
All	C	490	6.74	7	1.73	0	8
All	D	459	7.19	8	1.48	0	8

Table 8. Study A – Descriptive Statistics for DEC IW (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	208	5.72	6	2.10	0	8
1	B	157	6.14	7	2.17	0	8
1	C	207	6.02	7	2.20	0	8
1	D	164	6.67	7	1.80	0	8
2	A	200	7.47	8	1.14	3	8
2	B	172	7.20	8	1.54	0	8
2	C	144	7.14	7	1.01	2	8
2	D	167	7.29	8	1.34	0	8
3	A	197	7.61	8	1.23	0	8
3	B	132	7.42	8	1.46	2	8
3	C	139	7.40	8	0.98	0	8
3	D	128	7.74	8	0.84	0	8

Vocabulary

Descriptive statistics are provided for the grade K–3 VOC Benchmark forms in Table 9; recall that the VOC forms contain grade-specific content. A ceiling effect was observed for all VOC forms, which aligns to its purpose as a screener of Tier One vocabulary (Beck, McKeown, & Kucan, 2013), which most students acquire without formal instruction in school. The maximum score a student can achieve on the VOC measure is 15.

Table 9. Study A – Descriptive Statistics for VOC (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
K	A	240	13.88	14	1.26	10	15
K	B	238	14.06	15	1.37	8	15
K	C	241	13.67	14	1.30	8	15
K	D	239	13.28	14	1.56	5	15
1	A	211	13.90	15	2.43	0	15
1	B	206	13.79	15	2.36	1	15
1	C	209	13.65	14	2.55	0	15
1	D	194	13.09	14	3.17	0	15
2	A	198	14.06	14	1.19	8	15
2	B	172	14.25	15	1.67	5	15
2	C	146	13.93	14	1.40	8	15
2	D	168	14.07	15	1.46	5	15
3	A	197	13.54	14	2.20	1	15
3	B	132	13.40	14	2.55	0	15
3	C	138	14.38	15	1.61	0	15
3	D	128	14.73	15	1.39	0	15

Comprehension

Descriptive statistics are provided for the grade K–3 Fiction and Nonfiction CS Benchmark forms in Tables 10 and 11; recall that the CS forms contain grade-specific content. The descriptive statistics provided are based on CS total score (see Appendix 2 for the determination of CS total score formula). While CS Fiction and Nonfiction texts have equivalent readability (a quantitative measure of text complexity), these results suggest students found some of the Fiction texts easier than the Nonfiction texts. This result may be due to exposure: younger students are more commonly exposed to narrative than expository texts and their structures (Duke, 2000; Yopp & Yopp, 2000; Hoffman, et al., 1994), and experience may be an important factor in genre-specific comprehension (Kamberelis, 1998; Kamil & Lane, 1997). Indeed, as grade level and presumably experience with nonfiction text increases, student performance on the nonfiction texts improves. The maximum score a student can achieve on the CS Fiction and Nonfiction measures is 14.

Table 10. Study A – Descriptive Statistics for CS Fiction (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	174	7.76	8	4.36	0	14
1	B	164	8.12	8	4.35	0	14
1	C	124	10.39	12	3.77	0	14
1	D	149	7.97	9	4.56	0	14
2	A	197	9.62	11	4.06	0	14
2	B	164	8.99	9	4.08	0	14
2	C	142	8.46	9	4.02	0	14
2	D	164	9.85	10	3.47	0	14
3	A	198	10.56	12	3.83	0	14
3	B	127	10.68	12	4.12	0	14
3	C	135	11.27	12	3.16	0	14
3	D	129	8.82	9	3.95	0	14

Table 11. Study A – Descriptive Statistics for CS Nonfiction (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	136	7.95	8	4.02	0	14
1	B	157	8.83	10	4.28	0	14
1	C	113	9.41	10	4.14	0	14
1	D	142	6.11	6	4.27	0	14
2	A	192	8.84	9	3.70	0	14
2	B	166	8.28	9	3.65	1	14
2	C	133	8.56	9	3.53	0	14
2	D	167	7.68	7	3.98	0	14
3	A	188	9.48	10	3.84	0	14
3	B	132	9.27	10	4.24	0	14
3	C	133	9.61	10	3.63	0	14
3	D	133	7.27	7	3.96	0	14

Study B: Inter-Rater Reliability Study

Purpose: Study B examines the inter-rater reliability (IRR) of the Burst:Reading assessments.

Recruitment: Schools from the Study A sample were invited to participate in an additional IRR study at the end of the 2013–2014 school year. One school agreed to participate (School 1), but due to scheduling conflicts could not provide enough students to achieve the target sample size in grades 2 and 3. Thus, an additional school (School 2) was recruited in the summer after the 2013–2014 school year to provide additional students in grades 2 and 3.

Participants: This field study was conducted around the 2013–2014 end-of-year benchmark administration period (April through June) and in July 2014 during a summer school session. In total, 84 students in kindergarten through grade 3 from two schools in two districts in two states participated in the study. There were 29 kindergarten students, 20 grade 1 students, 12 grade 2 students, and 23 grade 3 students.

In School 1, the primary test administrators were district employees who had undergone a half-day training in administering the assessments; this group was a subset of the testers who participated in Study A. The secondary, shadow scorers were Amplify employees who were Burst trainers. In School 2, both the primary test administrator and the secondary shadow scorer were Amplify employees: one was a Burst trainer and the other was a member of the research team.

Demographic information: Participants in this field study included students from the East South Central and Mountain regions of the United States (US Census Bureau, n.d.). The sample was comprised of 50 percent male, 17 percent female, 33 percent unspecified gender; 50 percent White, 14 percent Black, 2 percent Hispanic, and 33 percent unspecified race; 6 percent were in special education program; 51 percent were eligible for free or reduced priced lunch; 14 percent were not eligible for free or reduced priced lunch and 35 percent were not specified in free or reduced lunch category. Among the sample, 27 percent were far below benchmark on DIBELS Next composite score, 21 percent were below benchmark, 26 percent were above benchmark, 25 percent were unspecified on DIBELS Next composite score.

Experimental Design: Pairs of testers administered the Benchmark Burst:Reading assessments to each student. One tester served as the primary test administrator, reading all of the directions and questions to the student while scoring performance as usual, while the other tester sat in the background and shadow-scored along with the student's performance. The assessments were administered using paper-and-pencil forms, and students were tested individually in a pull-out setting with two forms of each grade-appropriate Burst:Reading Assessment. Note that not all forms for all grades were tested in the IRR study. School 1 had been randomly assigned to administer the EOY and ALT forms to students at the EOY testing period; the BOY and MOY forms were therefore administered in School 2 during the summer to include a broader range of assessment forms in the study.

Descriptive Statistics: Tables 12 through 22 provide descriptive statistics for each Burst:Reading Assessment Benchmark form, including mean, median, standard deviation (SD), minimum score, maximum score, and number of students tested. Results for DEC are provided for each form both across grades and by grade because the same forms are used across grades; results for each VOC and CS form are provided by grade because the forms for these measures are grade specific.

Patterns of results observed in Study B are similar to those described in Study A.

Table 12. Study B: Descriptive Statistics for DEC RW (All Grades).

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	16	6.16	7	1.76	1	8
All	B	13	5.58	5	2.16	1	8
All	C	37	6.19	7	1.80	0	8
All	D	37	6.45	7	1.86	0	8

Table 13. Study B: Descriptive Statistics for DEC RW (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	18	5.79	6	2.10	0	8
1	D	18	6.15	7	2.22	0	8
2	A	5	5.20	6	2.35	1	7
2	B	5	5.20	5	2.53	1	8
2	C	7	6.45	7	1.57	4	8
2	D	7	6.27	7	1.85	3	8
3	A	11	6.59	7	1.26	4	8
3	B	8	5.81	5	1.94	2	8
3	C	12	6.63	7	1.34	3	8
3	D	12	6.95	7	1.18	4	8

Table 14. Study B: Descriptive Statistics for DEC LC (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	16	4.59	4.50	2.39	0	8
All	B	12	4.75	5.50	2.44	0	8
All	C	37	5.28	6.00	2.02	0	8
All	D	37	5.05	5.50	1.92	0	8

Table 15. Study B: Descriptive Statistics for DEC LC (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	18	4.19	5	2.08	0	8
1	D	18	4.50	5	2.02	0	8
2	A	5	3.50	4	2.42	0	7
2	B	5	3.80	3	2.82	0	8
2	C	7	5.64	7	1.96	3	8
2	D	7	5.27	5	1.95	2	8
3	A	11	5.09	6	2.27	2	8
3	B	7	5.43	6	1.95	2	8
3	C	12	6.63	7	0.68	6	8
3	D	12	5.68	6	1.60	3	8

Table 16. Study B: Descriptive Statistics for DEC AP (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	16	6.53	7	2.17	1	8
All	B	11	5.64	6	1.22	3	7
All	C	37	6.19	7	2.42	1	8
All	D	37	5.02	6	2.50	0	8

Table 17. Study B: Descriptive Statistics for DEC AP (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	18	4.93	6.00	2.84	1	8
1	D	18	3.46	4.00	2.45	0	7
2	A	5	6.00	7.00	2.75	1	8
2	B	4	5.38	5.00	1.06	4	7
2	C	7	7.00	7.00	1.18	5	8
2	D	7	5.73	6.00	1.62	3	8
3	A	11	6.77	7.00	1.88	1	8
3	B	7	5.79	6.00	1.31	3	7
3	C	12	7.45	8.00	1.23	3	8
3	D	12	6.83	7.50	1.38	4	8

Table 18. Study B: Descriptive Statistics for DEC IW (All Grades)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
All	A	16	7.72	8	0.77	5	8
All	B	13	7.08	8	1.87	1	8
All	C	37	7.00	7	0.90	4	8
All	D	37	7.46	8	0.91	4	8

Table 19. Study B: Descriptive Statistics for DEC IW (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	18	6.93	7	1.11	4	8
1	D	18	7.32	8	0.95	5	8
2	A	5	7.50	8	1.08	5	8
2	B	5	6.50	8	2.68	1	8
2	C	7	7.09	7	0.70	6	8
2	D	7	7.18	8	1.25	4	8
3	A	11	7.82	8	0.59	6	8
3	B	8	7.44	8	1.09	5	8
3	C	12	7.05	7	0.69	6	8
3	D	12	7.83	8	0.38	7	8

Table 20. Study B: Descriptive Statistics for VOC (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
K	A	0	n/a	n/a	n/a	n/a	n/a
K	B	0	n/a	n/a	n/a	n/a	n/a
K	C	28	12.94	13.50	1.57	10	15
K	D	28	13.27	13.00	1.10	11	15
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	18	14.19	15.00	1.11	11	15
1	D	18	14.00	14.00	0.85	12	15
2	A	5	14.20	14.00	0.79	13	15
2	B	5	14.80	15.00	0.42	14	15
2	C	7	14.67	15.00	0.65	13	15
2	D	7	13.70	14.00	1.34	11	15
3	A	11	14.64	15.00	0.49	14	15
3	B	8	14.50	15.00	0.73	13	15
3	C	12	14.74	15.00	0.73	12	15
3	D	12	15.00	15.00	0.00	15	15

Table 21. Study B: Descriptive Statistics for CS Fiction (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	13	10.27	11.00	3.24	5	14
1	D	12	10.06	11.00	3.13	4	14
2	A	5	11.00	11.00	2.00	8	14
2	B	3	7.50	5.50	4.32	4	13
2	C	7	8.00	7.00	4.22	2	13
2	D	7	11.46	12.00	2.33	6	14
3	A	10	11.16	12.00	2.81	6	14
3	B	7	11.29	12.00	3.07	6	14
3	C	12	12.21	13.00	2.04	8	14
3	D	12	10.32	10.00	2.38	6	14

Table 22. Study B: Descriptive Statistics for CS Nonfiction (by Grade)

Grade	Form	n	Mean	Median	SD	Minimum	Maximum
1	A	0	n/a	n/a	n/a	n/a	n/a
1	B	0	n/a	n/a	n/a	n/a	n/a
1	C	10	8.73	9.00	3.80	1	14
1	D	9	6.29	6.00	3.02	2	13
2	A	4	6.75	7.00	1.50	5	8
2	B	3	9.50	9.50	1.64	7	12
2	C	7	7.00	6.00	3.46	2	12
2	D	7	9.00	10.00	3.34	4	14
3	A	10	8.95	9.50	3.07	5	14
3	B	7	10.50	10.00	2.24	7	14
3	C	12	10.05	10.00	2.72	2	14
3	D	12	7.11	6.00	2.88	2	12

Reliability

Reliability is generally described as the consistency of a measuring instrument: reliability statistics present information about the precision of an instrument, expressed as a ratio. A test with perfect score precision has a reliability coefficient equal to 1, meaning that 100 percent of the variation among persons' scores is attributable to variation in the trait or skill the test measures, and none of the variation is attributable to error. Perfect reliability is unattainable in educational measurement; a test with a reliability coefficient of 0.90 is more likely. On such a test, 90 percent of the variation among students' scores is attributable to the trait or skill being measured, and 10 percent is attributable to errors of measurement. If the trait or skill were measured a second time, students' scores would fluctuate to some degree; that is, scores on the second test would not be perfectly consistent with the same students' initial scores.

Further, reliability is an essential characteristic of interim and formative assessments that are used for instructional decision-making; if results are spurious and unreliable, inappropriate decisions might be made. Salvia, Ysseldyke, & Bolt's (2013) standards for reliability were used to evaluate the reliability data for the Burst:Reading assessments. According to these standards, a minimum reliability of 0.60 is required to make educational decisions about groups of students, a minimum of 0.70 suggests adequate reliability generally, a minimum of 0.80 is required for screening decisions, and a minimum of 0.90 is required for important educational decisions concerning an individual student. Decisions made from early identification or screening measures, such as the Burst:Reading assessments, typically do not involve a high-stakes decision to change an individual student's placement or educational classification (Kaminski & Good, 1998).

This chapter provides details on three types of reliability evidence for Joule: internal consistency, inter-rater reliability, and alternate form reliability.

- Internal consistency reliability refers to the degree of confidence in the precision of scores from a single measurement.
- Inter-rater reliability estimates the degree to which different raters consistently estimate the same student's performance.
- Alternate form reliability indicates the extent to which test results generalize to different forms. Alternate forms of the test with different items should yield the same approximate scores.

Internal Consistency Reliability

Internal consistency reliability of the Burst:Reading assessments was estimated using Cronbach's alpha as a measure of internal consistency, based on classical test theory. Cronbach's alpha is the most widely used reliability coefficient that measures the degree of internal consistency/homogeneity between variables measuring one construct/concept, i.e., the degree to which different items measuring the same variable provide consistent results (Crocker & Algina, 1986). Study A provides data for internal consistency reliability analyses. Tables 23 through 33 provide Cronbach's alpha and sample sizes for each Burst:Reading Assessment Benchmark form. Results for DEC are provided for each form both across grades and by grade because the same forms are used across grades; results for each VOC and CS form are provided by grade because the forms for these measures are grade specific.

Decode

Regular Words

Table 23 shows Cronbach's alpha for each RW form across grades, and Table 24 shows the results for each form by grade. The mean and median Cronbach's alpha for all the RW forms across grades are 0.79 and 0.80, respectively. All Cronbach's alpha results for RW exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 23. Internal Consistency of DEC RW (All Grades)

Grade	Form	n	Cronbach's Alpha
All	A	608	0.81
All	B	458	0.79
All	C	491	0.75
All	D	461	0.80

Table 24. Internal Consistency of DEC RW (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	207	0.79
1	B	157	0.81
1	C	211	0.79
1	D	163	0.86
2	A	203	0.76
2	B	169	0.76
2	C	144	0.65
2	D	168	0.75
3	A	198	0.79
3	B	132	0.77
3	C	136	0.62
3	D	130	0.69

Letter Combinations

Table 25 shows Cronbach's alpha for each LC form across grades, and Table 26 shows the results for each form by grade. The mean and median Cronbach's alpha for all the LC forms across grades are 0.78 and 0.78, respectively. All Cronbach's alpha results for LC exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 25. Internal Consistency of DEC LC (All Grades)

Grade	Form	n	Cronbach's Alpha
All	A	605	0.79
All	B	457	0.82
All	C	489	0.76
All	D	462	0.76

Table 26. Internal Consistency of DEC LC (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	208	0.79
1	B	155	0.82
1	C	207	0.79
1	D	165	0.78
2	A	200	0.72
2	B	171	0.81
2	C	144	0.63
2	D	168	0.74
3	A	197	0.76
3	B	131	0.80
3	C	138	0.66
3	D	129	0.65

Advanced Phonics

Table 27 shows Cronbach's alpha for each AP form across grades, and Table 28 shows the results for each form by grade. The mean and median Cronbach's alpha for all the AP forms across grades are 0.87 and 0.86, respectively. All Cronbach's alpha results for AP exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 27. Internal Consistency of DEC AP (All Grades)

Grade	Form	n	Cronbach's Alpha
All	A	605	0.89
All	B	458	0.83
All	C	492	0.86
All	D	461	0.87

Table 28. Internal Consistency of DEC AP (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	206	0.89
1	B	157	0.85
1	C	209	0.87
1	D	164	0.89
2	A	200	0.76
2	B	170	0.74
2	C	146	0.63
2	D	168	0.78
3	A	199	0.81
3	B	131	0.81
3	C	137	0.72
3	D	129	0.63

Irregular Words

Table 29 shows Cronbach's alpha for each IW form across grades, and Table 30 shows the results for each form by grade. The mean and median Cronbach's alpha for all the IW forms across grades are 0.81 and 0.81, respectively. Most Cronbach's alpha results for IW exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 29. Internal Consistency of DEC IW (All Grades)

Grade	Form	n	Cronbach's Alpha
All	A	605	0.83
All	B	461	0.84
All	C	490	0.78
All	D	459	0.78

Table 30. Internal Consistency of DEC IW (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	208	0.79
1	B	157	0.82
1	C	207	0.82
1	D	164	0.79
2	A	200	0.72
2	B	172	0.81
2	C	144	0.40
2	D	167	0.74
3	A	197	0.87
3	B	132	0.86
3	C	139	0.59
3	D	128	0.76

Vocabulary

Internal consistency reliability results for VOC are provided in Table 31. The mean and median Cronbach's alpha results for all the VOC forms are 0.71 and 0.82, respectively. Lower Alpha values are observed in kindergarten than in the other grades, which may be due to the broader frequency range allowed for items in this grade than in other grades (see Burst Vocabulary – Design), leading to less homogeneous forms. Generally, Cronbach's alpha results for VOC exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 31. Internal Consistency of VOC (by Grade)

Grade	Form	n	Cronbach's Alpha
K	A	240	0.43
K	B	238	0.59
K	C	241	0.39
K	D	239	0.50
1	A	211	0.89
1	B	206	0.87
1	C	209	0.88
1	D	194	0.91
2	A	198	0.49
2	B	172	0.80
2	C	146	0.55
2	D	168	0.67
3	A	197	0.83
3	B	132	0.86
3	C	138	0.83
3	D	128	0.93

Comprehension

Fiction

Internal consistency reliability results for CS Fiction are provided in Table 32. The mean and median Cronbach's alpha for all the CS Fiction forms across grades are 0.82 and 0.82, respectively. All Cronbach's alpha results for CS Fiction exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 32. Internal Consistency of CS Fiction (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	174	0.83
1	B	164	0.84
1	C	124	0.81
1	D	149	0.86
2	A	197	0.82
2	B	164	0.80
2	C	142	0.79
2	D	164	0.75
3	A	198	0.84
3	B	127	0.88
3	C	135	0.77
3	D	129	0.80

Nonfiction

Internal consistency reliability results for CS Nonfiction are provided in Table 33. The mean and median Cronbach's alpha for all the CS Nonfiction forms across grades are 0.82 and 0.82, respectively. All Cronbach's alpha results for CS Nonfiction exceed the criterion of 0.60 for acceptable reliability suggested by Salvia, Ysseldyke, and Bolt (2013).

Table 33. Internal Consistency of CS Nonfiction (by Grade)

Grade	Form	n	Cronbach's Alpha
1	A	136	0.82
1	B	157	0.88
1	C	113	0.84
1	D	142	0.85
2	A	192	0.80
2	B	166	0.79
2	C	133	0.75
2	D	167	0.81
3	A	188	0.84
3	B	132	0.87
3	C	133	0.80
3	D	133	0.81

Inter-Rater Reliability

Inter-rater reliability (IRR) indicates the extent to which test results generalize across assessors. IRR is important for screening assessments such as Burst:Reading as: student reading performance should be scored in the same manner by any trained administrator, leading to the same outcome. Score fluctuations are attributable to sources of error via the assessors, including scoring mistakes and differing interpretations of scoring procedures and student responses. The IRR estimates reported here are based on two or more independent assessors simultaneously scoring student performance during a single test administration ("shadow-scoring").

Agreement between raters is typically evaluated using either Cohen's kappa (for nominal variables) or intra-class correlations (for ordinal, interval, or ratio variables; Hallgren, 2012). Because Burst:Reading Assessment scores are ordinal, we utilize the intra-class correlation (ICC) to evaluate IRR. ICC is one of the most commonly used statistics for assessing IRR for ordinal, interval, or ratio variables and is suitable for studies with two or more coders (Hallgren, 2012). Cicchetti (1994) provides commonly cited interpretations of agreement based on ICC values: ICC values less than 0.40 are poor, values between 0.40 and 0.59 are fair, values between 0.60 and 0.74 are good, and values between 0.75 and 1.00 are excellent. In addition, percent agreement within a reasonable range is provided for each measure. Percent agreement is calculated by dividing the number of instances of score agreement across rates by the total number of scores; thus, percent agreement can vary between 0 and 100 percent. For DEC, percent agreement within 1 point is provided (fluctuations within 1 point is reasonable for the DEC submeasures, which have total scores of 8 points); for VOC and CS, percent agreement within 2 points is provided (fluctuations within 2 points are reasonable given total score ranges that are larger than DEC submeasures; the total scores of CS and VOC are 14 and 15 points, respectively). The higher the percent agreement, the stronger the evidence for IRR.

Study B provides data for IRR analyses. Results of the IRR analyses for each of the Burst:Reading assessments are reported in Tables 34 through 40, including the number of assessments included in each analysis, the number of raters, ICCs, and percent agreement. Sample sizes are provided for the number of assessments rather than the number of unique students; because students were assessed multiple times by different raters with different forms, the number of unique students does not match the number of assessments. In general, raters agree very often when scoring the Burst:Reading assessments.

Decode

Regular Words

Results of the IRR analyses for RW are reported in Table 34. The intra-class correlation for overall sample was 0.76 and the median grade-specific intra-class correlation was 0.88, which suggest good-to-excellent agreement among raters. The percent agreement values were also high, suggesting that raters agree very often when scoring RW.

Table 34. Inter-Rater Reliability of DEC RW

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 1 Point
All	103	6	0.76	95.15%
1	36	2	0.93	97.22%
2	24	6	0.88	95.83%
3	43	4	0.68	93.02%

Letter Combinations

Results of the IRR analyses for LC are reported in Table 35. The intra-class correlation for overall sample was 0.70, and the median grade-specific intra-class correlation was 0.79, which suggests good agreement among raters. The percent agreement values were high, suggesting that raters agree very often when scoring LC.

Table 35. Inter-Rater Reliability of DEC LC

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 1 Point
All	102	6	0.70	94.12%
1	36	2	0.72	91.67%
2	24	6	0.84	95.83%
3	42	4	0.79	95.24%

Advanced Phonics

Results of the IRR analyses for AP are reported in Table 36. The intra-class correlation for overall sample was 0.83, and the median grade-specific intra-class correlation was 0.89, which suggests excellent agreement among raters. The percent agreement values were high, suggesting that raters agree very often when scoring AP.

Table 36. Inter-Rater Reliability of DEC AP

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 1 Point
All	101	6	0.83	95.05%
1	36	2	0.89	88.89%
2	23	6	0.92	100%
3	42	4	0.80	97.62%

Irregular Words

Results of the IRR analyses for IW are reported in Table 37. The intra-class correlation for overall sample was 0.72, and the median grade-specific intra-class correlation was 0.75, which suggests good agreement among raters. The percent agreement values were high, suggesting that raters agree very often when scoring IW.

Table 37. Inter-Rater Reliability of DEC IW

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 1 Point
All	103	6	0.72	100%
1	36	2	0.92	100%
2	24	6	0.75	100%
3	43	4	0.71	100%

Vocabulary

Results of the IRR analyses for VOC are reported in Table 38. The intra-class correlation for overall sample was 0.55, and the median grade-specific intra-class correlation was 0.59, suggesting fair agreement among raters. The grade 1 intra-class correlation was less than 0.4, the criterion for fair agreement, but the VOC measure generally shows high ceiling effects and less variability in scores, given that it identifies the very lowest performers in vocabulary. Intra-class correlations can be misleading if there is low variation in scores across students (Graham, Milanowski, & Miller, 2012), especially when the sample sizes are also small. However, the percent agreement values were high across grades, suggesting that raters very often agree when scoring VOC.

Table 38. Inter-Rater Reliability of VOC

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 2 Points
All	159	6	0.55	99.37%
K	56	2	0.63	98.21%
1	36	2	0.35	100%
2	24	6	0.54	100%
3	43	4	0.99	100%

Comprehension

Fiction

Results of the IRR analyses for CS Fiction are reported in Table 39. The intra-class correlation overall was 0.63, and the median grade-specific intra-class correlation result was 0.73, suggesting good agreement among raters. The percent agreement values were high, suggesting that raters often agree when scoring CS Fiction.

Table 39. Inter-Rater Reliability of CS Fiction

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 2 Points
All	87	6	0.63	94.25%
1	25	4	0.92	100%
2	21	6	0.59	90.48%
3	41	4	0.73	92.68%

Nonfiction

Results of the IRR analyses for CS Nonfiction are reported in Table 40. The intra-class correlation overall was 0.51, and the median grade-specific intra-class correlation was 0.38, suggesting fair to poor agreement among raters. Grade 1 and 2 intra-class correlations fall below the threshold of 0.4 for fair agreement, likely due to the relatively small sample sizes in grades 1 and 2 as compared to grade 3 (smaller sample sizes are associated with larger measurement error). The percent agreement values, however, were high, suggesting that raters very often agree when scoring CS Nonfiction.

Table 40. Inter-Rater Reliability of CS Nonfiction

Grade	Assessments (n)	Raters (n)	Intra-Class Correlation	Percent Agreement Within 2 Points
All	78	6	0.51	92.31%
1	19	4	0.32	100%
2	19	6	0.38	94.74%
3	40	4	0.78	92.50%

Alternate Form Reliability

Alternate form reliability indicates the extent to which test results generalize to different item samples. Students are tested with two different but equivalent (i.e., alternate) forms of the test, and scores from these two forms are correlated. Alternate forms of a test should give approximately equivalent scores. Administering alternate forms of the same measure may lead to practice effects due to the similarity of the items and administration procedures, but to a lesser degree than in test-retest reliability studies in which the same form is administered to students twice.

Ranges of mean score differences between forms, intra-class correlations (ICC), Pearson correlations, and percent agreement were used to evaluate alternate form reliability with the Study A sample:

- The range of mean score differences between forms quantifies differences in the difficulties of the alternate forms. Smaller differences are desirable as they suggest that the alternate forms are of similar difficulty.
- ICC is one of the most commonly used statistics for assessing agreement between raters or alternate forms for ordinal, interval, or ratio variables and is suitable for studies with two or more raters or forms (Hallgren, 2012). Criteria for the evaluation of ICC values describing IRR are the same as those provided earlier for inter-rater reliability (see Cicchetti, 1994).
- Pearson correlations were used to describe the strength of the relationship between student performance on alternate forms (Crocker & Algina, 1986). We provide the range and median of the Pearson correlations results for each of the measures. Higher correlations between forms provide evidence for alternate form equivalence.
- Percent agreement within a reasonably small score range is also provided as alternate form reliability evidence for each measure (Graham, et al., 2012). The methodology is the same as described previously for inter-rater reliability where percent agreement is calculated by dividing the number of times alternate form scores agree by the total number of score pairs. Percent agreement can vary between 0 and 100 percent; higher agreement percentages suggest stronger evidence alternate form equivalence.

Study A provides data for alternate form reliability analyses. Results of the alternate form analyses for each of the Burst:Reading assessments are reported in Tables 41 through 47. In general, fair to good agreement was observed between alternate forms.

Decode

Regular Words

Results of the alternate form analyses for RW are reported in Table 41. Intra-class correlations for alternate forms were 0.53 across grades and above 0.4 for each grade. The correlations between forms are moderate to strong, and percent agreement ranges from 45 to 69 percent. Descriptive statistics for RW (Tables 1 and 2) show that mean score differences between the forms are less than 0.30 points overall. Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting the alternate form reliability of RW.

Table 41. Alternate Form Reliability of DEC RW

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 1 Point
All	0.00–0.30	0.53	0.46–0.70	0.58	55.93%
1	0.02–0.87	0.46	0.37–0.61	0.53	44.91%
2	0.04–0.22	0.55	0.49–0.76	0.60	57.20%
3	0.04–0.74	0.58	0.58–0.81	0.59	68.69%

Letter Combinations

Results of the alternate form analyses for LC are reported in Table 42. Intra-class correlations for alternate forms were 0.56 for all students and were above 0.4 for each grade. The correlations between forms are moderate to strong, and percent agreement ranges from 42 to 58 percent.

Descriptive statistics for LC (Tables 3 and 4) show that mean score differences between the forms are less than 0.84 points overall. Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting the alternate form reliability of LC.

Table 42. Alternate Form Reliability of DEC LC

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 1 Point
All	0.14–0.84	0.56	0.58–0.80	0.62	50.27%
1	0.01–0.94	0.52	0.52–0.76	0.58	42.22%
2	0.08–0.68	0.58	0.59–0.78	0.63	52.47%
3	0.38–1.24	0.51	0.41–0.83	0.54	57.87%

Advanced Phonics

Results of the alternate form analyses for AP are reported in Table 43. Intra-class correlations for alternate forms were 0.63 for all students and were above 0.4 for each grade. The correlations between forms are moderate to strong, and percent agreement ranges from 44 to 66 percent.

Descriptive statistics for AP (Tables 5 and 6) show that the mean score differences between the forms are less than 0.66 points overall. Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting the alternate form reliability of AP.

Table 43. Alternate Form Reliability of DEC AP

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 1 Point
All	0.02–0.66	0.63	0.64–0.75	0.69	53.28%
1	0.10–0.79	0.62	0.64–0.74	0.65	43.98%
2	0.23–1.15	0.43	0.30–0.63	0.47	52.08%
3	0.22–1.06	0.46	0.15–0.80	0.45	66.36%

Irregular Words

Results of the alternate form analyses for IW are reported in Table 44. Intra-class correlations for alternate forms were 0.63 for all students and were above 0.4 for each grade. The correlations between forms are moderate to strong, and percent agreement ranges from 58 to 91 percent.

Descriptive statistics for DEC IW (Tables 7 and 8) show that the mean score differences between the forms are less than 0.45 points overall. Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting alternate form reliability of IW.

Table 44. Alternate Form Reliability of DEC IW

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 1 Point
All	0.02–0.45	0.63	0.60–0.80	0.64	74.97%
1	0.12–0.95	0.60	0.58–0.77	0.64	57.99%
2	0.06–0.33	0.52	0.37–0.79	0.47	79.47%
3	0.02–0.34	0.63	0.45–0.90	0.58	90.74%

Vocabulary

Results of the alternate form analyses for VOC are reported in Table 45. The median intra-class correlation for alternate forms of the VOC measure was 0.54. The correlations between forms are moderate to strong, which suggests that a student's performance on alternate forms is highly similar. Percent agreement was high overall and for each grade, indicating that the VOC forms are generally of equivalent difficulty. Relatively lower correlations among forms were observed in kindergarten, potentially due to the broader frequency range allowed for items in this grade compared to other grades (see Burst Vocabulary – Design), leading to less homogeneous forms. Descriptive statistics for VOC (Table 9) show that the maximum mean score difference between the forms is 1.33 points (out of a total of 15 points). Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting alternate form reliability of VOC.

Table 45. Alternate Form Reliability of VOC

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 2 Points
K	0.18–0.78	0.23	0.13–0.42	0.22	71.43%
1	0.11–0.81	0.66	0.28–0.86	0.51	82.17%
2	0.01–0.32	0.43	0.25–0.67	0.44	88.43%
3	0.14–1.33	0.65	0.59–0.93	0.69	91.08%

Comprehension Skills

Fiction

Results of the alternate form analyses for CS Fiction are reported in Table 46. The median intra-class correlation for alternate forms is 0.47. The correlations between forms are moderate to strong, and percent agreement ranges from 48 to 51 percent. Descriptive statistics for CS Fiction (Table 10) show that the maximum mean score difference between the forms is 2.63 points (out of a total of 14 points). Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting alternate form reliability of CS Fiction.

Table 46. Alternate Form Reliability of CS Fiction

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 2 Points
1	0.15–2.63	0.45	0.39–0.71	0.46	47.72%
2	0.23–1.39	0.47	0.47–0.77	0.52	50.94%
3	0.12–2.45	0.52	0.47–0.84	0.63	48.85%

Nonfiction

Results of the alternate form analyses for CS Nonfiction are reported in Table 47. The median intra- class correlation for alternate forms is 0.49. The correlations between forms are moderate to strong, and percent agreement ranges from 42 to 50 percent. Descriptive statistics for CS Nonfiction (Table 11 show that the maximum mean score difference between the forms is 3.30 points (out of a total of 14 points). Mean score differences, intra-class correlations, Pearson correlations, and percent agreement suggest fair to good evidence supporting alternate form reliability of CS Nonfiction.

Table 47. Alternate Form Reliability of CS Nonfiction

Grade	Range of Mean Score Differences Between Forms	Intra-Class Correlation	Range of Correlations Between Forms	Median Correlation Between Forms	Percent Agreement Within 2 Points
1	0.58–3.30	0.46	0.54–0.79	0.64	50.20%
2	0.28–1.16	0.49	0.23–0.70	0.57	48.13%
3	0.13–2.34	0.49	0.50–0.82	0.55	42.20%

Validity

The validity of a test is the degree to which it measures the construct that it claims to measure. Validity is formally defined as the degree to which evidence and theory support the interpretation of test scores according to test usage (American Educational Research Association, 1999). In other words, validity represents our degree of confidence that interpretations of test scores accurately represent what we believe they do (e.g., high scores on a comprehension assessment actually represent high comprehension skill). In this sense, validity is a way to describe a test's accuracy or utility.

Validity is not "proven" but rather evidence is collected to strengthen the assertion that a test accurately measures the desired construct(s). Validity was traditionally considered a property assessments themselves possessed; it was categorized as content-, construct-, and criterion validity. The current view, however, considers a more unified treatment under which validity evidence is collected to support test score interpretations (Kane, 2001; Messick, 1989) and may be captured under a more general heading of evidence for construct validity. Assessing the validity of a test involves the use of data and other information both internal and external to the test instrument itself.

To facilitate discussion and demonstration, evidence for the construct validity and criterion validity of the Burst:Reading Assessment is presented via concurrent and prediction results. Criterion-related validity is the extent to which student performance on the assessment procedure being validated can estimate student performance on a criterion measure (Salvia, et al., 2013). Criterion-related validity includes concurrent and predictive validity. Evidence for the concurrent or predictive validity of an assessment refers to the degree to which current outcomes are associated with outcomes on an external, conceptually-related, instrument administered near-concurrently (concurrent validity evidence) or subsequently (predictive validity evidence).

Concurrent validity was evaluated for students administered the Burst:Reading assessments and two external measures within a two-month time period. Predictive validity was evaluated for students administered the Burst:Reading assessments followed by two external measures, at an interval of greater than two months apart; specifically, students in the predictive validity analyses took the Burst:Reading assessments at MOY and DIBELS Next and SEL at EOY approximately 3–4 months apart. These analyses provide an estimate of the linear relationship between Burst:Reading Assessment measure (or sub measure) scores and scores on external measures, covering a similar academic domain. Predictive correlations are attenuated by time due to learning that occurs in the interim between testing occasions; both predictive and concurrent correlations are attenuated by differences in test content specifications. Two criterion measures of similar constructs were

used in the concurrent and predictive validity analyses: DIBELS Next (Dynamic Measurement Group, 2010) and STAR Early Literacy (SEL; Renaissance Learning, 2001):

- DIBELS Next is a set of screening measures used three times per year to assess early literacy skills, including phonemic awareness, phonics, accurate and fluent reading of connected text, vocabulary, and reading comprehension for students in kindergarten through grade 6. DIBELS Next includes the following measures: Letter Naming Fluency (LNF), a general indicator of risk; First Sound Fluency (FSF), a measure of phonemic awareness; Phoneme Segmentation Fluency (PSF), a measure of phonemic awareness; Nonsense Word Fluency (NWF), a measure of alphabetic principle and basic phonics; DIBELS Oral Reading Fluency (DORF), a measure of advanced phonics and word attack skills, accuracy and fluency with connected text, and reading comprehension; and DAZE, a measure of reading comprehension. The measures that a student is administered depends on the student's grade and the benchmark period (e.g., students are not administered DORF until the middle of grade 1). A student's overall composite score is calculated based on his or her scores on grade-specific measures to provide an overall indication of reading skill. We used this composite score as a criterion measure in validity analyses for the Burst:Reading assessments. The reported reliability and validity evidence for each of the DIBELS Next measures indicates that DIBELS Next is adequate for use as a criterion measure (Good, et al., 2013).
- SEL is a computer-adaptive interim assessment of early literacy skills, designed to assess literacy skills and concepts in kindergarten through grade 3. SEL assesses 41 different skill sets in three key domains (Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, and Numbers and Operations) and 10 sub-domains of early literacy and numeracy. Student performance is reported using skill and domain scores; the Rasch ability scale is transformed into a scaled score and a proficiency score. We used the scaled score as a criterion measure in validity analyses for the Burst:Reading assessments because the scaled score is an overall indicator of student reading skills. The reported reliability and validity evidence for SEL indicates that SEL is adequate for use as a criterion measure (Renaissance Learning, 2014).

The majority of students in Study A were administered DIBELS Next during both the middle- and end- of-year benchmark periods and SEL during the end-of-year benchmark period of the 2013–2014 school year. To evaluate concurrent validity evidence, Burst:Reading Assessment results at MOY and EOY were correlated with DIBELS Next results from the same administration period; in addition, Burst:Reading Assessment Results at EOY were correlated with SEL results at EOY. To evaluate predictive validity evidence, Burst:Reading Assessment results at MOY were correlated with both DIBELS Next and SEL results at EOY.

Concurrent Validity

Pearson correlations were used to characterize the concurrent relationship of the Burst:Reading assessments with DIBELS Next at both MOY and EOY and with SEL at EOY. The medians and ranges of the Pearson correlation coefficients (after removing multivariate outliers) and the sample sizes are provided across all forms for each grade in Tables 48 through 54. Form-specific and administration period (MOY or EOY) specific results for each Burst:Reading Assessment are provided in Appendix 4[4].

In most cases, moderate to strong relationships were found among the measures. Generally, the Burst:Reading assessments were more closely related to DIBELS Next than to SEL, likely due to similarities in testing modality between the Burst:Reading assessments and DIBELS Next. The Burst:Reading assessments and DIBELS Next (except for the DAZE measure) are both administered in a one-on-one setting in which a test administrator elicits open-ended responses from students, while SEL is administered to students on a computer using a multiple-choice response format. Specifically, DEC submeasures were more strongly related to DIBELS Next than to SEL, reflecting the former assessment's more intensive focus on the alphabetic principle and phonics. VOC and CS measures demonstrated similar correlations to both of the criterion measures.

Decode

Regular Words

Concurrent validity results for RW are provided in Table 48. Correlations with DIBELS Next were typically higher than those with SEL for RW. The median correlations with SEL ranged from 0.31 to 0.57 while the median overall correlations with DIBELS Next ranged from 0.58 to 0.71. Overall, the results show moderate to strong relationships among the measures.

SEL typically focuses assessment items on word reading skills for beginning readers while it puts a greater focus on vocabulary and comprehension skills for more advanced readers. It is likely that the stronger correlations observed between SEL and RW in grade 1 than in grades 2 and 3 are due to closer alignment of the content of the two assessments earlier on.

4 Significant mean differences are reported between MOY and EOY, however, those differences are typically very small (mean absolute difference = 0.41; median absolute difference = 0.31) and the shape of the distributions is typically similar, as indicated by the Komolgorov-Smirnov tests. Therefore, results from the MOY and EOY administration periods are collapsed and analyzed together for the purpose of providing concurrent validity evidence.

Table 48. Concurrent Validity of DEC-RW With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.55 (0.21–0.62)	155 (70–191)	0.59 (0.58–0.66)	325 (244–417)
1	0.57 (0.13–0.64)	66 (23–79)	0.71 (0.64–0.74)	108 (82–136)
2	0.31 (0.05–0.62)	44 (23–67)	0.58 (0.50–0.64)	120 (105–144)
3	0.42 (0.30–0.43)	38 (23–56)	0.60 (0.31–0.71)	96 (57–137)

Letter Combinations

Concurrent validity results for LC are provided in Table 49. Correlations with DIBELS Next were typically higher than those with SEL for LC. The median correlations with SEL ranged from 0.37 to 0.60 while the median overall correlations with DIBELS Next ranged from 0.52 to 0.73. Moderate to strong relationships are observed among the measures in most cases. Stronger relationships are seen between LC and SEL at the lower grades than in grade 3, likely because SEL places less of an emphasis on decoding for more advanced readers.

Table 49. Concurrent Validity of DEC LC With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.60 (0.32–0.82)	154 (69–189)	0.64 (0.54–0.68)	324 (248–413)
1	0.56 (0.33–0.69)	66 (24–79)	0.73 (0.67–0.76)	106 (85–136)
2	0.59 (0.38–0.89)	44 (21–67)	0.52 (0.34–0.62)	122 (105–141)
3	0.37 (–0.10–0.51)	39 (22–58)	0.66 (0.45–0.74)	96 (57–136)

Advanced Phonics

Concurrent validity results for AP are provided in Table 50. Correlations with DIBELS Next were typically higher than those with SEL for AP. The median correlations with SEL ranged from 0.28 to 0.57 while the median correlations with DIBELS Next ranged from 0.57 to 0.77. Moderate to strong relationships are observed among the measures in most cases. As with LC, weaker relationships with SEL are observed for AP in grades 2 and 3, which is again likely due to increasing divergence in the content of the assessments as grade level increases.

Table 50. Concurrent Validity of DEC AP With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.56 (0.55–0.71)	156 (68–192)	0.72 (0.67–0.77)	326 (246–413)
1	0.57 (0.55–0.63)	67 (22–81)	0.77 (0.68–0.88)	108 (84–135)
2	0.28 (0.11–0.38)	44 (22–68)	0.68 (0.64–0.77)	123 (105–140)
3	0.43 (–0.12–0.63)	38 (24–54)	0.57 (0.27–0.74)	96 (57–138)

Irregular Words

Concurrent validity results for IW are provided in Table 51. Correlations with DIBELS Next were typically higher than those with SEL for IW. The median correlations with SEL ranged from 0.18 to 0.47 while the median correlations with DIBELS Next ranged from 0.46 to 0.74. Moderate to strong relationships are observed among the measures in most cases. Again, as with LC and AP, weaker relationships with SEL are observed for IW in grades 2 and 3, which is likely explained by increasing divergence in the content of the assessments as grade level increases.

Table 51. Concurrent Validity of DEC IW With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.48 (0.32–0.61)	156 (68–193)	0.62 (0.53–0.70)	326 (245–414)
1	0.47 (0.39–0.59)	68 (22–81)	0.74 (0.55–0.77)	107 (84–137)
2	0.18 (0.04–0.34)	44 (22–66)	0.63 (0.54–0.74)	122 (105–141)
3	0.24 (–0.04–0.63)	40 (24–57)	0.46 (–0.09–0.65)	98 (56–136)

Vocabulary

Concurrent validity results for VOC are provided in Table 52. The median correlations with SEL ranged from 0.20 to 0.40 while the median overall correlations with DIBELS Next ranged from 0.24 to 0.50. These results suggest moderate relationships among the measures in most cases. Concurrent validity evidence in grade 3 was stronger than in other grades. The DIBELS Next Composite Score reflects different combinations of measures based on a student's grade and time of year: while DIBELS next focuses on phonemic awareness, decoding skills, and accurate and fluent reading in grades K–2, the shift focuses towards accurate and fluent reading with comprehension in grade 3 and beyond. Because vocabulary knowledge is such a large component of reading comprehension, the stronger correlations observed between VOC and DIBELS Next in grade 3 compared to grades K, 1, and 2 are unsurprising. A similar pattern is observed for SEL, which also shifts in focus from decoding towards comprehension as readers become more advanced.

Table 52. Concurrent Validity of VOC with SEL and DIBELS Next

Grade	Correlations with SEL	SEL (n)	Correlations with DIBELS Next	DIBELS Next (n)
K	0.25 (0.19–0.27)	84 (65–94)	0.24 (0.12–0.31)	166 (149–172)
1	0.20 (0.11–0.29)	68 (23–80)	0.39 (0.29–0.57)	131 (114–140)
2	0.30 (0.04–0.48)	45 (23–67)	0.40 (0.27–0.41)	123 (105–140)
3	0.40 (0.08–0.86)	40 (23–56)	0.50 (0.31–0.67)	97 (56–136)

Comprehension

Fiction

Concurrent validity results for CS Fiction are provided in Table 53. The median correlations with SEL ranged from 0.39 to 0.54 while the median overall correlations with DIBELS Next ranged from 0.53 to 0.66. These results suggest moderate relationships among the measures in most cases.

Table 53. Concurrent Validity of CS Fiction With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	0.53 (0.49–0.66)	41 (19–68)	0.54 (0.40–0.63)	97 (87–111)
2	0.39 (0.28–0.49)	46 (24–64)	0.53 (0.24–0.61)	118 (102–140)
3	0.54 (0.07–0.59)	38 (24–56)	0.66 (0.44–0.77)	93 (57–137)

Nonfiction

Concurrent validity results for CS Nonfiction are provided in Table 54. The median correlations with SEL ranged from 0.38 to 0.56 while the median overall correlations with DIBELS Next ranged from 0.48 to 0.53. These results suggest moderate relationships among the measures in most cases.

Table 54. Concurrent Validity of CS Nonfiction With SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	0.56 (0.53–0.62)	42 (17–66)	0.52 (0.31–0.60)	88 (80–103)
2	0.38 (0.16–0.85)	44 (19–67)	0.48 (0.06–0.63)	118 (100–133)
3	0.47 (0.28–0.55)	36 (24–59)	0.53 (0.33–0.81)	94 (61–127)

Predictive Validity

Pearson correlations were used to characterize the predictive relationship between the Burst:Reading assessments at MOY and DIBELS Next and SEL at EOY. The medians and ranges of the Pearson correlation coefficients (after removing multivariate outliers) and the sample sizes are provided across all forms for each grade in Tables 55 through 61. Form-specific results for each Burst:Reading Assessment are provided in Appendix 5; results are unavailable for some forms for which the sample size of students who took both predictor and criterion measures was too small.

In most cases, moderate relationships were found among the measures. Generally, the Burst:Reading assessments were more closely related to DIBELS Next than to SEL, likely due to similarities in testing modality between the Burst:Reading assessments and DIBELS Next. The Burst:Reading assessments and DIBELS Next (except for the DAZE measure) are both administered in a one-on-one setting in which a test administrator elicits open-ended responses from students, while SEL is administered to students on a computer using a multiple-choice response format. In addition, the DEC submeasures were more strongly related to DIBELS Next than to SEL, reflecting DIBELS Next's strong focus on decoding skills, whereas VOC and CS demonstrated more similar correlations to both of the criterion measures. These results suggest good predictive validity evidence for the Burst:Reading assessments, both overall and for each individual measure.

Decode

Regular Words

Predictive validity results for RW are provided in Table 55. Correlations with DIBELS Next were typically higher than those with SEL for RW. The median correlations with SEL ranged from 0.48 to 0.60 while the median overall correlations with DIBELS Next ranged from 0.53 to 0.78. These results suggest moderate to strong relationships among the measures in most cases.

Table 55. Predictive Validity of DEC-RW to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.60 (0.50–0.71)	180 (145–376)	0.62 (0.61–0.64)	206 (169–384)
1	0.52 (0.29–0.76)	62 (54–132)	0.70 (0.59–0.71)	66 (61–129)
2	0.48 (0.33–0.65)	61 (41–130)	0.53 (0.44–0.66)	74 (53–132)
3	0.60 (0.11–0.73)	56 (49–114)	0.78 (0.61–0.79)	62 (50–117)

Letter Combinations

Predictive validity results for LC are provided in Table 56. Correlations with DIBELS Next were typically higher than those with SEL for LC. The median correlations with SEL ranged from 0.44 to 0.52 while the median overall correlations with DIBELS Next ranged from 0.57 to 0.71. These results suggest moderate to strong relationships among the measures in most cases.

Table 56. Predictive Validity of DEC LC to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.54 (0.41–0.71)	183 (145–377)	0.61 (0.54–0.73)	207 (166–378)
1	0.52 (0.37–0.63)	62 (52–131)	0.71 (0.64–0.73)	64 (61–125)
2	0.44 (0.37–0.73)	62 (41–128)	0.57 (0.48–0.61)	77 (52–132)
3	0.52 (0.24–0.55)	56 (47–112)	0.71 (0.63–0.71)	64 (50–115)

Advanced Phonics

Predictive validity results for AP are provided in Table 57. Correlations with DIBELS Next were typically higher than those with SEL for AP. The median correlations with SEL ranged from 0.46 to 0.55 while the median overall correlations with DIBELS Next ranged from 0.62 to 0.80. These results suggest moderate to strong relationships among the measures in most cases.

Table 57. Predictive Validity of DEC AP to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.66 (0.6–0.79)	180 (143–366)	0.73 (0.65–0.83)	209 (164–379)
1	0.46 (0.37–0.7)	62 (53–132)	0.80 (0.75–0.85)	63 (59–125)
2	0.55 (0.38–0.65)	62 (42–124)	0.71 (0.66–0.79)	78 (54–131)
3	0.50 (0.12–0.82)	54 (49–113)	0.62 (0.55–0.80)	63 (52–118)

Irregular Words

Predictive validity results for IW are provided in Table 58. Correlations with DIBELS Next were typically higher than those with SEL for IW. The median correlations with SEL ranged from 0.44 to 0.56 while the median correlations with DIBELS Next ranged from 0.58 to 0.79. These results suggest moderate to strong relationships among the measures in most cases. Lower correlations with SEL and DIBELS are observed in grade 3 because SEL and DIBELS emphasize decoding skills more for beginning readers than for advanced readers.

Table 58. Predictive Validity of DEC IW to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	0.62 (0.61–0.67)	181 (145–371)	0.69 (0.65–0.70)	203 (169–382)
1	0.56 (0.47–0.76)	62 (54–129)	0.79 (0.76–0.80)	62 (61–128)
2	0.45 (0.40–0.53)	63 (42–126)	0.69 (0.44–0.74)	77 (54–132)
3	0.44 (–0.06–0.74)	56 (48–114)	0.58 (0.14–0.76)	65 (53–120)

Vocabulary

Predictive validity results for VOC are provided in Table 59. The median correlations with SEL ranged from 0.26 to 0.46 while the median correlations with DIBELS Next ranged from 0.26 to 0.72. These results suggest weak to strong relationships among the measures. Generally, correlations with DIBELS Next were stronger but more variable than those with SEL for VOC. Predictive validity evidence in grade 3 was stronger than in other grades. The increase in correlations with both DIBELS and SEL as grades increase may be explained by the skills targeted by each of these measures based on grade level or reading proficiency. Both DIBELS and SEL place more of a focus on phonemic awareness and decoding skills for beginning readers and on vocabulary and comprehension for more advanced readers; thus, VOC is more related to SEL and particularly to DIBELS as grades increase.

Table 59. Predictive Validity of VOC to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
K	0.36 (0.32–0.43)	63 (33–158)	0.26 (0.06–0.31)	94 (58–155)
1	0.26 (0.13–0.40)	69 (53–134)	0.37 (0.27–0.40)	82 (64–133)
2	0.30 (0.09–0.50)	63 (43–127)	0.35 (0.18–0.45)	79 (53–129)
3	0.46 (0.16–0.86)	56 (48–114)	0.72 (0.25–0.74)	65 (53–118)

Comprehension

Fiction

Predictive validity results for CS Fiction are provided in Table 60. Correlations with SEL and DIBELS Next were generally comparable, although higher correlations were observed with DIBELS Next in grade 3, likely due to the increased focus on comprehension in DIBELS as students are first administered DAZE in this grade. The median correlations with SEL ranged from 0.53 to 0.59 while the median overall correlations with DIBELS Next ranged from 0.56 to 0.75. These results suggest moderate to strong relationships among the measures.

Table 60. Predictive Validity of CS Fiction to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	0.55 (0.30–0.71)	52 (38–112)	0.58 (0.48–0.67)	62 (38–107)
2	0.53 (0.05–0.72)	57 (41–125)	0.56 (0.33–0.79)	67 (50–125)
3	0.59 (0.31–0.70)	52 (45–114)	0.75 (0.50–0.79)	62 (51–117)

Nonfiction

Predictive validity results for CS Nonfiction are provided in Table 61. Correlations with SEL and DIBELS Next were generally comparable, although higher correlations were observed with DIBELS Next in grade 3, likely due to the increased focus on comprehension in DIBELS as students are first administered DAZE in this grade. The median correlations with SEL ranged from 0.43 to 0.56 while the median overall correlations with DIBELS Next ranged from 0.42 to 0.74. These results suggest moderate to strong relationships among the measures in most cases.

Table 61. Predictive Validity of CS Nonfiction to SEL and DIBELS Next

Grade	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	0.51 (0.46–0.63)	50 (34–85)	0.42 (0.33–0.70)	64 (34–82)
2	0.43 (0.24–0.69)	57 (41–123)	0.49 (0.29–0.73)	74 (48–121)
3	0.56 (0.33–0.69)	56 (47–107)	0.74 (0.15–0.74)	66 (54–108)

Cut Points

Burst:Reading Assessment cut points for determining intervention need and priority within the Burst:Reading Intervention were set using student performance results from Study A. The cut points for BOY and MOY are the same as the cut points proposed for EOY because: (1) the field study suggests similar student performance at MOY and EOY (see Appendix 3 for details); (2) the score range is too small to have three different cut points; (3) to allow room for growth for students with lower skill ratings at BOY; (4) to align with the purpose of the assessment, which is to identify students who need intervention. The procedures to set cut points were the following:

1. Students' scale scores on SEL were categorized into three groups based on the SEL cut points: less than 25th percentile (25%), 25th to 40th percentile (25–40%), and greater than 40th percentile (40%). These groups help educators identify students who require intervention to accelerate growth and move toward proficiency. This aligns with the purpose of the Burst:Reading assessments.
2. The percentages of students in each SEL group out of the total number of students in each analysis as well as mean and median performance on the Burst:Reading assessments for each SEL group were calculated.
3. The primary consideration in setting the cut points was the results of the contrasting group analysis (Cizek & Bunch, 2007). Under this method, mean and median Burst:Reading Assessment performance is calculated for each SEL group. The midpoint of these mean and median values are next calculated to provide tentative thresholds between the SEL groups: 25 percent versus 25–40 percent, which correspond to tentative cut points for the 0/1 Burst:Reading Assessment skill ratings (Well Below Benchmark/Below Benchmark), and 25–40 percent versus ≥ 40 percent, used to set the tentative cut points for the 1/2 Burst:Reading Assessment skill ratings (Below Benchmark/At or Above Benchmark).
4. The secondary consideration was consistency of the percentages of students in SEL groups compared to the percentages of students in the Burst:Reading Assessment skill groups. In considering potential cut points, we tried to keep the percentages of students in each skill rating group consistent between the predictor to criterion measures. For example, 63 percent of students in our grade 1 sample scored at or above the 40th percentile on SEL; therefore, we tentatively set the grade 1 EOY benchmark goal so that 63 percent of students also had a skill rating of 2 (At or Above Benchmark).
5. Based on the above two considerations, the range of tentative cut points was submitted to examination of classification accuracy, specificity, sensitivity, logistic regression analyses results (i.e., likelihood of being at or above the 40th percentile on SEL), negative prediction value (i.e., the probability of being at or above the 40th percentile on SEL given At or Above Benchmark status on the Burst:Reading Assessment), marginal percentages (i.e., the percentages of students at or above the 40th percentile on SEL for a specific Burst:Reading Assessment score), and the score distributions of the three SEL groups.

6. There were other considerations specific to each Burst:Reading Assessment, including:
   - Preserving the three skill ratings for each measure by ensuring nonoverlapping cut points
   - Setting cut points below the maximum score of each measure to allow for minor student errors
   - Preferring lower cut points from the tentative cut points as suggested by the analytic methods, allowing for demonstration of growth within a grade
   - Considering theoretical knowledge about reading development when choosing between multiple possible cut points
7. We specified the cut points for the 0/1 and 1/2 skill ratings as an overall evaluative judgment of the above considerations. The final cut points for measure and grade are presented in Table 62.

Table 62. Cut Points for the Burst:Reading assessments

Assessment	Total Score	K	1	2	3
DEC RW	8	n/a	4/6	5/7	6/7
DEC LC	8	n/a	3/5	4/6	6/7
DEC AP	8	n/a	3/5	5/7	6/7
DEC IW	8	n/a	5/7	6/7	6/7
VOC	15	12/14	13/14	13/14	13/14
CS	14	n/a	6/9	6/9	8/11

Note. (1) The score before "/" is the cut point for 0/1 skill ratings and the score after "/" is the cut point for 1/2 skill ratings. The cut points should be interpreted as "at or above."

References

American Educational Research Association, Psychological Association, & National Council on Measurement in Education. (1999). **Standards for educational and psychological testing**. Washington, DC: Author.

Baron, J. & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. **Journal of Experimental Psychology: Human Perception & Performance, 2**(3), 386–393.

Barron, R. W. (1986). Word recognition in early reading: A review of the direct and indirect access hypotheses. **Cognition, 24**, 93–119.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). **Bringing words to life: Robust vocabulary instruction**. New York: Guilford Press.

Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials? **Psychology in the Schools**, 51(2), 198–215.

Biemiller, A. (2010). **Words worth teaching: Closing the vocabulary gap**. Columbus, OH: McGraw Hill.

Bybee, J. (1995). Diachronic and typological properties of morphology and their implications for representation. In L. E. Feldman (Ed.), **Morphological aspects of language processing** (pp. 225– 246). Hillsdale, NJ: Erlbaum.

Carlisle, J. (2003). Morphology matters in learning to read: A commentary. **Reading Psychology**, 24(3), 291–322.

Carnine, D., Silbert, J., Kame'enui, E., & Tarver, S. (2004). **Direct instruction reading** (4th ed.). Upper Saddle River, NJ: Prentice Hall/Merrill.

Catts, H. W., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. **Journal of Speech, Language, and Hearing Research**, 49(2), 278–293.

Chall, J. S. (1983). **Stages of reading development**. New York: McGraw Hill.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. **Psychological Assessment**, 6(4), 284.

Cizek, G. J., & Bunch, M. B. (2007). **Standard setting: A guide to establishing and evaluating performance standards on tests**. Thousand Oaks, CA: SAGE Publications Ltd.

Crocker, L., & Algina, J. (1986). **Introduction to classical and modern test theory**. Orlando, FL: Holt, Rinehart and Winston.

Dolch, E. (1948). **Problems in reading**. Champaign, IL: Garrard Press.

Duke, N.K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. **Reading Research Quarterly, 35**, 202–224.

Dynamic Measurement Group. (2010). DIBELS Next Benchmark Goals and Composite Scores. Retrieved from https://dibels.uoregon.edu/docs/DIBELSNextFormerBenchmarkGoals.pdf

Ehri, L. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. Metsala and L. Ehri. **Word recognition in beginning literacy** (pp. 3–40). Mahwah, NJ: Lawrence Erlbaum.

Embretson, S. E., & Reise, S.P. (2000). **Item response theory for psychologists**. Mahwah, NJ: Lawrence Erlbaum.

Good, R. H., Kaminski, R., Dewey, E., Walin, J., Powell-Smith, K., & Latimer, R. (2013). **DIBELS Next technical manual**. Eugene, OR; Dynamic Measurement Group, Inc.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. **Remedial and Special Education, 7**(1), 6–10.

Graham, M., Milanowski, A., & Miller, J. (2012). **Measuring and promoting inter-rater agreement of teacher and principal performance ratings**. Center for Educator Compensation Reform.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. **Tutorials in Quantitative Methods for Psychology**, 8(1), 23.

Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap by age 3. **American Educator, 27**(1), 4–9.

Hoffman, J.V., McCarthey, S.J., Abbott, J., Christian, C., Corman, L., Curry, C., Dressman, M., Elliott, B., Matherne, D., & Stahle, D. (1994). So what's new in the new basals? A focus on first grade. **Journal of Reading Behavior, 26**, 47–73.

Hudson, R. (1994). About 37% of word-tokens are nouns. **Language**, 331–339.

Johns, J. & Lenski, S. (1997). **Improving reading: A handbook of strategies** (2nd ed.). Dubuque, IA: Kendall Hunt.

Kamberelis, G. (1998). Relations between children's literacy diets and genre development: You write what you read. **Literacy Teaching and Learning, 3**, 7–53.

Kamil, M. L., & Lane, D. (1997). **A classroom study of the efficacy of using information text for first- grade reading instruction**. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kaminski, R., & Good, R. (1996). Toward a technology for assessing basic early literacy skills. **School Psychology Review, 25**(2), 215–227.

Kane, M. (2001). Current concerns in validity theory. **Journal of Educational Measurement, 38**(4), 319–342.

Kintsch, W. & Van Dijk, T. (1978). Toward a model of text comprehension and production. **Psychological Review, 85**(5), 363–394.

Linacre, J. M. (2014). **Winsteps® Rasch measurement computer program user's guide**. Beaverton, OR: Winsteps.com

McKoon, G. & Ratcliff, R. (1992). Inference during reading. **Psychological Review, 99**(3), 440–446. Messick, S. (1989). Validity. In R. L. Linn (Ed.), **Educational measurement** (3rd ed., pp. 13–103). New York: Macmillan.

Mogilner, A. & Mogilner, T. (2006). **Children's writer's word book** (2nd ed.). Cincinnati, OH: Writer's Digest Books.

Morrow, L. M. (1988). Retelling stories as a diagnostic tool. **Reexamining reading diagnosis: New trends and procedures**, 128–149.

Muter, V., Hulme, C., Snowling, M., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. **Developmental Psychology, 40**(5), 665–681.

Nation, I. (2001). **Learning vocabulary in another language**. Cambridge, UK: Cambridge University Press.

National Institute of Child Health and Human Development. (2000). **Report of the National Reading Panel Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction** (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Paris, S. G. (2005). Reinterpreting the development of reading skills. **Reading Research Quarterly**, 40(2), 184–202.

Pearson, P. D., & Dole, J. A. (1987). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. **The Elementary School Journal**, 151–165.

Perfetti, C., Bell, L., & Delaney, S. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. **Journal of Memory and Language, 27**, 1–22.

Renaissance Learning. (2001). STAR Early Literacy. Wisconsin Rapids, WI: Author. Renaissance Learning. (2014). STAR Early Literacy: Technical manual. Wisconsin Rapids, WI:

Author. Available by request to research@renlearn.com or online at http://doc.renlearn.com/KMNet/R004327609GJB0BF.pdf

Roth, F., Speece, D., & Cooper, D. (2002). A longitudinal analysis of the connection between oral language and early reading. **Journal of Educational Research, 95**(5), 259–273.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). **Assessment in special and inclusive education**. Independence, KY: Cengage Learning.

Samuels, S. (2002). Reading fluency: Its development and assessment. In S. J. Samuels & A. E. Farstrup (Eds.), **What research has to say about reading instruction** (pp. 166–183). Newark, DE: International Reading Association.

Samuels, S. & Flor, R. (1997). The importance of automaticity for developing expertise in reading. **Reading & Writing Quarterly**, 13(2), 107–121.

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? **Frontiers in psychology**, 3(34).

Snow, C. (2002). **Reading for understanding: Toward an R&D program in reading comprehension**. Santa Monica, CA: RAND Corporation.

Spache, G. (1974). **Good reading for poor readers**. Champaign, IL: Garrard Press.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. **Reading Research Quarterly, 21**(4), 360–405.

Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). The Lexile framework for reading technical report. Retrieved from Metametrics website https://www.lexile.com/research/9/.

Storch, S. & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. **Developmental Psychology, 38**(6), 934–947.

Taylor, S. E., Frackenpohl, H., & White, C. E. (1989). A revised core vocabulary. **EDL core vocabularies in reading, mathematics, science, and social studies.** Columbia, SC: Educational Developmental Laboratories, Inc.

Wolman, D. (2008). **Righting the mother tongue: From Olde English to email, the tangled story of English.** New York: HarperCollins.

Yopp, R.H., & Yopp, H.K. (2000). Sharing informational text with young children. **The Reading Teacher, 53**, 410–423.

Zeno, S., Ivens, S., Duvvuri, R., & Millard, R. (1995). **The educator's word frequency guide**. Brewster, NY: Touchstone Applied Science Associates.

Appendix 1. Demographic Comparison of Schools Nationwide, Schools Using the Burst:Reading Assessment, and Schools in Study A

Table 63. Demographic Comparison

	Schools Nationwide					Schools Using Burst:Reading Assessment in 2013–2014					Study A Schools
	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Overall
Sample Size (n)											
States	51	51	51	51	51	19	22	20	18	23	6
Districts	14,556	14,606	14,600	14,585	14,904	292	320	313	235	346	7
Schools	50,884	51,540	51,547	51,508	58,500	829	986	973	720	1133	7
Students	20,607,036	21,036,776	20,785,859	20,739,751	122,623,894	99,935	109,012	100,503	72,967	382,417	1,116
Geographic Region (%)											
Midwest	24.45	24.62	24.57	24.64	24.51	23.1	23.81	23.05	4.57	19.49	0.00
Northeast	15.80	15.91	15.88	15.78	15.69	0.24	0.30	0.31	0.42	0.31	14.29
South	34.30	34.12	34.19	34.20	34.24	9.15	8.48	7.89	5.96	7.95	42.86
West	25.44	25.35	25.36	25.37	25.57	67.51	67.41	68.75	89.06	72.24	42.86
Location Relative to Population Centers (%)											
City: large	15.53	15.52	15.31	15.26	15.43	17.45	19.37	19.88	20.50	19.29	0.00
City: mid-size	6.45	6.40	6.43	6.42	6.47	4.09	4.44	4.41	3.19	4.09	0.00
City: small	7.57	7.51	7.52	7.51	7.53	6.62	6.56	6.35	4.29	6.05	14.29

Table 63. Demographic Comparison

	Schools Nationwide					Schools Using Burst:Reading Assessment in 2013–2014					Study A Schools
	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Overall
Suburb: large	24.31	24.5	24.56	24.53	24.41	15.64	15.34	14.96	17.45	15.74	14.29
Suburb: mid-size	2.86	2.86	2.86	2.87	2.84	1.81	1.82	1.95	0.69	1.62	0.00
Suburb: small	1.86	1.89	1.88	1.87	1.87	0.96	0.91	0.72	0.69	0.82	0.00
Town: fringe	1.53	1.52	1.54	1.54	1.52	1.68	1.41	1.43	1.39	1.48	0.00
Town: distant	5.35	5.39	5.41	5.38	5.36	3.97	4.34	4.82	3.74	4.26	0.00
Town: remote	3.68	3.73	3.72	3.74	3.68	8.66	8.68	8.91	12.19	9.46	0.00
Rural: fringe	12.65	12.65	12.71	12.81	12.74	15.04	14.53	14.24	12.88	14.23	57.14
Rural: distant	11.55	11.46	11.45	11.48	11.52	11.55	11.10	10.96	8.73	10.68	14.29
Rural: remote	6.67	6.59	6.62	6.59	6.64	12.52	11.50	11.37	14.27	12.27	0.00
Missing	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
School Type and Characteristics (%)											
Schoolwide Title I	58.28	58.24	58.13	58.06	58.08	68.35	67.81	68.65	67.04	68.01	42.86
Charter school	6.05	6.03	5.97	5.92	6.01	2.65	3.03	2.97	3.60	3.04	28.57
Regular school	97.92	97.65	97.51	97.39	97.43	99.76	99.29	99.59	99.31	99.49	100
Special education school	1.13	1.23	1.33	1.40	1.35	0.12	0.40	0.10	0.28	0.23	0.00
Vocational school	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Alternative/other school	0.94	1.11	1.14	1.20	1.20	0.12	0.30	0.31	0.42	0.28	0.00

Table 63. Demographic Comparison

	Schools Nationwide					Schools Using Burst:Reading Assessment in 2013–2014					Study A Schools
	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Kindergarten	Grade 1	Grade 2	Grade 3	Overall	Overall
Pupil to teacher ratio	16.77	16.71	16.68	16.69	16.71	16.49	16.87	16.92	16.89	16.80	17.07
Percentage of Free/ reduced lunch	53.02	53.02	52.98	52.95	53.02	60.46	59.06	58.91	59.68	59.48	64.83
Student Characteristics (%)											
Male	51.69	51.73	51.77	51.78	51.77	51.46	51.34	51.42	51.25	51.37	48.30
Female	48.31	48.27	48.23	48.22	48.23	48.54	48.66	48.58	48.75	48.63	51.70
White	53.29	53.30	53.44	53.49	53.33	48.57	48.42	47.94	43.05	47.22	48.32
Black	15.55	15.59	15.54	15.56	15.62	8.74	8.82	8.33	4.55	7.79	26.93
Hispanic	22.16	22.15	22.05	21.99	22.08	35.48	35.54	36.59	44.91	37.74	15.39
Am. Indian/Alaskan Native	1.71	1.70	1.71	1.71	1.72	3.01	2.79	2.79	3.59	3.00	4.97
Asian	3.98	3.96	3.95	3.96	3.96	1.71	1.94	1.87	1.95	1.89	0.98
Hawaiian Native/ Pacific Islander	0.39	0.39	0.39	0.39	0.39	0.14	0.15	0.14	0.17	0.15	1.01
Two or more races	2.92	2.91	2.92	2.90	2.90	2.35	2.34	2.33	1.79	2.23	2.40

Appendix 2. Principal Component Analysis for Determining Comprehension Skills Total Score

Principal Component Analysis (PCA) was used to determine the formula for calculating the Comprehension Skills (CS) Total Score for both fiction and nonfiction measures based on retell/recall, literal question, and inferential question scores. PCA was used to reduce the explanatory variables to one or a few principal components in order to calculate a composite score. The procedure for conducting PCA is as follows:

1. Calculate correlations between different possible explanatory variables and extract the one principal component score for each person
2. Run a regression analysis in which the principal component score regresses on retell/recall, literal, and inferential scores
3. Round the regression loadings and use them as the loadings to compute a "composite" CS score

The results of this analysis showed that for all grades (i.e., grades 1–3) the CS total score formula should be Retell/Recall score *2 + Literal score *1 + Inferential score *1.

Because the retell/recall score is doubled in this formula, the original CS scoring rubric was adjusted in the Burst:Reading Assessment software so that the calculation would be automatic and transparent for assessment users (i.e., it appears to users that CS composite score = retell/recall score + literal score + inferential score).

Table 64. Retell/Recall Rubric Adjustment

Retell/Recall Interpretation	Original Retell/Recall Score	Adjusted Retell/Recall Score
Insufficient Retell/Recall	0	0
Basic Retell/Recall	1	2
Good Retell/Recall	2	4

Appendix 3. Student Performance Equivalence Across Administration Periods

Statistical analyses were conducted to evaluate the equivalence of student performance across MOY and EOY for those students participating in Study A. Results of these analyses determined whether to analyze the data from the two time periods together or separately.

One-way ANOVA analysis was conducted for each measure (or submeasure) by grade, treating alternate forms as a covariate or source of error, for the purpose of evaluating performance differences between students in the MOY Study A sample and those in the EOY Study A sample. With the exception of AP in grade 1, none of the measures show a score increase of more than one point between MOY and EOY, and the majority of the measures showed non-significant results on the F-test. Non-significant results suggest MOY and EOY performance results are equivalent. We also used the Kolmogorov-Smirnov (K-S) test to compare the distribution and mean of performance on the measures. Most of the results are non-significant; therefore, student performance from the MOY and EOY Study A samples were combined for the analyses described in this Research Report. Table 65 provides the specific results.

Table 65. Study A: MOY and EOY Equivalence Test

Measure	Grade	MOY Mean Score	EOY Mean Score	F-Statistics	p-Value for F-test	p-Value for K-S test
DEC RW	1	4.96	5.60	12.37	<0.01*	<0.01*
	2	5.98	6.19	1.67	0.20	0.60
	3	6.48	6.83	5.44	0.02*	0.17
DEC LC	1	3.93	4.44	8.08	<0.01*	0.03*
	2	5.24	5.25	0.00	0.95	0.65
	3	5.63	6.42	22.19	<0.01*	<0.01*
DEC AP	1	3.49	4.56	26.93	<0.01*	<0.01*
	2	6.01	6.24	2.09	0.15	0.42
	3	6.84	7.15	5.30	0.02*	0.38
DEC IW	1	5.75	6.58	28.66	<0.01*	<0.01*
	2	7.24	7.39	1.98	0.16	0.99
	3	7.48	7.66	3.25	0.07	0.94
VOC	K	13.64	13.82	3.94	0.05*	0.20
	1	13.43	13.90	6.23	0.01*	<0.01*
	2	13.98	14.29	6.86	<0.01*	0.09
	3	13.74	14.34	12.59	<0.01*	<0.01*
CS	1	8.29	8.17	0.22	0.64	0.79
	2	8.78	8.87	0.15	0.70	0.85
	3	9.40	10.11	9.37	<0.01*	<0.01*

* significant at $p < 0.05$.

Appendix 4. Concurrent Validity: Form- and Grade-Specific Results

Table 66. Concurrent Validity of DEC RW With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations With DIBELS Next at EOY	DIBELS Next at EOY (n)
All	A	0.62	70	0.66	417	0.66	408	0.62	9
	B	0.57	177	0.60	327	0.64	247	0.57	80
	C	0.21	133	0.58	323	0.65	182	0.41	141
	D	0.53	191	0.58	244	0.70	35	0.53	209
1	A	0.59	23	0.69	136	0.69	136	n/a	0
	B	0.56	79	0.73	93	0.77	67	0.76	26
	C	0.13	66	0.64	124	0.63	65	0.52	59
	D	0.64	67	0.74	82	n/a	0	0.74	82
2	A	0.17	23	0.58	144	0.58	143	n/a	1
	B	0.62	51	0.59	136	0.66	112	0.29	24
	C	0.05	38	0.50	105	0.59	58	0.12	47
	D	0.45	67	0.64	105	0.70	35	0.43	70
3	A	0.43	23	0.71	137	0.71	129	0.60	8
	B	0.30	46	0.63	98	0.68	68	0.36	30
	C	0.41	30	0.58	94	0.59	59	0.29	35
	D	0.43	56	0.31	57	n/a	0	0.31	57

Table 67. Concurrent Validity of DEC LC With SEL and DIBELS Next

Grade	Form	Correlations with SEL	SEL (n)	Correlations with DIBELS Next	DIBELS Next (n)	Correlations with DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations with DIBELS Next at EOY	DIBELS Next at EOY (n)
All	A	0.82	69	0.64	413	0.64	405	0.19	8
	B	0.73	176	0.64	327	0.63	249	0.73	78
	C	0.32	133	0.68	322	0.75	182	0.52	140
	D	0.47	189	0.54	248	0.76	35	0.53	213
1	A	0.69	24	0.75	136	0.75	136	n/a	0
	B	0.65	79	0.71	92	0.71	66	0.75	26
	C	0.33	63	0.76	121	0.71	65	0.70	56
	D	0.48	69	0.67	85	n/a	0	0.67	85
2	A	0.89	21	0.52	141	0.52	141	n/a	0
	B	0.77	50	0.62	138	0.68	115	0.66	23
	C	0.38	38	0.51	105	0.65	57	−0.06	48
	D	0.41	67	0.34	106	0.76	35	0.11	71
3	A	0.41	22	0.70	136	0.71	128	0.19	8
	B	0.51	46	0.74	97	0.78	68	0.34	29
	C	−0.10	32	0.62	96	0.63	60	0.16	36
	D	0.33	58	0.45	57	n/a	0	0.45	57

Table 68. Concurrent Validity of DEC AP With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations With DIBELS Next at EOY	DIBELS Next at EOY (n)
All	A	0.55	68	0.76	413	0.76	404	0.34	9
	B	0.71	175	0.69	329	0.70	251	0.76	78
	C	0.55	137	0.77	323	0.84	182	0.61	141
	D	0.57	192	0.67	246	0.89	35	0.64	211
1	A	0.55	22	0.80	135	0.80	135	n/a	0
	B	0.63	81	0.68	93	0.76	67	0.86	26
	C	0.57	66	0.88	122	0.87	65	0.72	57
	D	0.57	68	0.74	84	n/a	0	0.74	84
2	A	0.21	22	0.77	140	0.77	140	n/a	0
	B	0.38	48	0.68	139	0.68	116	0.08	23
	C	0.11	40	0.64	107	0.73	58	0.26	49
	D	0.34	68	0.67	105	0.89	35	0.39	70
3	A	−0.12	24	0.60	138	0.61	129	0.34	9
	B	0.42	47	0.74	97	0.78	68	0.49	29
	C	0.63	30	0.54	94	0.61	59	0.65	35
	D	0.44	54	0.27	57	n/a	0	0.27	57

Table 69. Concurrent Validity of DEC IW With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations With DIBELS Next at EOY	DIBELS Next at EOY (n)
All	A	0.47	68	0.64	414	0.64	405	n/a	9
	B	0.61	179	0.70	330	0.71	250	0.69	80
	C	0.32	133	0.61	323	0.66	183	0.40	140
	D	0.50	193	0.53	245	0.73	35	0.49	210
1	A	0.50	22	0.77	137	0.77	137	n/a	0
	B	0.59	81	0.74	93	0.77	66	0.82	27
	C	0.39	65	0.74	121	0.79	65	0.51	56
	D	0.44	70	0.55	84	n/a	0	0.55	84
2	A	0.06	22	0.58	141	0.59	140	n/a	1
	B	0.34	49	0.74	139	0.74	115	0.81	24
	C	0.04	38	0.54	105	0.53	58	0.27	47
	D	0.31	66	0.68	105	0.73	35	0.49	70
3	A	−0.04	24	0.53	136	0.54	128	n/a	8
	B	0.22	49	0.65	98	0.71	69	n/a	29
	C	0.63	32	0.38	97	0.25	60	0.24	37
	D	0.26	57	−0.09	56	n/a	0	−0.09	56

Table 70. Concurrent Validity of VOC With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations with DIBELS Next at EOY	DIBELS Next at EOY (n)
K	A	n/a	0	0.27	172	0.27	172	n/a	0
	B	0.19	65	0.12	172	0.07	140	0.22	32
	C	0.27	94	0.20	159	0.35	64	0.18	95
	D	0.25	84	0.31	149	0.34	35	0.15	114
1	A	0.11	23	0.34	140	0.33	139	n/a	1
	B	0.22	80	0.44	139	0.45	113	0.37	26
	C	0.18	66	0.29	123	0.34	66	0.25	57
	D	0.29	70	0.57	114	0.88	31	0.07	83
2	A	0.04	23	0.40	140	0.40	140	n/a	0
	B	0.31	52	0.41	139	0.41	115	0.01	24
	C	0.48	38	0.41	107	0.33	59	0.44	48
	D	0.28	67	0.27	105	0.40	35	0.06	70
3	A	0.08	23	0.67	136	0.66	128	0.47	8
	B	0.46	48	0.65	98	0.70	68	0.38	30
	C	0.86	33	0.36	96	0.27	60	0.51	36
	D	0.35	56	0.31	56	n/a	0	0.31	56

Table 71. Concurrent Validity of CS Fiction With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations With DIBELS Next at EOY	DIBELS Next at EOY (n)
1	A	0.52	19	0.63	111	0.63	111	n/a	0
	B	0.54	68	0.60	105	0.72	85	0.57	20
	C	0.49	28	0.48	89	0.57	40	0.44	49
	D	0.66	54	0.40	87	0.14	18	0.43	69
2	A	0.28	24	0.58	140	0.58	140	n/a	0
	B	0.32	50	0.61	131	0.70	107	0.04	24
	C	0.49	41	0.24	104	0.27	54	0.17	50
	D	0.46	64	0.49	102	0.68	35	0.18	67
3	A	0.07	24	0.77	137	0.78	129	0.55	8
	B	0.53	45	0.69	93	0.76	64	0.25	29
	C	0.59	30	0.63	93	0.57	58	0.71	35
	D	0.54	56	0.44	57	n/a	0	0.44	57

Table 72. Concurrent Validity of CS Nonfiction With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)	Correlations With DIBELS Next at MOY	DIBELS Next at MOY (n)	Correlations With DIBELS Next at EOY	DIBELS Next at EOY (n)
1	A	0.62	17	0.58	88	0.58	88	n/a	0
	B	0.56	66	0.60	103	0.73	84	0.55	19
	C	0.53	27	0.31	80	0.24	35	0.36	45
	D	0.55	56	0.47	89	0.43	20	0.47	69
2	A	0.85	19	0.52	133	0.52	133	n/a	0
	B	0.42	49	0.63	131	0.71	110	0.49	21
	C	0.16	38	0.06	100	0.35	53	–0.25	47
	D	0.34	67	0.45	105	0.70	36	0.11	69
3	A	0.28	24	0.81	127	0.81	119	0.78	8
	B	0.46	45	0.71	98	0.79	70	0.34	28
	C	0.55	27	0.35	91	0.26	60	0.21	31
	D	0.48	59	0.33	61	n/a	0	0.33	61

Appendix 5. Predictive Validity: Form- and Grade-Specific Results

Table 73. Predictive Validity of DEC RW With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	A	0.58	376	0.64	384
	B	0.61	193	0.61	206
	C	0.71	166	0.62	169
	D	0.50	145	n/a	0
1	A	0.52	132	0.70	129
	B	0.52	61	0.71	66
	C	0.76	63	0.59	61
	D	0.29	54	n/a	0
2	A	0.39	130	0.44	132
	B	0.58	67	0.66	74
	C	0.65	55	0.53	53
	D	0.33	41	n/a	0
3	A	0.63	114	0.78	117
	B	0.73	60	0.79	62
	C	0.57	49	0.61	50
	D	0.11	52	n/a	0

Table 74. Predictive Validity of DEC LC With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	A	0.52	377	0.61	378
	B	0.41	198	0.54	207
	C	0.71	168	0.73	166
	D	0.57	145	n/a	0
1	A	0.55	131	0.71	125
	B	0.37	60	0.64	64
	C	0.63	64	0.73	61
	D	0.49	52	n/a	0
2	A	0.37	128	0.48	132
	B	0.40	69	0.57	77
	C	0.73	55	0.61	52
	D	0.49	41	n/a	0
3	A	0.55	112	0.71	115
	B	0.50	61	0.71	64
	C	0.55	47	0.63	50
	D	0.24	51	n/a	0

Table 75. Predictive Validity of DEC AP With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	A	0.66	366	0.73	379
	B	0.60	198	0.65	209
	C	0.79	162	0.83	164
	D	0.65	143	n/a	0
1	A	0.47	132	0.80	125
	B	0.45	61	0.75	63
	C	0.70	64	0.85	59
	D	0.37	53	n/a	0
2	A	0.59	124	0.79	131
	B	0.52	70	0.66	78
	C	0.65	55	0.71	54
	D	0.38	42	n/a	0
3	A	0.55	113	0.62	118
	B	0.82	59	0.80	63
	C	0.45	49	0.55	52
	D	0.12	50	n/a	0

Table 76. Predictive Validity of DEC IW With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
All	A	0.63	371	0.65	382
	B	0.61	195	0.70	203
	C	0.67	167	0.69	169
	D	0.62	145	n/a	0
1	A	0.57	129	0.76	128
	B	0.55	60	0.79	62
	C	0.76	63	0.80	61
	D	0.47	54	n/a	0
2	A	0.45	126	0.69	132
	B	0.45	71	0.74	77
	C	0.40	55	0.44	54
	D	0.53	42	n/a	0
3	A	0.62	114	0.58	120
	B	0.74	61	0.76	65
	C	0.26	48	0.14	53
	D	−0.06	50	n/a	0

Table 77. Predictive Validity of VOC With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
K	A	0.39	158	0.26	155
	B	0.33	92	0.31	94
	C	0.32	33	0.06	58
	D	0.43	34	n/a	0
1	A	0.33	134	0.37	133
	B	0.40	77	0.27	82
	C	0.19	61	0.40	64
	D	0.13	53	n/a	0
2	A	0.36	127	0.35	129
	B	0.50	72	0.45	79
	C	0.09	54	0.18	53
	D	0.25	43	n/a	0
3	A	0.68	114	0.72	118
	B	0.86	62	0.74	65
	C	0.24	48	0.25	53
	D	0.16	49	n/a	0

Table 78. Predictive Validity of CS Fiction With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	A	0.51	112	0.58	107
	B	0.60	59	0.67	62
	C	0.30	38	0.48	38
	D	0.71	45	n/a	0
2	A	0.63	125	0.56	125
	B	0.72	62	0.79	67
	C	0.43	52	0.33	50
	D	0.05	41	n/a	0
3	A	0.69	114	0.75	117
	B	0.70	56	0.79	62
	C	0.49	45	0.50	51
	D	0.31	48	n/a	0

Table 79. Predictive Validity of CS Nonfiction With SEL and DIBELS Next

Grade	Form	Correlations With SEL	SEL (n)	Correlations With DIBELS Next	DIBELS Next (n)
1	A	0.50	85	0.42	82
	B	0.63	62	0.70	64
	C	0.52	34	0.33	34
	D	0.46	38	n/a	0
2	A	0.46	123	0.49	121
	B	0.69	68	0.73	74
	C	0.39	46	0.29	48
	D	0.24	41	n/a	0
3	A	0.69	107	0.74	108
	B	0.69	61	0.74	66
	C	0.43	47	0.15	54
	D	0.33	50	n/a	0

Appendix 6. Item Analysis by Form

Items in all of the Burst Reading assessments were analyzed using Item Response Theory (IRT; Embretson & Reise, 2000) in order to provide estimates of item difficulty, infit, and outfit statistics for the purpose of evaluating item performance. Infit statistics are more sensitive to unexpected patterns of observations by persons on items that roughly target their skill levels. Ideally, infit statistics should fall between 0.50 and 1.50 (Linacre, 2014). Values beyond that range suggest that the item might distort or degrade the measurement system. Outfit statistics are more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them. Ideally outfit statistics should fall between 0.50 and 1.50 (Linacre, 2014). Values beyond that range suggest that the item might distort or degrade the measurement system.

In addition, based on Classical Test Theory, p-values (mean item performance), Cronbach's alpha if the item is deleted, and adjusted point-biserial correlations are provided for each item. The point- biserial correlation is the Pearson correlation between the performance on an item and the person's raw scores or measures (or item marginal scores or measures). The adjusted point-biserial correlation is the point-biserial correlation excluding the current item from the raw score. Point-biserial correlations are crucial for evaluating whether the coding scheme and person responses accord with the requirement that "higher observations correspond to more of the latent variable" (and vice-versa; Linacre, 2014). The higher the value, the more consistent the item is with the rest of the test. Values lower than 0.20 indicate that an item is not very consistent and that high-ability people with score low on the item. The results suggest most of the items in Burst Reading Assessment perform well. Details of the item analyses for each form of each Burst Reading Assessment are provided in Table 80 to Table 135.

Table 80. DEC RW Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–0.76	0.79	0.81	0.62	0.79	0.50
Item 2	–2.08	0.77	0.57	0.79	0.79	0.53
Item 3	–1.73	0.66	0.51	0.75	0.77	0.61
Item 4	–0.69	0.79	0.75	0.61	0.79	0.51
Item 5	–2.13	0.79	0.64	0.80	0.79	0.51
Item 6	–2.35	0.62	0.41	0.82	0.78	0.61
Item 7	–1.81	0.88	0.75	0.76	0.80	0.46
Item 8	–1.21	0.86	0.78	0.68	0.79	0.48

Table 81. DEC RW Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.92	0.82	0.68	0.78	0.77	0.48
Item 2	–0.81	0.86	0.79	0.63	0.78	0.45
Item 3	–1.70	0.72	0.58	0.76	0.76	0.56
Item 4	–1.62	0.74	0.59	0.75	0.76	0.54
Item 5	–1.65	0.71	0.58	0.75	0.76	0.56
Item 6	–1.04	0.80	0.71	0.67	0.77	0.50
Item 7	–2.34	0.81	0.54	0.83	0.77	0.49
Item 8	–1.05	0.87	0.79	0.67	0.78	0.44

Table 82. DEC RW Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.14	0.80	0.59	0.82	0.73	0.46
Item 2	0.24	0.81	0.91	0.46	0.73	0.46
Item 3	–1.73	0.82	0.65	0.77	0.73	0.45
Item 4	–2.27	0.78	0.64	0.83	0.73	0.46
Item 5	–1.86	0.80	0.61	0.78	0.72	0.47
Item 6	–1.54	0.92	0.84	0.74	0.74	0.38
Item 7	–0.96	0.82	0.73	0.66	0.73	0.45
Item 8	–1.91	0.78	0.61	0.79	0.72	0.48

Table 83. DEC RW Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.92	0.76	0.65	0.78	0.79	0.52
Item 2	–2.44	0.65	0.47	0.83	0.78	0.58
Item 3	–1.98	0.79	0.62	0.79	0.79	0.50
Item 4	–1.27	0.70	0.63	0.70	0.78	0.56
Item 5	–1.79	0.81	0.72	0.76	0.79	0.47
Item 6	–2.26	0.78	0.61	0.82	0.79	0.50
Item 7	–1.50	0.81	0.72	0.73	0.80	0.47
Item 8	–1.53	0.66	0.59	0.73	0.78	0.59

Table 84. DEC LC Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.30	0.89	0.80	0.70	0.77	0.46
Item 2	–1.40	0.83	0.68	0.71	0.76	0.50
Item 3	–2.16	0.76	0.53	0.81	0.76	0.50
Item 4	–0.37	0.74	0.61	0.56	0.75	0.58
Item 5	–1.34	1.07	0.96	0.71	0.78	0.36
Item 6	0.45	0.81	0.72	0.43	0.76	0.52
Item 7	0.64	0.81	0.80	0.40	0.76	0.51
Item 8	–0.89	0.86	0.71	0.64	0.76	0.50

Table 85. DEC LC Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	0.26	0.79	0.65	0.46	0.80	0.59
Item 2	0.26	0.91	0.94	0.46	0.81	0.51
Item 3	–2.14	0.76	0.53	0.79	0.81	0.51
Item 4	–0.70	0.83	0.71	0.60	0.80	0.56
Item 5	0.36	0.88	0.86	0.44	0.81	0.53
Item 6	–1.23	0.81	0.63	0.68	0.80	0.56
Item 7	–1.03	0.84	0.70	0.65	0.80	0.55
Item 8	0.13	0.84	0.76	0.48	0.80	0.56

Table 86. DEC LC Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.01	1.07	1.02	0.67	0.76	0.32
Item 2	–0.73	0.85	0.76	0.62	0.74	0.47
Item 3	–1.12	0.72	0.60	0.68	0.72	0.56
Item 4	–0.54	0.79	0.70	0.59	0.73	0.51
Item 5	–0.54	0.81	0.72	0.59	0.73	0.49
Item 6	–0.51	0.81	0.74	0.59	0.73	0.49
Item 7	–1.81	0.86	0.85	0.78	0.74	0.43
Item 8	–2.64	0.78	0.62	0.87	0.75	0.44

Table 87. DEC LC Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–0.36	0.75	0.66	0.56	0.72	0.55
Item 2	–0.95	0.91	0.81	0.66	0.75	0.43
Item 3	–0.76	0.86	0.76	0.63	0.74	0.47
Item 4	–1.46	0.82	0.64	0.73	0.73	0.49
Item 5	–0.71	1.02	0.99	0.62	0.76	0.35
Item 6	–0.20	0.87	0.81	0.53	0.74	0.46
Item 7	–1.85	0.92	0.72	0.79	0.75	0.41
Item 8	–1.43	0.73	0.55	0.73	0.72	0.56

Table 88. DEC AP Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.87	0.66	0.47	0.74	0.87	0.70
Item 2	–2.04	0.58	0.39	0.76	0.87	0.74
Item 3	–1.54	0.58	0.47	0.71	0.87	0.74
Item 4	–2.09	0.55	0.40	0.76	0.87	0.75
Item 5	–1.36	0.59	0.47	0.69	0.87	0.73
Item 6	–3.42	0.87	0.52	0.87	0.89	0.54
Item 7	–1.57	0.77	0.64	0.71	0.88	0.64
Item 8	–0.12	0.90	0.91	0.54	0.89	0.52

Table 89. DEC AP Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–1.07	0.78	0.65	0.67	0.81	0.59
Item 2	0.05	0.73	0.62	0.50	0.81	0.58
Item 3	–0.62	0.92	0.86	0.60	0.83	0.48
Item 4	–0.96	0.60	0.49	0.65	0.80	0.69
Item 5	–2.05	0.77	0.58	0.78	0.81	0.58
Item 6	–1.93	0.65	0.43	0.77	0.80	0.66
Item 7	–0.89	1.28	1.27	0.64	0.85	0.30
Item 8	–2.09	0.59	0.39	0.79	0.80	0.68

Table 90. DEC AP Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.80	0.50	0.28	0.83	0.84	0.70
Item 2	–1.27	0.66	0.52	0.68	0.84	0.67
Item 3	–1.04	0.64	0.58	0.65	0.84	0.68
Item 4	–1.24	0.88	0.81	0.67	0.86	0.55
Item 5	–1.01	0.76	0.67	0.65	0.85	0.61
Item 6	–0.58	0.67	0.56	0.59	0.85	0.65
Item 7	–2.36	0.60	0.37	0.79	0.84	0.69
Item 8	–4.43	0.78	0.55	0.94	0.87	0.42

Table 91. DEC AP Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.24	0.69	0.45	0.79	0.85	0.64
Item 2	–0.75	0.72	0.58	0.61	0.85	0.66
Item 3	–1.33	0.90	0.70	0.69	0.86	0.58
Item 4	0.69	1.02	1.33	0.41	0.88	0.41
Item 5	–0.77	0.72	0.64	0.62	0.85	0.66
Item 6	–0.48	0.85	0.81	0.58	0.86	0.58
Item 7	–1.48	0.55	0.36	0.70	0.84	0.75
Item 8	–1.86	0.61	0.38	0.75	0.84	0.71

Table 92. DEC IW Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–5.31	0.58	0.38	0.97	0.83	0.39
Item 2	–2.03	0.64	0.64	0.76	0.81	0.57
Item 3	–2.51	0.48	0.42	0.81	0.80	0.68
Item 4	–4.06	0.54	0.27	0.92	0.81	0.56
Item 5	–1.60	0.58	0.55	0.72	0.79	0.70
Item 6	–3.62	0.47	0.25	0.89	0.80	0.64
Item 7	–5.72	0.59	0.14	0.98	0.84	0.35
Item 8	–3.32	0.48	0.31	0.87	0.80	0.65

Table 93. DEC IW Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–5.24	0.57	0.32	0.97	0.84	0.40
Item 2	–4.65	0.51	0.21	0.95	0.83	0.50
Item 3	–2.27	0.57	0.59	0.79	0.82	0.62
Item 4	–2.81	0.46	0.37	0.84	0.81	0.69
Item 5	–2.01	0.55	0.52	0.76	0.81	0.67
Item 6	–3.08	0.53	0.37	0.86	0.82	0.62
Item 7	–3.80	0.59	0.34	0.90	0.83	0.54
Item 8	–2.73	0.57	0.49	0.83	0.82	0.60

Table 94. DEC IW Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–3.78	0.69	0.51	0.93	0.76	0.47
Item 2	–4.00	0.62	0.32	0.94	0.76	0.51
Item 3	–0.50	0.97	1.09	0.59	0.81	0.32
Item 4	–2.09	0.51	0.45	0.80	0.73	0.63
Item 5	–3.62	0.63	0.38	0.92	0.76	0.53
Item 6	–2.11	0.66	0.58	0.80	0.75	0.51
Item 7	–3.90	0.63	0.31	0.93	0.76	0.52
Item 8	–2.35	0.57	0.47	0.83	0.74	0.60

Table 95. DEC IW Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.26	0.42	0.19	0.95	0.76	0.59
Item 2	–3.25	0.45	0.32	0.89	0.76	0.59
Item 3	–4.57	0.50	0.30	0.96	0.78	0.49
Item 4	–1.53	0.78	0.77	0.73	0.80	0.47
Item 5	–3.96	0.47	0.33	0.93	0.77	0.54
Item 6	–2.78	0.51	0.44	0.86	0.77	0.54
Item 7	–4.86	0.43	0.22	0.97	0.78	0.51
Item 8	–3.29	0.51	0.41	0.90	0.77	0.52

Table 96. VOC K Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.56	0.73	0.59	0.89	0.60	0.25
Item 2	–3.50	0.77	0.57	0.95	0.61	0.23
Item 3	–4.29	0.79	0.86	0.98	0.61	0.23
Item 4	–2.21	0.66	0.54	0.86	0.59	0.33
Item 5	–5.06	0.79	0.33	0.99	0.60	0.39
Item 6	–3.50	0.70	0.44	0.95	0.59	0.32
Item 7	–2.29	0.78	0.68	0.87	0.62	0.19
Item 8	–3.09	0.78	0.69	0.93	0.61	0.19
Item 9	–4.50	0.75	0.47	0.98	0.60	0.36
Item 10	–6.23	0.75	0.15	1.00	0.60	0.55
Item 11	–6.23	0.75	0.15	1.00	0.60	0.55
Item 12	–3.71	0.80	0.71	0.96	0.61	0.19
Item 13	–3.84	0.75	0.47	0.96	0.60	0.28
Item 14	–1.60	0.72	0.66	0.79	0.60	0.27
Item 15	–1.24	0.75	0.72	0.73	0.61	0.28

Table 97. VOC K Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.94	0.61	0.16	0.98	0.69	0.50
Item 2	–3.67	0.50	0.33	0.95	0.68	0.55
Item 3	–3.48	0.68	0.52	0.94	0.70	0.30
Item 4	–3.03	0.71	0.57	0.92	0.71	0.25
Item 5	–3.40	0.72	0.54	0.94	0.71	0.25
Item 6	–4.94	0.70	0.29	0.98	0.70	0.35
Item 7	–3.57	0.67	0.59	0.95	0.70	0.30
Item 8	–2.97	0.68	0.60	0.91	0.70	0.30
Item 9	–3.78	0.63	0.40	0.95	0.69	0.38
Item 10	–4.31	0.64	0.26	0.97	0.70	0.40
Item 11	–2.19	0.67	0.60	0.85	0.71	0.32
Item 12	–1.90	0.80	0.77	0.81	0.73	0.16
Item 13	–4.69	0.67	0.26	0.98	0.70	0.39
Item 14	–4.01	0.64	0.43	0.96	0.70	0.37
Item 15	–2.91	0.56	0.40	0.91	0.68	0.46

Table 98. VOC K Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.02	0.84	0.96	0.97	0.58	0.19
Item 2	–1.30	0.88	0.84	0.75	0.60	0.16
Item 3	–0.74	0.88	0.87	0.65	0.60	0.17
Item 4	–1.85	0.73	0.62	0.83	0.55	0.31
Item 5	–4.39	0.77	0.33	0.98	0.56	0.35
Item 6	–3.62	0.78	0.42	0.96	0.56	0.29
Item 7	–3.41	0.79	0.65	0.95	0.57	0.24
Item 8	–4.64	0.80	0.39	0.98	0.57	0.31
Item 9	–2.93	0.88	0.85	0.93	0.58	0.14
Item 10	–1.57	0.74	0.66	0.79	0.55	0.31
Item 11	–4.02	0.81	0.51	0.97	0.57	0.25
Item 12	–4.95	0.71	0.11	0.99	0.56	0.49
Item 13	–4.19	0.77	0.61	0.98	0.57	0.30
Item 14	–2.69	0.78	0.65	0.91	0.56	0.25
Item 15	–4.95	0.80	0.26	0.99	0.57	0.37

Table 99. VOC K Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.64	0.79	0.64	0.90	0.59	0.32
Item 2	–3.01	0.87	0.63	0.93	0.60	0.23
Item 3	–0.46	0.94	0.95	0.59	0.63	0.16
Item 4	–2.14	0.77	0.58	0.86	0.58	0.35
Item 5	–3.16	0.80	0.49	0.94	0.59	0.30
Item 6	–3.16	0.74	0.45	0.94	0.58	0.38
Item 7	–2.76	0.82	0.61	0.91	0.59	0.29
Item 8	–4.64	0.80	0.32	0.98	0.60	0.30
Item 9	–2.64	0.78	0.50	0.90	0.58	0.34
Item 10	–2.82	1.07	1.67	0.92	0.64	–0.03
Item 11	–0.35	0.90	0.90	0.57	0.62	0.21
Item 12	–3.75	0.85	0.46	0.96	0.60	0.25
Item 13	–5.38	0.72	0.09	0.99	0.60	0.44
Item 14	–2.59	0.86	0.74	0.90	0.60	0.24
Item 15	–3.01	0.75	0.43	0.93	0.58	0.37

Table 100. VOC G1 Form A Item Analysis.

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–3.08	0.55	0.44	0.90	0.90	0.59
Item 2	–4.81	0.40	0.13	0.97	0.90	0.72
Item 3	–5.26	0.28	0.04	0.98	0.90	0.77
Item 4	–3.50	0.45	0.27	0.92	0.89	0.70
Item 5	–3.80	0.54	0.28	0.93	0.90	0.64
Item 6	–4.46	0.29	0.08	0.96	0.89	0.83
Item 7	–2.00	0.71	0.80	0.81	0.91	0.36
Item 8	–3.92	0.56	0.34	0.94	0.90	0.62
Item 9	–4.31	0.39	0.25	0.95	0.89	0.75
Item 10	–2.94	0.69	0.56	0.89	0.90	0.47
Item 11	–3.69	0.54	0.36	0.93	0.90	0.63
Item 12	–3.50	0.57	0.37	0.92	0.90	0.60
Item 13	–2.74	0.52	0.45	0.87	0.90	0.60
Item 14	–3.80	0.40	0.24	0.93	0.89	0.75
Item 15	–3.92	0.83	0.63	0.94	0.90	0.41

Table 101. VOC G1 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.47	0.70	0.28	0.96	0.88	0.55
Item 2	–4.04	0.61	0.50	0.94	0.88	0.61
Item 3	–3.15	0.71	0.52	0.90	0.88	0.52
Item 4	–5.54	0.33	0.04	0.98	0.88	0.62
Item 5	–3.41	0.39	0.21	0.91	0.87	0.78
Item 6	–4.64	0.58	0.17	0.96	0.88	0.62
Item 7	–3.00	0.65	0.50	0.89	0.88	0.57
Item 8	–3.41	0.63	0.58	0.91	0.88	0.59
Item 9	–1.40	0.74	0.77	0.74	0.89	0.40
Item 10	–4.18	0.70	0.53	0.95	0.88	0.55
Item 11	–2.67	0.62	0.53	0.87	0.88	0.56
Item 12	–3.41	0.40	0.25	0.91	0.87	0.77
Item 13	–5.03	0.53	0.09	0.97	0.88	0.65
Item 14	–5.03	1.08	1.34	0.97	0.89	0.25
Item 15	–2.73	0.51	0.40	0.87	0.87	0.65

Table 102. VOC G1 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–2.20	0.71	0.63	0.83	0.89	0.48
Item 2	–4.14	0.47	0.30	0.95	0.88	0.72
Item 3	–4.28	0.38	0.16	0.95	0.88	0.77
Item 4	–1.63	0.68	0.66	0.77	0.90	0.43
Item 5	–3.56	0.56	0.41	0.92	0.88	0.67
Item 6	–3.05	0.81	0.73	0.90	0.89	0.47
Item 7	–3.77	0.60	0.33	0.93	0.88	0.66
Item 8	–3.88	0.57	0.25	0.94	0.88	0.67
Item 9	–1.55	0.74	0.70	0.76	0.90	0.41
Item 10	–3.88	0.57	0.26	0.94	0.88	0.68
Item 11	–3.77	0.85	0.69	0.93	0.89	0.49
Item 12	–4.60	0.50	0.17	0.96	0.88	0.67
Item 13	–3.88	0.47	0.20	0.94	0.88	0.74
Item 14	–3.56	0.69	0.50	0.92	0.88	0.59
Item 15	–4.01	0.69	0.46	0.94	0.88	0.59

Table 103. VOC G1 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–3.52	0.53	0.24	0.91	0.91	0.79
Item 2	–3.75	0.72	0.49	0.92	0.91	0.69
Item 3	–3.22	0.66	0.45	0.90	0.91	0.73
Item 4	–3.63	0.59	0.21	0.92	0.91	0.76
Item 5	–3.87	0.59	0.69	0.93	0.91	0.73
Item 6	–1.56	0.97	1.15	0.76	0.92	0.37
Item 7	–0.06	0.88	0.84	0.55	0.93	0.35
Item 8	–2.80	0.77	0.54	0.87	0.91	0.65
Item 9	–3.42	1.33	0.67	0.91	0.91	0.46
Item 10	–2.21	0.88	0.74	0.83	0.91	0.52
Item 11	–3.63	0.51	0.56	0.92	0.91	0.78
Item 12	–3.42	0.66	0.45	0.91	0.91	0.73
Item 13	–2.80	0.74	0.75	0.87	0.91	0.67
Item 14	–3.87	0.46	0.20	0.93	0.91	0.80
Item 15	–3.42	0.33	0.13	0.91	0.90	0.88

Table 104. VOC G2 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.51	0.53	0.13	0.98	0.67	0.55
Item 2	–3.36	0.67	0.52	0.94	0.69	0.32
Item 3	–5.54	0.67	0.12	0.99	0.68	0.50
Item 4	–1.64	0.70	0.68	0.78	0.71	0.26
Item 5	–0.89	0.84	0.89	0.66	0.73	0.22
Item 6	–4.29	0.53	0.16	0.97	0.67	0.53
Item 7	–3.68	0.64	0.37	0.95	0.68	0.37
Item 8	–4.51	0.53	0.13	0.98	0.67	0.55
Item 9	–3.26	0.70	0.57	0.93	0.69	0.28
Item 10	–5.54	0.70	0.32	0.99	0.69	0.44
Item 11	–3.02	0.83	0.74	0.92	0.71	0.14
Item 12	–5.09	0.57	0.10	0.98	0.68	0.56
Item 13	–5.09	0.67	0.18	0.98	0.68	0.44
Item 14	–5.54	0.70	0.32	0.99	0.69	0.44
Item 15	–3.81	0.64	0.39	0.96	0.68	0.37

Table 105. VOC G2 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–3.68	0.51	0.40	0.93	0.85	0.51
Item 2	–3.57	0.51	0.40	0.92	0.85	0.51
Item 3	–4.21	0.55	0.45	0.95	0.85	0.46
Item 4	–4.76	0.36	0.11	0.97	0.85	0.69
Item 5	–4.76	0.77	0.51	0.97	0.86	0.23
Item 6	–2.77	0.55	0.51	0.87	0.86	0.48
Item 7	–4.38	0.39	0.17	0.95	0.84	0.66
Item 8	–4.56	0.44	0.26	0.96	0.85	0.59
Item 9	–4.56	0.39	0.17	0.96	0.85	0.64
Item 10	–6.08	0.44	0.05	0.99	0.85	0.58
Item 11	–3.80	0.49	0.30	0.93	0.85	0.53
Item 12	–3.09	0.47	0.38	0.89	0.85	0.56
Item 13	–6.08	0.59	0.67	0.99	0.86	0.34
Item 14	–6.08	0.39	0.04	0.99	0.85	0.60
Item 15	–3.57	0.49	0.41	0.92	0.85	0.54

Table 106. VOC G2 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–5.21	0.80	0.87	0.99	0.72	0.39
Item 2	–5.21	0.72	0.12	0.99	0.72	0.56
Item 3	–3.59	0.67	0.37	0.95	0.71	0.41
Item 4	–1.68	0.69	0.65	0.78	0.72	0.37
Item 5	–3.44	0.77	0.62	0.94	0.73	0.27
Item 6	–2.97	0.63	0.40	0.91	0.71	0.42
Item 7	–2.54	0.70	0.59	0.88	0.72	0.33
Item 8	–3.94	0.78	0.58	0.96	0.73	0.28
Item 9	–3.94	0.71	0.45	0.96	0.72	0.39
Item 10	–4.75	0.76	0.36	0.98	0.72	0.40
Item 11	–2.54	0.70	0.54	0.88	0.72	0.33
Item 12	–2.78	0.69	0.64	0.90	0.72	0.34
Item 13	–2.25	0.69	0.55	0.85	0.72	0.33
Item 14	–3.31	0.67	0.52	0.93	0.71	0.38
Item 15	–3.75	0.75	0.72	0.95	0.72	0.30

Table 107. VOC G2 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.32	0.52	0.28	0.96	0.77	0.55
Item 2	–5.17	0.34	0.04	0.98	0.76	0.74
Item 3	–4.13	0.47	0.46	0.96	0.76	0.59
Item 4	–3.53	0.63	0.42	0.94	0.77	0.42
Item 5	–3.95	0.85	0.55	0.95	0.79	0.23
Item 6	–1.51	0.73	0.74	0.75	0.80	0.30
Item 7	–4.55	0.38	0.17	0.97	0.76	0.68
Item 8	–3.66	0.52	0.35	0.94	0.76	0.55
Item 9	–5.17	0.34	0.04	0.98	0.76	0.74
Item 10	–5.17	0.51	0.24	0.98	0.77	0.54
Item 11	–4.82	0.49	0.10	0.98	0.77	0.62
Item 12	–3.95	0.76	0.49	0.95	0.78	0.31
Item 13	–1.51	0.77	0.87	0.75	0.81	0.22
Item 14	–4.55	0.81	0.57	0.97	0.78	0.28
Item 15	–3.11	0.66	0.43	0.91	0.78	0.36

Table 108. VOC G3 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–6.96	0.31	0.02	0.99	0.85	0.38
Item 2	–3.78	0.46	0.33	0.93	0.83	0.74
Item 3	–3.67	0.82	0.79	0.93	0.84	0.52
Item 4	–5.11	0.52	0.84	0.97	0.85	0.55
Item 5	–5.74	0.30	0.03	0.98	0.85	0.58
Item 6	–3.47	0.78	0.46	0.92	0.84	0.56
Item 7	–3.57	0.59	0.32	0.92	0.84	0.67
Item 8	–4.49	0.49	0.11	0.96	0.84	0.70
Item 9	–4.03	0.93	0.66	0.94	0.85	0.46
Item 10	–1.97	0.60	0.47	0.80	0.84	0.56
Item 11	–3.47	0.61	0.47	0.92	0.84	0.65
Item 12	–3.19	0.67	0.39	0.90	0.84	0.61
Item 13	0.20	1.11	2.21	0.49	0.88	0.17
Item 14	–3.03	0.61	0.35	0.89	0.84	0.65
Item 15	–3.11	1.03	0.67	0.90	0.85	0.39

Table 109. VOC G3 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–4.72	0.58	0.10	0.96	0.88	0.67
Item 2	–3.53	0.62	0.43	0.92	0.88	0.68
Item 3	–3.25	0.43	0.21	0.90	0.87	0.79
Item 4	–4.23	0.35	0.08	0.95	0.87	0.81
Item 5	–3.68	0.74	0.38	0.93	0.88	0.61
Item 6	–2.41	0.65	0.45	0.84	0.88	0.61
Item 7	–3.13	0.67	0.45	0.90	0.88	0.63
Item 8	–3.53	1.01	0.82	0.92	0.88	0.46
Item 9	–3.53	0.80	0.59	0.92	0.88	0.58
Item 10	–3.13	0.72	0.58	0.90	0.88	0.61
Item 11	0.17	1.10	1.85	0.50	0.91	0.20
Item 12	–2.50	0.71	0.42	0.85	0.88	0.59
Item 13	–4.23	0.85	1.14	0.95	0.88	0.51
Item 14	–4.72	0.78	0.44	0.96	0.88	0.51
Item 15	–3.68	0.68	0.35	0.93	0.88	0.64

Table 110. VOC G3 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–5.52	0.46	0.06	0.99	0.88	0.84
Item 2	–4.10	0.43	0.36	0.96	0.88	0.62
Item 3	–3.01	0.50	0.43	0.91	0.89	0.48
Item 4	–4.34	0.38	0.18	0.96	0.88	0.71
Item 5	–4.34	0.54	0.37	0.96	0.89	0.54
Item 6	–3.53	0.46	0.35	0.94	0.89	0.55
Item 7	–3.01	0.45	0.37	0.91	0.89	0.52
Item 8	–3.38	0.52	0.48	0.93	0.89	0.47
Item 9	–5.52	0.46	0.06	0.99	0.88	0.84
Item 10	–2.61	0.69	0.67	0.88	0.91	0.25
Item 11	–5.52	0.46	0.06	0.99	0.88	0.84
Item 12	–4.34	0.35	0.13	0.96	0.88	0.76
Item 13	–5.01	0.45	0.11	0.98	0.88	0.75
Item 14	–4.64	0.49	0.31	0.97	0.88	0.64
Item 15	–4.34	0.33	0.12	0.96	0.88	0.78

Table 111. VOC G3 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Item 1	–5.35	0.38	0.13	0.98	0.96	0.96
Item 2	–4.45	0.29	0.16	0.97	0.96	0.78
Item 3	–4.83	0.37	0.24	0.98	0.96	0.78
Item 4	–5.35	0.38	0.13	0.98	0.96	0.96
Item 5	–5.35	0.38	0.13	0.98	0.96	0.96
Item 6	–5.35	0.38	0.13	0.98	0.96	0.96
Item 7	–5.35	0.38	0.13	0.98	0.96	0.96
Item 8	–5.35	0.38	0.13	0.98	0.96	0.96
Item 9	–4.45	0.33	0.23	0.97	0.96	0.73
Item 10	–5.35	0.38	0.13	0.98	0.96	0.96
Item 11	–4.83	0.37	0.24	0.98	0.96	0.78
Item 12	–4.45	0.33	0.23	0.97	0.96	0.73
Item 13	–2.90	0.52	0.51	0.91	0.98	0.44
Item 14	–4.83	0.37	0.24	0.98	0.96	0.78
Item 15	–5.35	0.38	0.13	0.98	0.96	0.96

Table 112. CS Fiction G1 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.48	1.03	1.03	0.73	0.81	0.44
Literal Item 2	0.44	0.89	0.76	0.43	0.80	0.51
Literal Item 3	–1.27	0.89	0.72	0.70	0.80	0.52
Literal Item 4	–1.52	0.82	0.77	0.73	0.80	0.54
Literal Item 5	–2.67	0.67	0.65	0.86	0.81	0.50
Inferential Item 1	–0.30	0.79	0.65	0.55	0.80	0.59
Inferential Item 2	0.99	0.78	0.67	0.35	0.80	0.54
Inferential Item 3	0.96	0.86	0.83	0.35	0.81	0.48
Inferential Item 4	0.23	0.86	0.73	0.47	0.80	0.55
Inferential Item 5	–1.27	1.09	1.45	0.70	0.82	0.39

Table 113. CS Fiction G1 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.98	0.90	0.96	0.78	0.82	0.51
Literal Item 2	1.70	0.82	1.36	0.25	0.83	0.45
Literal Item 3	–0.70	0.84	0.73	0.61	0.81	0.57
Literal Item 4	–1.38	0.88	0.78	0.71	0.82	0.54
Literal Item 5	–1.16	0.96	0.79	0.68	0.82	0.51
Inferential Item 1	0.07	0.78	0.66	0.49	0.81	0.57
Inferential Item 2	0.84	0.85	0.70	0.37	0.82	0.50
Inferential Item 3	–1.47	0.78	0.60	0.72	0.81	0.59
Inferential Item 4	–0.08	1.02	0.87	0.52	0.82	0.47
Inferential Item 5	–1.76	0.84	0.73	0.76	0.82	0.54

Table 114. CS Fiction G1 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.77	0.69	0.36	0.87	0.82	0.64
Literal Item 2	–2.3	0.61	0.46	0.83	0.81	0.68
Literal Item 3	–2.39	0.65	0.50	0.83	0.81	0.66
Literal Item 4	–2.48	0.72	0.44	0.84	0.82	0.63
Literal Item 5	–2.77	0.66	0.31	0.87	0.82	0.66
Inferential Item 1	–2.87	1.04	0.84	0.87	0.83	0.45
Inferential Item 2	0.23	1.02	1.05	0.48	0.85	0.30
Inferential Item 3	–1.38	0.90	0.76	0.72	0.83	0.48
Inferential Item 4	–1.77	0.83	0.67	0.77	0.82	0.54
Inferential Item 5	–0.71	0.84	0.74	0.63	0.83	0.47

Table 115. CS Fiction G1 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.97	1.04	0.84	0.77	0.85	0.49
Literal Item 2	–0.50	0.95	0.86	0.58	0.85	0.56
Literal Item 3	–1.55	0.71	0.56	0.72	0.84	0.65
Literal Item 4	–1.28	0.78	0.87	0.69	0.84	0.62
Literal Item 5	–0.64	0.63	0.45	0.60	0.83	0.71
Inferential Item 1	0.72	0.76	0.56	0.40	0.84	0.59
Inferential Item 2	2.27	0.90	0.79	0.21	0.86	0.41
Inferential Item 3	–0.83	1.07	0.81	0.63	0.85	0.52
Inferential Item 4	1.33	0.80	1.22	0.32	0.85	0.49
Inferential Item 5	–1.73	0.72	0.93	0.74	0.84	0.63

Table 116. CS Fiction G2 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.83	0.92	0.59	0.87	0.80	0.50
Literal Item 2	–2.35	0.69	0.43	0.83	0.79	0.62
Literal Item 3	–2.97	0.62	0.34	0.88	0.79	0.61
Literal Item 4	–0.72	0.81	0.73	0.63	0.79	0.53
Literal Item 5	–2.70	0.81	0.80	0.86	0.79	0.53
Inferential Item 1	0.84	0.91	0.87	0.37	0.81	0.37
Inferential Item 2	0.00	0.91	0.93	0.51	0.81	0.42
Inferential Item 3	–0.78	0.96	0.91	0.64	0.80	0.46
Inferential Item 4	–1.05	0.77	0.57	0.68	0.79	0.58
Inferential Item 5	–1.19	1.00	0.91	0.70	0.80	0.46

Table 117. CS Fiction G2 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.44	0.93	1.16	0.85	0.79	0.41
Literal Item 2	–1.62	0.88	0.67	0.76	0.78	0.50
Literal Item 3	–2.08	0.66	0.49	0.81	0.77	0.61
Literal Item 4	–0.68	0.75	0.62	0.62	0.77	0.57
Literal Item 5	–1.48	0.88	0.68	0.74	0.78	0.50
Inferential Item 1	–0.90	0.86	0.74	0.66	0.78	0.50
Inferential Item 2	0.35	1.00	0.92	0.45	0.80	0.36
Inferential Item 3	–1.71	0.83	0.61	0.77	0.78	0.53
Inferential Item 4	–1.06	0.91	0.77	0.68	0.79	0.47
Inferential Item 5	0.07	1.00	1.02	0.49	0.80	0.35

Table 118. CS Fiction G2 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.56	0.76	0.54	0.76	0.74	0.55
Literal Item 2	–1.61	0.82	0.61	0.76	0.75	0.49
Literal Item 3	–2.00	0.96	1.01	0.81	0.77	0.34
Literal Item 4	–0.75	0.77	0.68	0.63	0.74	0.56
Literal Item 5	–1.37	0.94	0.83	0.73	0.76	0.43
Inferential Item 1	0.23	1.18	1.22	0.46	0.78	0.27
Inferential Item 2	0.08	0.67	0.56	0.49	0.73	0.62
Inferential Item 3	–0.16	0.83	0.74	0.53	0.75	0.52
Inferential Item 4	–0.39	1.04	0.99	0.57	0.77	0.38
Inferential Item 5	–0.67	1.14	1.07	0.62	0.78	0.31

Table 119. CS Fiction G2 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.98	0.64	0.31	0.90	0.73	0.56
Literal Item 2	–2.81	0.69	0.65	0.89	0.73	0.52
Literal Item 3	–2.09	0.86	0.83	0.83	0.73	0.45
Literal Item 4	–1.88	0.76	0.64	0.80	0.72	0.54
Literal Item 5	–2.45	1.00	0.75	0.86	0.75	0.36
Inferential Item 1	–0.79	0.95	0.87	0.64	0.75	0.35
Inferential Item 2	0.26	0.97	0.99	0.46	0.76	0.29
Inferential Item 3	–1.05	1.08	1.07	0.69	0.76	0.27
Inferential Item 4	–1.13	0.77	0.71	0.70	0.72	0.51
Inferential Item 5	–2.32	0.71	0.52	0.85	0.72	0.55

Table 120. CS Fiction G3 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–3.65	0.58	0.21	0.92	0.84	0.64
Literal Item 2	–2.73	0.75	0.48	0.86	0.84	0.61
Literal Item 3	–2.86	0.64	0.51	0.86	0.84	0.65
Literal Item 4	–2.43	0.80	0.60	0.83	0.84	0.59
Literal Item 5	–1.17	0.78	0.70	0.69	0.84	0.55
Inferential Item 1	–1.78	0.80	0.61	0.76	0.84	0.57
Inferential Item 2	–1.25	0.86	0.77	0.70	0.85	0.50
Inferential Item 3	–0.84	0.81	0.74	0.64	0.85	0.51
Inferential Item 4	–3.07	0.58	0.36	0.88	0.83	0.67
Inferential Item 5	–0.37	0.80	0.73	0.57	0.85	0.49

Table 121. CS Fiction G3 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.63	0.61	0.36	0.84	0.88	0.73
Literal Item 2	–2.54	0.70	0.72	0.83	0.88	0.67
Literal Item 3	–1.30	0.96	0.88	0.71	0.89	0.51
Literal Item 4	–0.36	0.91	1.09	0.58	0.90	0.43
Literal Item 5	–2.54	0.63	0.41	0.83	0.88	0.72
Inferential Item 1	–2.74	0.52	0.26	0.84	0.88	0.78
Inferential Item 2	–2.10	0.59	0.37	0.79	0.88	0.74
Inferential Item 3	–1.43	0.55	0.45	0.72	0.88	0.73
Inferential Item 4	–1.17	0.73	0.64	0.69	0.88	0.62
Inferential Item 5	–1.57	0.88	0.73	0.74	0.89	0.56

Table 122. CS Fiction G3 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–3.66	0.58	0.57	0.93	0.76	0.60
Literal Item 2	–2.22	0.76	0.64	0.83	0.76	0.49
Literal Item 3	–1.49	0.72	0.61	0.74	0.76	0.50
Literal Item 4	–3.24	0.64	0.37	0.91	0.76	0.59
Literal Item 5	–1.28	0.80	0.72	0.72	0.78	0.41
Inferential Item 1	–3.01	0.80	0.49	0.90	0.77	0.50
Inferential Item 2	–1.65	0.78	0.67	0.77	0.77	0.46
Inferential Item 3	–1.71	0.76	0.63	0.77	0.77	0.48
Inferential Item 4	–0.67	0.96	0.93	0.62	0.80	0.28
Inferential Item 5	–2.37	0.77	0.68	0.85	0.77	0.48

Table 123. CS Fiction G3 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.15	0.84	0.60	0.81	0.79	0.52
Literal Item 2	–2.23	0.85	1.11	0.82	0.80	0.48
Literal Item 3	–0.96	0.94	0.95	0.66	0.80	0.49
Literal Item 4	–2.54	0.93	1.77	0.85	0.81	0.40
Literal Item 5	–2.08	0.79	0.61	0.80	0.79	0.54
Inferential Item 1	0.29	0.92	0.77	0.46	0.80	0.49
Inferential Item 2	–1.51	0.70	0.51	0.73	0.78	0.63
Inferential Item 3	0.52	0.74	0.63	0.42	0.79	0.54
Inferential Item 4	1.17	0.77	0.62	0.32	0.80	0.49
Inferential Item 5	–0.56	1.10	1.18	0.60	0.81	0.39

Table 124. CS Nonfiction G1 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	−0.76	0.85	0.72	0.64	0.77	0.53
Literal Item 2	−1.08	0.81	0.64	0.69	0.77	0.56
Literal Item 3	−0.01	0.89	0.80	0.51	0.78	0.48
Literal Item 4	−2.15	0.89	0.63	0.83	0.78	0.46
Literal Item 5	−0.38	0.93	0.79	0.57	0.78	0.47
Inferential Item 1	−0.21	0.85	0.75	0.54	0.77	0.52
Inferential Item 2	−0.38	0.99	0.94	0.57	0.78	0.43
Inferential Item 3	−0.59	1.02	1.06	0.61	0.79	0.41
Inferential Item 4	0.24	1.06	0.93	0.46	0.79	0.37
Inferential Item 5	−0.90	0.94	0.82	0.66	0.78	0.47

Table 125. CS Nonfiction G1 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	−0.24	1.14	1.09	0.56	0.88	0.41
Literal Item 2	−3.34	0.57	0.24	0.89	0.86	0.65
Literal Item 3	−2.40	0.65	0.41	0.82	0.85	0.69
Literal Item 4	−1.58	0.61	0.38	0.74	0.85	0.74
Literal Item 5	−1.87	0.77	0.44	0.77	0.85	0.67
Inferential Item 1	−2.00	0.91	0.74	0.78	0.86	0.59
Inferential Item 2	−0.63	0.75	0.63	0.62	0.86	0.63
Inferential Item 3	1.21	0.84	0.88	0.35	0.87	0.43
Inferential Item 4	−1.16	0.92	0.74	0.69	0.86	0.59
Inferential Item 5	−0.82	0.79	0.66	0.64	0.86	0.63

Table 126. CS Nonfiction G1 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.18	0.68	0.45	0.81	0.83	0.65
Literal Item 2	–0.92	0.76	0.61	0.65	0.83	0.60
Literal Item 3	–1.50	0.73	0.51	0.73	0.83	0.63
Literal Item 4	–1.64	0.70	0.55	0.75	0.83	0.64
Literal Item 5	–1.43	0.96	0.86	0.72	0.84	0.48
Inferential Item 1	–0.74	0.89	0.78	0.63	0.84	0.49
Inferential Item 2	–2.46	0.80	0.63	0.83	0.83	0.56
Inferential Item 3	–0.63	0.87	0.70	0.61	0.84	0.51
Inferential Item 4	–1.04	0.91	0.76	0.67	0.84	0.51
Inferential Item 5	–0.24	0.95	0.87	0.55	0.84	0.46

Table 127. CS Nonfiction G1 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.56	0.67	0.93	0.73	0.83	0.58
Literal Item 2	–0.07	0.83	0.63	0.51	0.82	0.59
Literal Item 3	1.21	0.89	0.72	0.32	0.83	0.52
Literal Item 4	0.24	0.93	0.77	0.47	0.83	0.53
Literal Item 5	–0.57	0.81	0.65	0.59	0.83	0.58
Inferential Item 1	–0.75	1.00	0.96	0.62	0.83	0.48
Inferential Item 2	1.31	0.91	1.19	0.31	0.83	0.48
Inferential Item 3	1.01	0.88	0.77	0.35	0.83	0.53
Inferential Item 4	1.16	0.82	0.64	0.33	0.83	0.56
Inferential Item 5	0.46	0.92	0.76	0.43	0.83	0.54

Table 128. CS Nonfiction G2 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.70	0.96	0.71	0.76	0.78	0.45
Literal Item 2	–2.97	0.84	0.84	0.89	0.79	0.40
Literal Item 3	–2.23	0.69	0.38	0.82	0.77	0.58
Literal Item 4	–2.60	0.92	0.67	0.86	0.79	0.41
Literal Item 5	–2.00	0.80	0.68	0.79	0.78	0.52
Inferential Item 1	–0.54	1.03	0.94	0.59	0.79	0.42
Inferential Item 2	0.80	0.82	0.98	0.37	0.78	0.45
Inferential Item 3	0.16	0.82	0.75	0.47	0.78	0.51
Inferential Item 4	–1.66	0.87	0.72	0.75	0.78	0.50
Inferential Item 5	–0.34	0.89	0.93	0.56	0.78	0.50

Table 129. CS Nonfiction G2 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.58	1.29	1.44	0.76	0.78	0.22
Literal Item 2	–0.79	0.83	0.64	0.64	0.74	0.54
Literal Item 3	–1.37	0.69	0.63	0.73	0.73	0.60
Literal Item 4	–0.79	0.87	0.77	0.64	0.74	0.51
Literal Item 5	–2.57	0.95	0.81	0.87	0.76	0.34
Inferential Item 1	–0.69	1.03	0.90	0.62	0.76	0.40
Inferential Item 2	–1.71	0.90	1.34	0.77	0.76	0.40
Inferential Item 3	0.27	0.85	0.84	0.45	0.75	0.48
Inferential Item 4	1.53	0.82	0.85	0.25	0.76	0.39
Inferential Item 5	–0.04	0.89	0.78	0.51	0.74	0.49

Table 130. CS Nonfiction G2 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	0.11	0.88	0.86	0.48	0.72	0.42
Literal Item 2	–0.01	0.95	0.84	0.50	0.73	0.39
Literal Item 3	–1.86	0.81	0.57	0.80	0.71	0.49
Literal Item 4	–1.68	0.91	0.83	0.78	0.72	0.42
Literal Item 5	–3.18	1.10	2.14	0.92	0.75	0.14
Inferential Item 1	1.00	0.81	0.66	0.33	0.72	0.43
Inferential Item 2	–2.12	0.69	0.51	0.83	0.71	0.54
Inferential Item 3	–1.57	1.06	1.09	0.76	0.74	0.31
Inferential Item 4	–0.25	0.89	0.81	0.55	0.72	0.43
Inferential Item 5	–0.63	0.86	0.75	0.61	0.71	0.46

Table 131. CS Nonfiction G2 Form D Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.29	0.86	0.77	0.71	0.77	0.48
Literal Item 2	–1.85	0.99	0.95	0.79	0.78	0.34
Literal Item 3	–2.31	0.97	1.16	0.84	0.79	0.29
Literal Item 4	–1.21	0.87	0.77	0.70	0.77	0.48
Literal Item 5	–1.06	0.82	0.66	0.67	0.76	0.52
Inferential Item 1	0.27	0.97	1.00	0.45	0.77	0.41
Inferential Item 2	–0.24	0.80	0.76	0.54	0.76	0.54
Inferential Item 3	0.14	0.94	0.92	0.47	0.77	0.45
Inferential Item 4	0.48	0.81	0.70	0.41	0.76	0.53
Inferential Item 5	–0.07	0.87	0.91	0.51	0.76	0.49

Table 132. CS Nonfiction G3 Form A Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–2.87	0.70	0.37	0.87	0.81	0.58
Literal Item 2	–2.94	0.65	0.35	0.88	0.81	0.59
Literal Item 3	–1.27	0.92	0.85	0.71	0.81	0.51
Literal Item 4	–2.61	0.70	0.40	0.85	0.81	0.60
Literal Item 5	–0.69	0.80	0.69	0.62	0.81	0.56
Inferential Item 1	–0.13	0.87	0.78	0.53	0.82	0.49
Inferential Item 2	–0.26	1.03	1.03	0.55	0.83	0.39
Inferential Item 3	–1.20	0.91	0.81	0.69	0.81	0.51
Inferential Item 4	–0.87	0.81	0.66	0.65	0.81	0.56
Inferential Item 5	–0.46	0.89	0.80	0.58	0.82	0.50

Table 133. CS Nonfiction G3 Form B Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.23	1.28	1.34	0.68	0.87	0.32
Literal Item 2	–1.76	0.69	0.49	0.75	0.84	0.65
Literal Item 3	–1.12	0.80	0.63	0.66	0.85	0.59
Literal Item 4	–1.76	0.76	0.52	0.75	0.85	0.61
Literal Item 5	–2.02	0.74	0.57	0.78	0.85	0.60
Inferential Item 1	–0.75	0.76	0.64	0.61	0.85	0.61
Inferential Item 2	–0.65	0.74	0.71	0.60	0.85	0.61
Inferential Item 3	–0.19	0.88	0.84	0.53	0.86	0.50
Inferential Item 4	–1.46	0.71	0.53	0.71	0.84	0.65
Inferential Item 5	–1.76	0.73	0.54	0.75	0.85	0.63

Table 134. CS Nonfiction G3 Form C Item Analysis

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.85	0.86	0.68	0.79	0.79	0.48
Literal Item 2	–1.20	0.84	0.72	0.7	0.79	0.50
Literal Item 3	–2.11	0.81	0.67	0.81	0.79	0.49
Literal Item 4	–2.32	0.86	0.71	0.84	0.79	0.47
Literal Item 5	–1.20	0.90	0.81	0.70	0.79	0.45
Inferential Item 1	–0.63	0.85	0.81	0.61	0.79	0.48
Inferential Item 2	–1.35	0.76	0.61	0.72	0.78	0.55
Inferential Item 3	–1.05	0.87	0.80	0.67	0.79	0.47
Inferential Item 4	–0.86	0.86	0.75	0.64	0.79	0.47
Inferential Item 5	–1.30	0.90	0.87	0.71	0.79	0.45

Table 135. CS Nonfiction G3 Form D Item Analysis.

Items	IRT Difficulty	Infit	Outfit	p-Value	Alpha If Deleted	Adjusted Point-Biserial Correlation
Literal Item 1	–1.33	1.02	0.97	0.72	0.78	0.34
Literal Item 2	0.37	0.78	0.88	0.43	0.76	0.56
Literal Item 3	–1.14	0.86	0.81	0.69	0.77	0.45
Literal Item 4	0.11	0.82	0.68	0.47	0.76	0.56
Literal Item 5	–1.33	0.94	0.93	0.72	0.78	0.38
Inferential Item 1	–0.06	0.97	0.95	0.50	0.77	0.45
Inferential Item 2	–0.06	0.99	0.89	0.50	0.77	0.43
Inferential Item 3	–0.39	0.99	1.00	0.56	0.78	0.42
Inferential Item 4	0.41	0.82	0.74	0.42	0.76	0.55
Inferential Item 5	–0.18	0.96	0.86	0.53	0.77	0.45

For more information visit amplify.com

Corporate:
55 Washington Street
Suite 900
Brooklyn, NY 11201-1071
(212) 796-2200

Sales Inquiries:
(866) 212-8688 • amplify.com

Amplify.