

mCLASS®

DIBELS® 8th Edition

Administration and Scoring Guide



Amplify.

Amplify.

© 2019 Amplify Education, Inc. All trademarks and copyrights are the property of Amplify or its licensors.



© University of Oregon, 2018-2019. 8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®). Eugene, OR: University of Oregon. Available: dibels.uoregon.edu

ISBN: 978-1-64089-021-3

Table of Contents

Introduction to DIBELS® 8th Edition	5
Chapter 1: Administration Instructions and Scoring Procedures	32
Letter Naming Fluency (LNF)	38
Phonemic Segmentation Fluency (PSF)	42
Nonsense Word Fluency (NWF)	48
Word Reading Fluency (WRF)	54
Oral Reading Fluency (ORF)	57
Maze	62
Interpreting DIBELS 8th Edition Scores	70
DIBELS 8th Edition Normative Information	74
Reliability of DIBELS 8th Edition	80
Validity of DIBELS 8th Edition	122
References	179
Appendix A: ORF Benchmark Passage Statistics	187
Appendix B: Maze Benchmark Passage Statistics	191
Appendix C: DIBELS 8th Edition Pronunciation Guide.	195
Appendix D: Administration and Scoring Fidelity Checklists	199
Appendix E: Composite Score Calculation Guide.	207

Introduction to DIBELS® 8th Edition

This manual is a compendium of information regarding DIBELS 8th Edition. It details the nature and purpose of DIBELS 8, how DIBELS 8 differs from previous editions of DIBELS, how to administer and score DIBELS 8 subtests, how to use DIBELS 8 data to inform instructional decision-making, and the technical qualities of DIBELS 8. It also provides appendices for composite scores, risk classifications, and percentile ranks. This manual begins with an overview of the history of the DIBELS assessment system.

DIBELS History

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) consists of a set of measures for assessing reading skills. DIBELS began as a series of short tests that assessed early childhood literacy in kindergarten and first grade (Meyer, 2000). Over the years, DIBELS has gone through several editions, expanding the range of skills assessed and grades in which it can be used. DIBELS is now in its 8th Edition, which offers reading measures for Grades K-8.

DIBELS began as Dynamic Indicators of Basic Skills (DIBS; Shinn, 1989, 1998). Inspired by Deno's (1986) definition of curriculum-based measurement (CBM), DIBS, and DIBELS after it, was an attempt to ground classroom assessment practices and decision making in measurement science. With the support of a federal grant, the first DIBELS measures intended for use in the elementary grades (i.e., kindergarten and first grade) were developed as part of Dr. Ruth Kaminski's doctoral thesis in 1992 at the University of Oregon, where Dr. Roland Good served as her advisor. The measures were Letter Naming Fluency, Picture Naming Fluency, and Phonemic Segmentation Fluency. In the years since, the evolution of DIBELS measures and their interpretation has involved a number of University of Oregon faculty in addition to Dr. Good and Dr. Kaminski, including, but not limited to, Dr. Edward Kame'enui, Dr. Mark Shinn, and Dr. Deborah Simmons. In addition, numerous University of Oregon graduate students have contributed to the rich history of DIBELS research and development, including Dr. Sylvia Barnes Smith, Dr. Rebecca Briggs, Dr. Kelli Cummings, Dr. Deborah Laimon, and Dr. Kelly Powell-Smith, among others.

Updated editions of DIBELS have been released every several years beginning in 1996. Before DIBELS 8th Edition, the last update (DIBELS Next) was in 2010 and before that in 2002 (DIBELS 6th Edition). Over the years, subtests have come (e.g., Nonsense Word Reading Fluency, Oral Reading Fluency) and gone (e.g., Picture Naming Fluency, Initial Sound Fluency). DIBELS 8th Edition continues the legacy of development and research that has been ongoing

at the University of Oregon since the late 1980s. It introduces several changes, including new features such as measures spanning kindergarten through eighth grade, a new DIBELS measure (Word Reading Fluency), and modern measurement approaches to scoring, as well as the retirement of two existing measures (First Sound Fluency and Retell Fluency).

Dimensions of Reading Assessed by DIBELS 8

DIBELS 8th Edition offers six subtests designed to assess component skills involved in reading: Letter Naming Fluency (LNF), Phonemic Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Word Reading Fluency (WRF), Oral Reading Fluency (ORF), and Maze. These subtests are aligned to four of the five “Big Ideas” in reading identified by the National Reading Panel (National Institute of Child Health and Human Development, 2000), including phonological awareness, phonics (or the alphabetic principle), fluency, and comprehension (Riedel, 2007; see Table 1.1). In many ways the DIBELS subtests represent not only the constructs in the National Reading Panel Report (NICHD, 2000), but also a developmental continuum. As a result, the subtests included change across grades in a manner that parallels student development and instructional foci (Adams, 1990; Chall, 1996; Ehri, 2005; Paris & Hamilton, 2009).

Table 1.1 The Big Ideas in Reading and DIBELS 8 Subtests

Pattern	LNF	PSF	NWF	WRF	ORF	Maze
Phonemic awareness		X				
Alphabetic principle			X	X	X	
Accuracy and fluency with text				X	X	X
Vocabulary						
Comprehension					X	X

Some DIBELS 8 subtests are also aligned to subskills of reading that are associated with risk for dyslexia and other word reading disabilities. The International Dyslexia Association (IDA) recommends universal screening of students in kindergarten through second grade (IDA, 2019). Consistent with IDA recommendations, DIBELS 8 offers LNF, PSF, and NWF subtests as dyslexia screening measures of rapid naming (or processing speed), phonemic awareness, and letter-sound correspondence for use in kindergarten and first grade. Also consistent with IDA recommendations, DIBELS 8 offers real and nonsense word measures (NWF, WRF, and ORF) as dyslexia screening measures.

Description of DIBELS 8

DIBELS 8th Edition takes a curriculum-based measurement (CBM) approach to assessing reading. It is intended for assessing reading skills from the beginning of kindergarten through the end of eighth grade. DIBELS 8 subtests are designed as brief, easily administered measures of reading. Five of the subtests (LNF, PSF, NWF, WRF, and ORF) are 60-second measures designed to be administered individually in a quiet setting. The sixth subtest, Maze, is a 3-minute measure designed to be administered in group settings. Because DIBELS subtests are timed measures, efficiency in reading skills is considered as well as accuracy. The subtests offered in specific grades are aligned to curriculum and instruction typical for each grade, as well as to recommendations made by the IDA (see Figure 1.1).

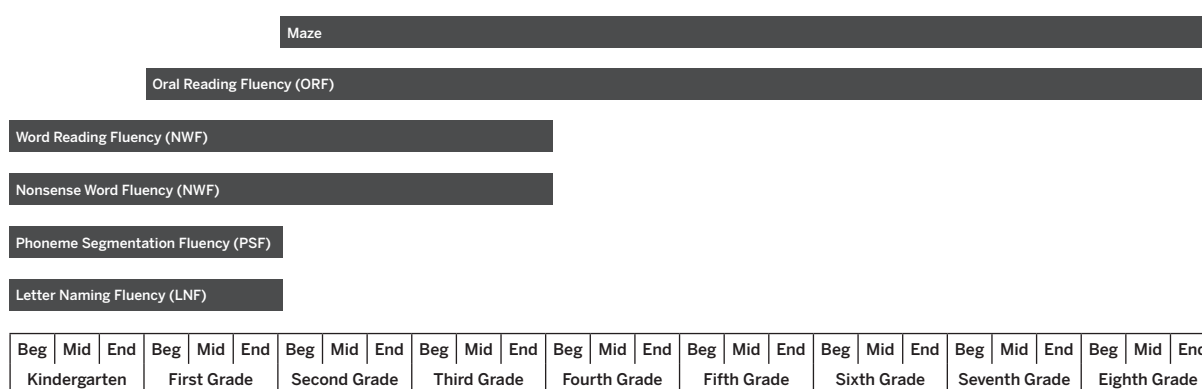


Figure 1.1 DIBELS 8th Edition Timeline of Subtest Availability by Grade

To maintain efficiency of benchmark assessment procedures, we have instituted new discontinuation rules to save time and avoid student frustration during benchmark assessment. As a result, total administration time varies by grade and by student skill (see Table 1.2).

Table 1.2 Administration Time in Minutes by Grade and Administration Type

Pattern	K	1	2-3	4-8
Individual	4-6	5-7	4	2
Group	NA	NA	5	5

Note. Ranges are provided in grades where rules exist for discontinuing a benchmark assessment. Only Maze is administered in a group setting. NA = not applicable.

Letter Naming Fluency (LNF). LNF is a standardized, individually-administered test that provides a measure of risk for reading achievement. LNF is based on research by Marston and Magnusson (1988) and is administered to students in the beginning of kindergarten through the end of first grade.

For LNF, students are presented with a page of 100 uppercase and lowercase letters arranged in a random order and are asked to name as many letters as they can. Students are given 1 minute to provide letter names. If a student does not know a letter name, the examiner provides the letter name and marks the letter name incorrect. The LNF measure has 3 benchmark forms for each grade in which it is available. As in previous editions, alternate progress-monitoring forms are not provided for LNF because it serves solely as a risk indicator.

Phonemic Segmentation Fluency (PSF). PSF is a standardized, individually-administered measure of phonological awareness. PSF is a good predictor of reading achievement and is administered to students in the beginning of kindergarten through the end of first grade.

PSF assesses students' ability to fluently segment two- to six-phoneme words into their individual phonemes. In PSF, the examiner orally presents a series of words and asks a student to verbally produce the individual phonemes for each word. For example, if the examiner said "sat," and the student said "/s/ /a/ /t/," the student would receive three points for the word. After each response, the examiner presents the next word. Students are given 1 minute to segment the words into phonemes. The PSF measure has 3 benchmark forms and 20 alternate progress-monitoring forms for each grade in which it is available.

Nonsense Word Fluency (NWF). NWF is a standardized, individually-administered measure of the alphabetic principle. NWF is seen as a "pure" measure of the alphabetic principle, because vocabulary and sight word knowledge cannot play a role in recognizing nonsense words. NWF is administered to students in the beginning of kindergarten through the end of third grade.

NWF assesses students' ability to decode words based on the alphabetic principle. For NWF, students are presented with an 8.5-inch x 11-inch sheet of paper with nonsense words (e.g., sig, ral) and asked to verbally produce (a) the whole nonsense word or (b) individual letter sounds. For example, if the stimulus word is "hap", a student could say the nonsense word as a whole or "/h/ /a/ /p/" to receive three letter sounds correct. On DIBELS 6th Edition, if the nonsense word was read as a whole (either initially or after sounding out), the student received credit for one whole word read correctly. On DIBELS Next, the student only received credit for reading the nonsense word correctly if it was read as a whole in the initial attempt. DIBELS 8th Edition reverts to the DIBELS 6th Edition practice because it more

accurately captures students' knowledge of sound-spelling patterns and the ability to blend sounds into words, which is the primary intent of NWF. Students are given 1 minute to read or sound out as many nonsense words as they can. The NWF measure has 3 benchmark forms and 20 alternate progress-monitoring forms for each grade in which it is available.

Word Reading Fluency (WRF). The new WRF subtest involves reading real words out of context. Inspired by other CBMs that incorporate WRF, most especially easyCBM (Alonzo & Tindal, 2007), it is a standardized, individually-administered measure of accuracy and fluency in reading “sight” words. Sight words include words with irregular pronunciations (non-decodable words like “the” and “was” and “of”) as well as common words with regular pronunciations (decodable words like “in” and “we” and “no”). WRF is administered to students from the beginning of kindergarten through the end of third grade.

In WRF, students are presented with an 8.5-inch x 11-inch sheet of paper with real words and asked to verbally produce the whole word. Students must blend words to receive credit. In contrast to NWF, no credit is given for individual letter sounds. Students are given 1 minute to read as many words as they can, and the final score is the number of words read correctly within 1 minute. The WRF measure has 3 benchmark forms and 20 alternate progress-monitoring forms for each grade in which it is available.

Oral Reading Fluency (ORF). ORF is a standardized, individually-administered measure of accuracy and fluency with connected text. ORF is administered to students in the beginning of first grade through the end of eighth grade.

ORF assesses a student's ability to read words in connected text. In ORF, the examiner presents the student with a passage and asks the student to read the passage aloud for 1 minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The final score is the number of words read correctly (and self-corrected) within 1 minute. The ORF measure has 3 benchmark forms and 20 alternate progress-monitoring forms for each grade in which it is available.

Maze. Maze is a standardized, group-administered measure of reading comprehension. Maze is administered to students in the beginning of second grade through the end of eighth grade.

In Maze, the examiner presents students with a passage that has every seventh word removed and replaced with three options. In third through eighth grade, the first and last sentence are left intact, and in second grade, the first two sentences and last sentence are left intact. The final score is one-half the number of overt errors subtracted from the number of maze words selected correctly within 3 minutes. Skipped items are treated as errors, but

items not reached are not counted as errors. The Maze measure has 3 benchmark forms for each grade in which it is available. We will release a limited number of progress monitoring forms by fall 2020; the number will be limited because reading comprehension does not improve as rapidly as the other components of reading assessed by DIBELS and therefore monitoring is conducted less often.

Appropriate Uses of DIBELS 8

DIBELS 8 measures are designed to be used from the beginning of kindergarten through the end of eighth grade. Although DIBELS 8 can be used for off-grade assessment, it has not been validated for this use.

DIBELS 8 subtests were developed and researched as indicators of risk and progress in overall reading, as well as risk for dyslexia and other reading difficulties. DIBELS 8 has three principal uses: to identify students who may be at risk of reading difficulties by screening up to three times per year, to document students' progress of reading skills as a consequence of special intervention programs through progress monitoring, and to provide minimum levels of performance for all students to reach to be considered on track for becoming a reader through benchmark goals and timelines. DIBELS 8 benchmark forms were validated as screening measures administered at the beginning, middle, and end of a school year. Additional forms have been validated for use in progress monitoring and are provided for most measures.

DIBELS was also designed as a means to evaluate the effectiveness of intervention for those students receiving support, in order to inform changes in intervention strategy as necessary to improve student learning and growth. Similarly, DIBELS was designed for use in research on reading development, especially the development of readers at risk.

DIBELS 8 can be used to make judgments about the instructional needs and responsiveness of individual students and regarding the efficacy of curriculum and instructional practices. It has not been designed to render judgments regarding teacher effectiveness or school progress.

Rationale for and Innovations in DIBELS 8

DIBELS 8 was developed consistent with best practices in educational measurement (AERA, APA, & NCME, 2014). The most recent standards for educational testing suggest that test creators must renorm tests “with sufficient frequency to permit continued accurate and appropriate score interpretations” (AERA et al., 2014, p. 104). This standard is interpreted as meaning that not only should norms be updated regularly (as DIBELS Next norms have

been), but related validity evidence must also be updated, especially when used for critical instructional decision-making. Given that validity data on the last edition of DIBELS was almost a decade old and educational practices have shifted during that decade (for example, the introduction of Common Core Standards and a new generation of state tests), a new edition of DIBELS was deemed necessary. In addition, a new edition provided an opportunity to improve DIBELS in several ways. These innovations are summarized in this section.

Expanded grade levels. DIBELS 8 can be used in kindergarten through eighth grade. The expansion through eighth grade means that DIBELS can now be used in schools with a wider range of grade configurations: K-3, K-5, K-8, 5-8, 6-8, etc.

Consistent subtests within grade. DIBELS 8 subtests used at any point during a given grade are available for all benchmark periods in that grade (see Figure 1.1). This availability supports users who may want or be required to have consistent data across all three benchmark assessment periods in the school year.

Discontinue benchmark rules. To maintain efficiency of benchmark assessment procedures, we have instituted new discontinuation rules in kindergarten and first grade (see Table 1.3). These rules are intended to save time and avoid student frustration during benchmark assessment. For example, in the beginning of the school year, if a kindergarten student cannot segment any phonemes on PSF, or if a first grade student cannot read any words on WRF, the administrator does not need to administer the remaining subtests (NWF and WRF in kindergarten and ORF in first grade). In this way, administration remains efficient, while still yielding information on more able readers.

Table 1.3 Discontinue Benchmark Rules

Grade and Season	Benchmark Discontinue Rule	Scoring
Kindergarten, Fall	If PSF is discontinued, do not administer NWF and WRF.	Enter 0 for PSF. Do not enter scores for the remaining subtests: NWF and WRF
Kindergarten, Winter	If NWF is discontinued, do not administer WRF.	Enter 0 for NWF. Do not enter scores for the remaining subtest: WRF
First grade, Fall	If WRF is discontinued, do not administer ORF.	Enter 0 for WRF. Do not enter scores for the remaining subtest: ORF

These rules are intended to save time and to spare the student unnecessary frustration. The rules were derived from a national field trial that indicated students who scored 0 for the indicated assessments in the periods specified above were extremely unlikely to get any items correct on the remaining subtests. Nonetheless, examiners have the option of administering the remaining subtests based on professional judgment.

Font type and size. The font chosen for DIBELS 8th Edition was informed by research on the effect of fonts for children with and without word reading disabilities like dyslexia. Although a great deal of research has explored the effects of different fonts, including “dyslexia friendly” fonts like Dyslexie and Open Dyslexic, very few of these studies used rigorous scientific methods. The few peer-reviewed studies that have employed randomized trials have yielded equivocal results. Dyslexia-friendly fonts have no discernible effects on readers with and without dyslexia and other word reading disabilities. For example, children with and without dyslexia showed no significant differences in reading speed or reading accuracy when the Dyslexie, Times New Roman, and Arial fonts were compared, especially when spacing of letters was controlled across fonts (Duranovic, Senka, & Babic-Gavric, 2018; Marinus et al., 2016; Wery & Diliberto, 2017). More promising is research that shows that the spacing of letters, which co-varies with font-size, does affect reading speed and comprehension for all readers.

As a result, we paid a great deal of attention to font sizes in the development of DIBELS 8th Edition (see next section), but the ultimate choice of font was guided by the distinguishability of letters. Of paramount concern was that the capital i (I) be easily distinguishable from a lowercase L (l). To accomplish this aim, it was necessary to use a font with serifs, which are the slight projections on letters in some fonts. For example, in Arial font, which does not have serifs, the uppercase i and lowercase L are nearly indistinguishable: I, l.

However, no serif font represents the letters a, g, j, and q in the forms they are more commonly taught in the primary grades: a, a, g, j, and q. In fact, any font that represents one of these letters as they are typically taught represents other letters in a less typical form. Given that no font fulfilled all of these practical considerations, we opted for the Rockwell font, which is similar to the more familiar Times New Roman but has slightly thicker serifs and a more typical form of lower-case G (see Table 1.4).

Table 1.4 DIBELS 8th Edition Fonts

Times New Roman	Rockwell
i, I, l, L	i, I, l, L
a, g, j, q	a, g, j, q

Rockwell is used for all subtests except for Oral Reading Fluency (ORF) and Maze, where Times New Roman is used instead. Times New Roman is used when subtest probes involve reading in context because research has shown that young readers, including those with dyslexia and other word reading difficulties, prefer familiar fonts (Kuster, van Weerdenburg, Gompel, & Bosman, 2018; Wery & Diliberto, 2017). This same research indicated that font and preference did not affect overall performance for any group of readers. As a result, we used the more familiar Times New Roman to minimize any potential interference in meaning-making that a less familiar font might cause.

Font sizes for DIBELS 8th Edition were informed by research on the effects of font sizes for children with and without word reading disabilities like dyslexia. For all readers, larger font sizes promote faster reading speeds up to a “critical” font size when increases in font size no longer result in faster reading (O’Brien, Mansfield, & Legge, 2005). This critical font decreases in size with grade level for all readers, suggesting that font sizes can be safely decreased each year for all readers. Although readers with dyslexia benefit from larger font sizes than readers without dyslexia, their ability to read smaller font each year efficiently progresses in a similar fashion. Additional research demonstrates that smaller font sizes and longer line lengths can also interfere with primary grade readers’ ability to comprehend text (Katzir, Hershko, & Halamish, 2013). However, for intermediate grade readers larger font sizes interfered with comprehension, while line lengths had no effect. Spacing between lines had no effect for either group of readers. In general, research indicates that larger print results in younger readers reading faster and comprehending better, but that there are diminishing and even no benefits the older a reader is (Hughes & Wilkins, 2000; Katzir et al., 2013; O’Brien et al., 2005; Wilkins, Cleave, Grayson, & Wilson, 2009). Thus, DIBELS 8th Edition font sizes start at 24pt in Kindergarten and get slowly, but progressively smaller until fifth grade (see Table 1.5). ORF font sizes are slightly smaller than font sizes for the other DIBELS 8th Edition subtests to keep passages from taking up more than the front and back of a single page without resorting to overly narrow margins.

Table 1.1 The Big Ideas in Reading and DIBELS 8 Subtests

Subtest	K	1	2	3	4	5+
LNF	24pt	22pt	20pt	18pt	NA	NA
NWF	24pt	22pt	20pt	18pt	NA	NA
WRF	24pt	22pt	20pt	18pt	NA	NA

Table 1.1 The Big Ideas in Reading and DIBELS 8 Subtests

Subtest	K	1	2	3	4	5+
ORF	24pt	22pt	20pt	18pt	NA	NA
Maze	NA	20pt	18pt	16pt	14pt	13pt

Note. NA = not applicable.

Letter Naming Fluency improvements. For DIBELS 8th Edition, LNF now accounts for how frequently letters appear in both uppercase and lowercase forms. To better control differences in difficulty between forms, consistent rules are used in both kindergarten and first grade regarding when less frequent letters can appear on the forms. Each form in both grades begins with a sampling of the 20 most frequently seen letters (Jones & Mewhort, 2004), thereby preventing students from getting frustrated by forms that begin with rarer letters, such as X or q. The kindergarten version of LNF also only assesses the 40 most commonly seen uppercase and lowercase letters, while the first grade version assesses 49 uppercase and lowercase letters.

LNF *excludes* three letters on all forms: uppercase and lowercase W and lowercase L. Although these are obviously important letters for students to know, they introduce real problems in a fluency assessment. W is the only letter with a multi-syllabic name: three syllables to be exact. As a result, any time W appears, it takes three times as long to name as other letters, which negatively affects a student's LNF score. The lowercase L (l) was eliminated because it is easily confused with both the uppercase I and the number 1. Not only does this visual similarity pose problems for students, but it has also historically created scoring problems for the adult administering the assessment. By avoiding these letters, each included item (or letter) is equally challenging, other than in terms of its frequency in printed language.

Phonemic Segmentation Fluency improvements. In DIBELS 8th Edition, PSF accounts for both word frequency and the number of phonemes in a word. All forms draw only from the 2,500 most frequent words in English (Balota et al., 2007) to minimize vocabulary familiarity from interfering with student performance. In addition, to better control differences in difficulty between forms, consistent rules are used in both grades regarding where less frequent words can appear on the forms. Moreover, spelling patterns are ordered in terms of the number of phonemes, proceeding from two phoneme words to words with progressively more phonemes.

In kindergarten, the first 20% of items have two phonemes, while the remaining 80% have

three phonemes. In this way, PSF now avoids the distinct floor effects (i.e., many students scoring zero) in kindergarten that have plagued previous versions and, thus, eliminates the need for a separate measure of initial sound fluency. In first grade, the progression in difficulty is a bit more rapid, with the first 13% of items having two phonemes and then increasing in phonemes with additional increases after every eight items.

Nonsense Word Fluency improvements. In DIBELS 8th Edition, NWF now accounts for the frequency of spelling patterns (Jones & Mewhort, 2004; Norvig, 2012). As a result, all forms utilize only phonetically regular letter combinations that actually appear in English. Thus, students will no longer be asked to decode nonsense words like “fev” or “kaj”, and nonsense words like “kex” will appear less often than ones like “lat”.

DIBELS 8th Edition also expands the spelling patterns assessed beyond simply consonant-vowel-consonant (CVC) after kindergarten. While kindergarten forms are limited to CVC patterns, the first grade forms also include vowel-consonant (VC) spelling patterns. In addition, the latter half of first grade forms include additional spelling patterns typically taught in first grade, thus increasing the instructional relevance of this DIBELS subtest. DIBELS 8th Edition also now offers NWF in second and third grade by including more complex phonics patterns in these grades. As a result, DIBELS NWF forms provide instructionally relevant information even for students who are at minimal risk in kindergarten through third grade. New spelling patterns included in first through third grade appear in Table 1.6 below.

Table 1. Examples of First through Third Grade NWF Spelling Patterns

Pattern	Grade introduced	Example non-word
CVCe	1	bace
CVr(C)	1	zart
CVCC	1	melb
CCVC	1	scap
CCVCC	1	brold
(C)CVVC(C)	2	geap
CVCCy	2	foddy
(C)V CVC(C)	3	cotalm
(C)VC CVC(C)	3	fudlerk

An additional improvement to NWF is that we have reverted to scoring words recoded correctly (WRC; DIBELS 6th Edition practice) rather than whole words read (WWR; DIBELS Next practice). Whereas with WWR students only received credit if they correctly read a nonsense word at first sight (i.e., without sounding out), with WRC they also receive credit if they blend a nonsense word after sounding out the component sounds. Because both methods of scoring predict student risk, in DIBELS 8th Edition, students receive credit for blending nonsense words whether they sound them out first or not. In addition, with WWR the information about students' ability to blend words was lost if students first verbalized the sounds prior to blending them into words. In this case students would receive no credit for whole words read even though they blended sounds into words. Since the main purpose of NWF is to assess readers' understanding of the alphabetic principle and sound-symbol correspondence, WRC was deemed the more appropriate scoring method.

Word Reading Fluency innovative features. WRF targets real words based on age of acquisition in students' vocabulary (Brysbaert & Biemiller, 2017) and their frequency in written text (Balota et al., 2007). WRF assesses only words that are typically acquired orally in or before a given grade. This reduces the likelihood that students will encounter words on the assessment that they have never heard before and are not yet expected to know.

In addition, each form starts with a sample of the most frequent words seen in text and then moves on to less frequent words in the latter half of the form. In this way, WRF yields instructionally relevant information both for students at risk and students at minimal risk.

Finally, DIBELS WRF accounts for word complexity, as measured by the number of syllables in a word. All forms include one-syllable words. Grades 1-3 include two-syllable words, and Grades 2-3 include two-syllable and three-syllable words. In Grade 3, we also included words with more than three syllables, but again only those that are typically acquired by Grade 3 and are frequently seen in print.

These features ensure the instructional relevance of DIBELS WRF results for all students. Importantly, our research, as well as that of others (Clemens, Shapiro, & Thoemmes, 2011; Fuchs, Fuchs, & Compton, 2004; Smith, Cummings, Nese, Alonzo, Fien, & Baker, 2014), has shown that the inclusion of WRF helps to identify students at risk who might otherwise be missed by other DIBELS subtests.

Oral Reading Fluency improvements. DIBELS 8th Edition marks the first time that DIBELS ORF requires the administration of only one passage per benchmark period. Research has shown that administering more than one passage does little to improve the reliability and validity of ORF, meaning that the minimal benefits of administering three passages just does not warrant the additional administration time (Baker et al., 2015; Petscher & Kim, 2011).

Rather, a single passage works just as well, and reduces the testing burden for both students and assessors.

An additional unique and exciting feature of DIBELS 8th Edition ORF passages is that they were written by experienced and aspiring children's authors, most of whom have previous experience writing for students and have previously published short stories. The authors have diverse backgrounds, come from across the US, and have experience writing in a range of genres. As a result, ORF passages are not only more engaging for both students and assessors, but also read as more authentic and appropriate for the grades in which they appear.

Maze improvements. Maze has now been informed by research that shows consistently that maze measures tend to assess low-level comprehension (e.g., January & Ardoin, 2012; Shanahan, Kamil, & Tobin, 1982). To make DIBELS maze measures more informative, we undertook several innovations. First, as with ORF, maze passages are written by experienced and aspiring authors. Second, more work has gone into the selection of distractors, and this work is described in the section on development. Third, the formatting of Maze was revised to make reading the passages easier on the eye, reflecting research that suggests that overly long lines can cause disfluency and interfere with reading comprehension for young readers (e.g., Dyson & Haselgrove, 2001; Katzir et al., 2013). Finally, maze measures will be available in second through eighth grade instead of only third through sixth.

Retirement of subtests. Both First Sound Fluency (FSF) and Retell Fluency (RTF) have been removed from DIBELS 8 as subtests for several reasons. First and most critically, both subtests add time to the administration of DIBELS without adding much useful information for screening or instructional planning. Thus, the information yielded through these measures relative to the time spent administering them was not deemed as worthwhile as it was for the other DIBELS subtests.

Additional factors that played into the decision to drop FSF were its redundancy and constrained nature. Given our modifications to PSF, FSF was deemed more redundant with PSF than it had been in the past. In addition, First Sound Fluency taps a very constrained aspect of phonemic awareness: the detection of initial phonemes. Given that this phonemic awareness skill is mastered quite quickly, especially in the presence of instructional intervention, FSF was not a good candidate for administering in all three benchmark periods in kindergarten. In other words, its best use was incompatible with the new design specifications of DIBELS 8, which requires the same subtests to be available throughout a grade.

Additional factors that played into the decision to drop RTF were questions regarding its

validity and the new availability of Maze in lower grades. DIBELS users have often questioned whether a words-per-minute rate for retelling captures comprehension adequately. When measured as a rate, factors unrelated to comprehension can radically affect scores. For example, students with speech impediments like stutters will produce fewer words in their retell regardless of their level of comprehension. Similarly, students learning English who have more limited expressive vocabulary than receptive vocabulary also have a tendency to score lower than their English-only speaking counterparts regardless of their level of comprehension. In other words, RTF was as much a measure of expressive language fluency as it was of reading comprehension. The influence of expressive language fluency is undesirable in a measure of reading comprehension.

Furthermore, recent research has demonstrated that retell in the absence of word reading fluency is almost useless. Word reading accounts for almost all the variability in first grade reading comprehension measures (e.g., Lonigan & Burgess, 2017; Lonigan, Burgess, & Schatschneider, 2018). In fact, word reading sets a hard limit on whether a student can read enough text to build a mental representation of what is read. For students who score below the risk cut-score in first grade, which is 7 or fewer words at the beginning of the year and 32 or fewer words at the end of the year, very little of substance has been read. As a result, only the students with the most advanced reading skills will be able to give a retelling that yields any useful information. However, this picture begins to change quickly past first grade. Consequently, Maze, which has superior predictive powers to retell fluency, is now available from Grade 2 onward.

Development of DIBELS 8

DIBELS 8 was developed consistent with the most recent standards in educational measurement (AERA et al., 2014). These standards provide criteria for test development that promote the validity of interpretations of test scores. The development process outlined by the standards includes detailing the intended uses of a test, specifying content and format requirements, and using specifications to create item pools, inform item selection, and guide assignment of items to forms. These standards also recommend an iterative approach to development decisions and evidence gathering.

Consistent with these standards, the intended uses of DIBELS 8 were defined (as noted in the section of this manual on Appropriate Uses). Prior to determining specifications, DIBELS researchers performed a comprehensive literature review of critiques and limitations of DIBELS and other reading CBMs. Researchers also consulted with DIBELS Data System (DDS) customer service at the University of Oregon to gain an understanding of which aspects of DIBELS first-hand users reported as the most valuable and the least valuable, as

well as what they frequently ask for that DIBELS did not yet offer. As a result, several new goals were identified for DIBELS 8.

Increase the utility of NWF by expanding the spelling patterns assessed and grades in which it is available. Research (e.g., January, Ardoin, Christ, Eckert, & White, 2016) has demonstrated that NWF can be a more useful tool for screening and monitoring progress when patterns assessed move beyond CVC words. Research has also shown that this utility extends beyond kindergarten and first grade. As a result, expanding both the spelling patterns assessed by NWF and grades in which NWF was assessed became a goal.

Pay attention to order effects. Research (e.g., Burns et al., 2009) has shown the order in which items appear on fluency measures affects reading rate. Specifically, a form that begins with easier items and in which items become progressively more difficult allows for maximal performance by students, essentially by allowing them to gain momentum. In contrast, when item difficulty is more randomly distributed, the rate is adversely affected. As a result, using progressive difficulty as a principle in item assignments to forms became a goal for LNF, PSF, NWF, and WRF.

Compensate for form effects on oral reading fluency and maze through equating. A wealth of research (e.g., Baker et al., 2015; Cummings, Park, & Bauer Schaper, 2013; Santi, Barr, Khalaf, & Francis, 2016) has demonstrated that ORF is subject to form effects that can obscure the actual progress of readers. Form effects are average difficulty effects of reading passages that persist despite the efforts with DIBELS and other reading CBMs to tightly control passage equivalence through readability formulas and passage piloting. These effects have been well studied for ORF and were presumed to affect Maze equally, as well as the other DIBELS subtests to a lesser extent. As a result, equating alternate forms for DIBELS subtests became a goal, with the equating of ORF and Maze taking first priority. Given that all students are assessed with benchmark forms, the equating of benchmark forms also took precedence over the equating of all alternate forms.

Add a word reading fluency measure. Research (e.g., Fuchs et al., 2004; January et al., 2016; Smith et al., 2014; Wise et al., 2010) has also shown that NWF and ORF do not capture all struggling readers. Particularly in the early elementary grades, WRF improves prediction of students at risk, as well as monitoring of progress. As a result, the incorporation of WRF became a priority.

Provide consistent subtests within a grade. DDS customers frequently expressed a desire for consistency in subtests available within a grade. The lack of consistency, especially in kindergarten through second grade, made tracking the progress of all students during an academic year more challenging.

Validate specifically for use as a dyslexia screening tool. Perhaps the number one question of DDS customers in the last few years has been whether DIBELS is a valid screening assessment for dyslexia. While DIBELS has always been validated as a screener of risk for reading difficulties, it had never been specifically validated as a screening measure for word reading disabilities, including dyslexia. Of particular concern was the use of LNF as a measure of processing speed via rapid automatized naming (RAN). As a result, this new use of LNF, and DIBELS in general, informed development decisions as never before.

Letter Naming Fluency development process. The item pool for LNF consists of the uppercase and lowercase versions of all English letters, with the exception of the lowercase L (l) and both uppercase and lowercase W. Although these are obviously important letters for students to know, they were excluded to better align LNF to its increasing use as a RAN measure of processing speed. For the same reason, the kindergarten item pool was further limited to the 40 most frequently seen uppercase and lowercase letters. The kindergarten pool therefore includes all lowercase letters except z and q, while the first grade pool includes the remaining nine letters.

To better control for differences in difficulty between forms, consistent rules are used in both kindergarten and first grade regarding when less frequent letters can appear on the forms. Each form in both grades begins with a sampling of the 20 most frequently seen letters, thereby preventing students from getting frustrated by forms that begin with rarer letters. Uppercase and lowercase letter frequency was determined based on the average frequency from five large corpora, as reported in Jones & Mewhort (2004). The 49 letters in the item pool were then grouped by average frequency into ten bins of five items each (with the exception of the final group, which includes only four items). In kindergarten, three sets of the top 20 items and two sets of items 21-40 were combined to create a total item pool of 100 letters per form. In Grade 1, two sets of the 49-item pool, plus two additional, randomly selected letters were combined to create a total item pool of 100 letters per form.

Each item was then assigned a random number using the default random number generator available in the statistical programming language R (R Core Team, 2018). Next, items were sorted so that each row included one letter from each bin, with the relative position of the individual letters in each bin determined by the random numbers assigned to each letter. Letters with a lower random number appeared before letters with a higher random number. Within each row, letters were strategically positioned so that the first row presented bins in decreasing order of frequency, and subsequent rows ordered the bins in varying combinations of difficulty.

This process, including the generation of a new set of random numbers, was repeated 50 times per grade to generate a pool of 50 potential forms. Multiple research staff

reviewed each form, from which three were selected as the benchmark forms, and an additional 20 were selected as the progress monitoring forms by eliminating forms in which the same letter occurred more than once in succession or in which sequences of letters spelled English words.

Phonemic Segmentation Fluency development process. To minimize the effect of vocabulary familiarity, all forms draw from the 2,500 most frequent two- to six-phoneme words in English, based on data from the English Lexicon Project (Balota et al., 2007). The initial item pool included all words from the English Lexicon Project that a) were identified as one of 2,500 most frequent words in both the Kučera & Francis (1967) word frequency list and the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996); b) had at least one meaning known by at least 50% of second grade students (Dale & O'Rourke, 1981); and c) had an adult-rated age of acquisition less than or equal to 7 (Brysbaert & Biemiller, 2017). This resulted in an initial pool of 662 words. We then removed homonyms (e.g., two, hear), potentially sensitive words (e.g., fight, hit), and rare two-phoneme words (i.e., those with an average word frequency > 200). This resulted in a first grade item pool of 594 words. For the kindergarten item pool, we further removed four- through six-phoneme words, resulting in a kindergarten item pool of 295 words.

To better control differences in difficulty between forms, consistent rules were used in both grades regarding where less frequent words can appear on the forms. Moreover, spelling patterns were ordered in terms of the number of phonemes, proceeding from two-phoneme words to words with progressively more phonemes. In kindergarten, the first six items have two phonemes, while the remaining 24 have three phonemes. In this way, PSF now avoids the distinct floor effects (i.e., many students scoring zero) in kindergarten that have plagued previous versions and, thus, eliminates the need for a separate measure of initial sound fluency. In first grade, the progression in difficulty is more rapid, with the first four items having two phonemes, the second six having three phonemes, and each subsequent group of six words having one more phoneme than the previous group.

All words in the final PSF item pool were assigned a random number using the default random number generator available in the statistical programming language R (R Core Team, 2018). This number was used to select words for inclusion in each section of the form (e.g., in kindergarten, the six two-phoneme words with the lowest random numbers were selected first, followed by the 24 three-phoneme words with the lowest random numbers). This process, including the generation of a new set of random numbers, was repeated 25 times per grade to create a pool of 25 potential forms. Multiple research staff reviewed each form, from which three were selected as the benchmark forms, and an additional 20 were selected as the progress monitoring forms.

Nonsense Word Fluency development process. The NWF item pool for DIBELS 8th Edition differs from previous versions of NWF in two important respects. First, all items now respect the English order and word position rules of individual letter combinations (Jones & Mewhort, 2004; Norvig, 2012), meaning that only phonetically regular letter combinations that actually appear in English are used: students are no longer asked to decode nonsense words such as *fev* or *kaj*. Second, the spelling patterns assessed have expanded beyond just the consonant-vowel-consonant (CVC) and vowel-consonant (VC) patterns used in previous versions. Although kindergarten forms are limited to CVC patterns, first grade forms include additional spelling patterns (described below) that are typically taught in first grade. DIBELS 8th Edition also now offers NWF in second and third grades, and includes additional, more complex phonics patterns in these grades.

The NWF item pool was created by first compiling lists of legal word parts in English, including various onsets and rimes. A total of 78 onset patterns were identified, including single letter onsets (e.g., *b*, *s*), blends (e.g., *bl*, *tr*), digraphs (e.g., *ch*, *kn*), trigraphs (e.g., *str*, *thr*), and VC onsets (e.g., *am*, *ev*). An additional 219 rimes were identified, including VC rimes (e.g., *ab*, *in*), vowel-consonant-e (VCe) rimes (e.g., *abe*, *ide*), vowel-r (Vr) rimes (e.g., *ar*, *ir*), vowel-r-consonant (VrC) rimes (e.g., *arm*, *ort*), vowel-consonant-consonant (VCC) rimes (e.g., *est*, *olk*), and vowel-vowel-consonant (VVC) rimes (e.g., *aid*, *eed*). These lists were then cross-combined in all possible legal English combinations and matched to a list of 31,845 real words and a separate list of 704 nonsense words that are pronounced like either a real word or a proper name or were deemed inappropriate or difficult to pronounce. Items on either list were removed from the item pool.

Finally, we used frequency counts of English letter n-grams (Norvig, 2012) to weight the relative frequency with which various word parts should appear on each form, computed a total frequency estimate for each nonsense word, and divided the total pool of nonsense words into quartiles based on that frequency estimate. Words with the least frequently appearing combinations of letters (i.e., those in the bottom quartile) were dropped from the pool, resulting in a total NWF item pool of 79,314 nonsense words.

Items are arranged in five columns, and forms include between 75 (in kindergarten) and 100 (Grade 3) nonsense words. Form templates were created for each grade, which specified a particular sequence and relative frequency of each spelling pattern. In kindergarten, all 75 items are CVC words. In the first grade template, the first five rows (a total of 25 items) consist entirely of CVC and VC nonsense words. In the next four rows, half of the items are again CVC or VC nonsense words, and the other half are silent-e (VCe) and r-controlled (CVrC) nonsense words. In the next three rows, consonant blends (CVCC) and digraphs (CCVC) are introduced: 1/3 of items (i.e., 5 words) are VC and CVC, 1/3 are VCe and CVrC, and 1/3 are CVCC and CCVC. In the final three lines, more complex patterns (i.e., CCVCC and CCCVC)

are introduced: 1/3 of items 1/3 are CVCe and CVrC, 1/3 are CVCC and CCVC, and 1/3 are CCVCC and CCCVC. The templates for second and third grades followed similar patterns but introduced additional spelling patterns: vowel digraphs (Grade 2), short vowel words ending in Y (Grade 2), and two-syllable words (Grade 3).

As with the other subtests, all nonsense words in the final NWF item pool were assigned a random number using the default random number generator available in the statistical programming language R (R Core Team, 2018). This number was used to select words for inclusion in each section of the form (e.g., in Grade 1, the 15 CVC and VC nonsense words with the lowest random numbers were selected first, and then the next 10 CVC and VC nonsense words were combined with the 10 CVCe and CVrC words with the lowest random number to complete the second section). Within each section, items were further randomized to ensure each word type appeared in a variety of positions. This process was repeated for each section, and then 36 times per grade (including the generation of a new set of random numbers) to create a pool of 36 potential forms. Multiple research staff reviewed each form, from which three were selected as the benchmark forms, and an additional 20 were selected as the progress monitoring forms.

Word Reading Fluency development process. Four different word lists contributed to the definition of the item pool for WRF. The Dale and O'Rourke (1981) word list is the only known list of words with age of acquisition determined by actual assessment with children. Because they assessed only fourth grade students and above, we supplemented their list with the far more recent work by Brysbaert and Biemiller (2017).

Importantly, Biemiller estimated in earlier work (2010), that words known by 80% or more of fourth graders were likely to be known by 50% or more of second graders, thereby allowing for extension of the Dale and O'Rourke grade of typical acquisition down to second grade. These researchers asked adults to retrospectively estimate the age at which they knew words on the Dale and O'Rourke list. Results showed remarkable agreement between the children (test-based) and adults (retrospective) in terms of age of acquisition. As a result, we used the more fine-grained information from the newer list to further winnow down the list used for each grade.

Two word frequency lists were also used in creating the WRF pool. The Kučera and Francis (1967) word list is widely used for its comprehensiveness and availability. Nonetheless, it is an older list and the English language changes constantly. Thus, we also used the newer Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996), which includes internet-based texts and is used very commonly as well. We consulted these word frequency lists together with age of vocabulary acquisition lists because most word frequency corpora, including the two we used, do not restrict themselves to children's texts. The

combination of all four lists ensures that the words chosen have been frequent over several decades and are age- and grade-appropriate.

To create the WRF item pool, we began with a list of about 40,000 English words with a test-based age of acquisition rating (Dale & O'Rourke, 1981). We then excluded items that met any of four criteria words whose meaning was known by less than 50% of fourth grade students (Dale & O'Rourke, 1981). Next, we eliminated words with an adult-rated age of acquisition greater than 9 (Brysbaert & Biemiller, 2017). From this more limited pool, we further narrowed down the list to words with a frequency rating in both the Kučera & Francis (1967) word frequency list and the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996) that was greater than 7,500 per million words. Finally, words with at least one meaning identified as potentially inappropriate or distracting for students were omitted from the list. This resulted in a total item pool of 2,065 words for Grade 3.

Additional restrictions were imposed on the item pools for kindergarten through second grade. For all three grades, words had to be known by at least 50% of second grade students (Brysbaert & Biemiller, 2017). In Grade 2, words also had to have a frequency rating in the top 5,000 in both the Kučera & Francis (1967) and Lund & Burgess (1996) frequency norms, have an adult-rated age of acquisition less than or equal to 8, and could only be up to three syllables in length. The total item pool for Grade 2 was 1,111 words. In Grade 1, these criteria were further constrained. Namely, words had to have a frequency rating in the top 2,500 in both the Kučera & Francis (1967) and Lund & Burgess (1996) frequency norms, have an adult-rated age of acquisition less than or equal to 7, and be one or two syllables in length. The total item pool for Grade 1 was 652 words.

Finally, in kindergarten, the words had to have a frequency rating in the top 1,000 in both the Kučera & Francis (1967) and Lund & Burgess (1996) frequency norms, have an adult-rated age of acquisition less than or equal to 6 (Brysbaert & Biemiller, 2017), and could only be one syllable in length. The total item pool for kindergarten was 242 words.

In each grade, the item pool was grouped into three bins based on relative frequency. In kindergarten, words rated as one of the 50 most frequent words by both Kučera & Francis (1967) and Lund & Burgess (1996) were placed in the first bin, words with a frequency rating between 51 and 300 were placed in the second bin, and words with a frequency rating between 301 and 1,000 were placed in the third bin. In Grade 1, words rated as one of the 50 most frequent words by both Kučera & Francis (1967) and Lund & Burgess (1996) were placed in the first bin, words with a frequency rating between 51 and 1,000 were placed in the second bin, and words with a frequency rating between 1,001 and 2,500 were placed in the third bin.

In Grade 2, words rated as one of the 300 most frequent words by both Kučera & Francis (1967) and Lund & Burgess (1996) were placed in the first bin, words with a frequency rating between 301 and 2,500 were placed in the second bin, and words with a frequency rating between 2,501 and 5,000 were placed in the third bin. In Grade 3, words rated as one of the 1,000 most frequent words by both Kučera & Francis (1967) and Lund & Burgess (1996) were placed in the first bin, words with a frequency rating between 1,001 and 5,000 were placed in the second bin, and words with a frequency rating between 5,001 and 7,500 were placed in the third bin.

Items were then assigned a random number using the default random number generator available in the statistical programming language R (R Core Team, 2018) and arranged by frequency bin and random number. In kindergarten, the 15 words in the first frequency bin with the lowest random numbers were selected as the top three rows of the form, the 35 words in the second frequency bin with the lowest random numbers were selected as the next seven rows, and the 35 words in the third frequency bin with the lowest random numbers were selected as the last seven rows. In Grade 1, the 15 words in the first frequency bin with the lowest random numbers were selected as the top three rows of the form, the 45 words in the second frequency bin with the lowest random numbers were selected as the next nine rows, and the 45 words in the third frequency bin with the lowest random numbers were selected as the last nine rows.

In Grade 2, the 20 words in the first frequency bin with the lowest random numbers were selected as the top four rows of the form, the 55 words in the second frequency bin with the lowest random numbers were selected as the next 11 rows, and the 55 words in the third frequency bin with the lowest random numbers were selected as the last 11 rows. In Grade 3, the 30 words in the first frequency bin with the lowest random numbers were selected as the top six rows of the form, the 55 words in the second frequency bin with the lowest random numbers were selected as the next 11 rows, and the 55 words in the third frequency bin with the lowest random numbers were selected as the last 11 rows.

This process, including the generation of a new set of random numbers, was repeated 30 (in kindergarten and Grade 1) to 40 (in Grades 2 and 3) times per grade to create a pool of potential forms. Multiple research staff reviewed each form, from which three were selected as the benchmark forms, and an additional 20 were selected as the progress monitoring forms.

Oral Reading Fluency development process. Rather than hiring item writers to author the new ORF passages, we hired published and aspiring short story authors: Rose Gowen, Kristen Havens, Sarah Meacham, Ben Seipel, Bob Thurber, Tina Truitt, and Andrew Wilson. Rose Gowen is an American writer and mother of two living in Montreal who has been published in the *American Poetry Review*, *Night Train*, and *McSweeney's* among other venues and attended the 2018 Bread Loaf Writers' Conference. Kristen Havens is a writer and editor living in Los Angeles, CA, who has written for many clients including the Special Olympics, has received multiple honorable mentions from *Glimmer Train*, and was nominated for the PEN/Robert J. Dau Short Story Prize for Emerging Writers. Sarah Meacham is a writer, anthropologist, and mother living in Los Angeles, CA, who was a staff writer for the UCLA Division of Social Sciences and External Affairs and worked with the Strategic Education Research Partnership in Boston Public Schools. Ben Seipel is an Assistant Professor at California State University, Chico, and is an aspiring author who taught Spanish in K-12 in Minnesota for many years. Bob Thurber is an author and father living in North Attleboro, MA, who has published two novels and innumerable short stories, appeared in over 50 short story anthologies, and won more than 20 writing awards. Tina Truitt is an author, mother of three, and preschool teacher living in Cherry Hill, NJ, who has published two books, including a children's multicultural, bilingual picture book about teamwork. Andrew L. Wilson is an author and editor living in Eugene, OR, who has published poetry and short stories in a wide range of venues, including *Exquisite Corpse* and *In Posse Review*, and has edited the online literary journal *Linnaean Street* as well as academic books and technical reports. The authors come from diverse socio-economic and cultural backgrounds.

Authors were given detailed specifications to guide them in writing their assigned passages, which included narrative and informational texts for multiple grade levels. Specifications for passage length and Flesch-Kincaid grade-level readability were also provided (see Table 1.7). Authors were coached to represent diverse experiences in terms of culture, geography, and locale, as well as to avoid hackneyed and culturally sensitive topics.

In addition, authors were asked to give each passage a relatively short title that did not give away the ending, as well as use standard English formatting and grammar and grade-level appropriate topics and vocabulary. Narrative texts were required to have a discrete beginning, middle, and end, with multiple episodes or events in the middle. Informational texts were required to have a clear introduction and conclusion with intermediate paragraphs that

provided supporting details, and where possible utilize text structures frequently used in the elementary grades (i.e., compare-contrast, cause-effect, problem-solution, and sequence). Authors were also asked to avoid dialogue, headings, slang, italics, and bold font, as well as content that could be considered religious, controversial, or offensive to some cultures. Finally, they were encouraged to refrain from writing passages that were too funny or emotional, consistently similar in style and tone, or overly arcane or familiar in topic.

Table 1.7 DIBELS 8th Edition Oral Reading Fluency Passage Writing Specifications

Grade	Required length in words	Target Flesch-Kincaid grade level
1	150-200	1.5-2.0
2	150-200	2.5-3.0
3	175-225	3.5-4.0
4	175-225	4.5-5.0
5	200-250	5.5-6.0
6	200-250	6.5-7.0
7	250-300	7.5-8.0
8	250-300	8.5-9.0

Once passages were turned in by the authors, the DIBELS 8th Edition development team reviewed them for consistency with the specifications. In cases where passages diverged from these specifications, passages were revised by the DIBELS 8th team in cases where the passage was deemed salvageable. Others were discarded at this stage. Grade level was determined by readability level (i.e., Flesch-Kincaid grade level).

Next, all passages were reviewed by a team of external reviewers who were parents and/or former teachers with experience with K-8 students and settings. Reviewers were trained by familiarizing them with oral reading fluency measures and the purposes of the review, as well as the criteria by which they would evaluate stories. They reviewed the passages for grade-level appropriateness of their vocabulary, syntax, sentence length, and overall content, as well as the background knowledge required for comprehension.

They also indicated when passages were likely to evoke an emotional reaction from readers that might interfere with reading rate (e.g., laughing out loud, gasping in surprise).

In addition, they were asked to rate passages for how accessible and enjoyable they were for slow and struggling readers, helping to ensure that the first few sentences were not overly difficult and provided a hook (or schema) that supported comprehension. Furthermore, they reviewed passages for potential bias, indicating whenever they judged a passage as potentially offensive to readers or teachers based on gender, ethnicity, race, national origin, religion, disability status, sexual orientation, and geographical region. They were also asked to rate potential for bias due to passage topic and tone, especially bias toward students from backgrounds typically under-represented in children's texts. Note that potential bias in ORF passages was also addressed through sensitivity analyses of classification accuracy for readers for different backgrounds. This information can be found in Chapter 6.

Finally, reviewers indicated if a given passage might be as or more appropriate for other grade levels. Importantly, the training emphasized that reliability of ratings was not a goal and diversity of opinions was perfectly acceptable.

Once passages had been reviewed by two or more of the panel members, DIBELS 8th Edition researchers analyzed ratings and revisited all passages where reviewers noted one or more problems. In some cases, passages were immediately discarded. Judgments regarding vocabulary inappropriateness were supplemented with checks of word frequencies and age of acquisition, and in cases where the inappropriateness was confirmed, a more grade-appropriate substitution was made. Judgments regarding syntactic complexity resulted almost uniformly in similar revisions.

Of particular importance was the content appropriateness, which resulted in passages being considered for assignment to higher and lower grade levels than their readability would suggest. These judgments were sometimes based on the background knowledge required to comprehend a passage, but also often relied on the sophistication of literary and rhetorical devices and overall conceptual complexity. In such cases, some effort was made to increase or decrease readability to improve apparent “fit” with the new grade level assignment. Nonetheless, current consensus is that the appropriate grade level of reading material is more than a matter of strict readability. Thus, given that oral reading fluency is intended to act as an indicator of reading comprehension (rather than strictly of decoding skill efficiency), some passages were assigned to higher and lower grade levels even when readability did not strictly match the assigned grade (see Appendix A).

Finally, all passages were field-tested in their targeted grade levels. Passages where reviewers disagreed about text complexity and grade appropriateness were field-tested in multiple grades. The final assignment of passages to grades and benchmark periods was based on student performance on the passages, the predictive validity of specific passages in a given grade, and maintaining a balance of narrative and informational texts. We increased

the diversity of narrative and informational subgenres represented across the intermediate and middle grades. Among the subgenres we included in these grades are fantasy, science fiction, western, and mystery passages. For informational texts, we increasingly varied topics across content areas (e.g., life sciences, earth sciences, ancient history, modern history, biography) and also varied text structures (e.g., compare and contrast, cause and effect, description, problem and solution, and procedural). In many cases, it is difficult to categorize a passage as narrative or informational; for instance, a passage written in the first person about an informational topic could be considered narrative, informational, or an amalgam of both depending on its particular style. Nonetheless, in Grades 1-5, we required that narrative passages make up more than half the passages with an average balance of 60% narrative to 40% informational. In Grades 6-8, we relaxed this requirement and selected more informational passages with an average balance of 40% narrative to 60% informational. Key text readability and complexity statistics are reported for all benchmark passages in Appendix A.

Maze development process. Maze passages were developed in the same manner as ORF passages but went through a few additional steps of development. First, passages were lengthened to reach typical lengths found in other CBMs and in previous DIBELS editions to allow for enough items for appropriate measurement of readers with better fluency and comprehension.

Second, following common rules, the first and last sentences of every passage were left intact, except in Grade 2 where the second sentence was also left intact to allow for better establishment of a situation model for the passage (Kintsch, 1998). Third, beginning with the third word of the second sentence (or third sentence in Grade 2), every seventh word was deleted with a few caveats. If the seventh word was a proper noun or number, then the eighth word was deleted. If the seventh word was highly specialized (e.g., an uncommon scientific term for a given grade), it would not be deleted unless it had occurred previously in the passage. Also, hyphenated words were treated as one word.

Third, the deleted word became one of the answer choices, and two distractors were written for each deleted word. Each distractor was written by a different DIBELS 8th Edition researcher according to a number of rules informed by research. Distractors could not begin with the same letter as the correct word (Conoyer et al., 2017). Distractors were also kept to within two letters in length of the correct answer, although this rule was relaxed in the upper grades (i.e., Grade 5 and beyond). When the deleted word was a noun, verb, or adjective, distractors had to be grammatically correct. For instance, if the word to be chosen followed “an”, then the distractors had to begin with a vowel. When the deleted word was a contraction, all distractors also had to be contractions and tense agreement was deemed unimportant. Different forms of the same word were never used as distractors (e.g., “be”, “is”, and “are”). For all other parts of speech, grammatical correctness was not a requirement because it was

found to result in repetitive distractors. For example, when the deleted word was an article, requiring grammatical correctness resulted in the answer choices always being “a”, “an”, and “the.” It was deemed undesirable to have answer choices repeat too frequently. Finally, in Grade 5 and up, one of the distractors was required to have semantic similarity to the correct word. That is, it could make sense in a given sentence but not in the story as a whole.

Once distractors were written, they were reviewed by another DIBELS 8th Edition researcher, who would make corrections when rules were violated. If the reviewer found a particular item to be inordinately difficult, the item was brought to a subset of researchers for discussion and potential revision. Finally, the answer choices were reordered so that they were always listed alphabetically.

Benchmark passages were selected from the resulting pool using rules that balanced readability, text complexity, and Lexile ratings (see Table 1.8). In order to balance these factors, readability grade levels were permitted to go above grade level in all but second grade. Key text readability and complexity statistics are reported for all benchmark passages in Appendix B.

Table 1.8 DIBELS 8th Edition Maze Benchmark Passage Selection Specifications

Grade	Required length in words	Target Flesch-Kincaid grade level	Lexile	Coh-Metrix narrativity score
2	350+	2.0-2.9	500L – 600L	80+
3	350+	3.0-4.9	500L – 600L	70+
4	400+	4.0-5.9	700L – 900L	60–90
5	400+	5.0-7.5	800L – 1000L	50–80
6	400+	6.0-8.5	900L – 1100L	20–70
7	450+	7.0+	900L – 1100L	20–70
8	450+	8.0+	1000L – 1200L	< 70

Summary

This chapter has laid out the history and most recent developments of DIBELS measures. As noted throughout, DIBELS researchers based decisions about DIBELS 8th Edition on the research literature, user feedback, and ongoing research conducted by the University of Oregon (UO). Research into the properties of DIBELS and how to improve its usefulness is ongoing at UO. Regular addendums to this manual will keep DIBELS 8 users up-to-date on the features and technical qualities of DIBELS.

Chapter 1: Administration Instructions and Scoring Procedures

DIBELS® 8th Edition is intended for use with students enrolled in kindergarten through eighth grade. Subtests can be administered to students with or without reading difficulties and disabilities, with frequency of assessment adjusted based on the assessment purpose (e.g., universal screening, progress monitoring).

Examiners who give and interpret DIBELS 8th Edition must receive training in standardized administration and scoring procedures. Standardization ensures reliable scores and allows for comparisons between results and research-determined criteria.

The next section presents general guidelines for administering DIBELS. That section is followed by specific instructions for administering and scoring the five DIBELS 8th Edition subtests: (a) letter naming fluency, (b) phonemic segmentation fluency, (c) nonsense word fluency, (d) word reading fluency, and (e) oral reading fluency. Specific materials required for each subtest are listed in the subtest descriptions. Throughout this chapter, **bold Arial font** is used to indicate scripted directions or prompts provided to the student.

DIBELS 8th Edition General Guidelines

There are a number of common features across the DIBELS 8th Edition individually-administered subtests. For each subtest, the following are provided:

- Applicable grades: the grades for which a subtest is designed;
- Objective: the activity in which the student engages, including administration time;
- Uses: the uses for which the subtest is designed.

In addition, a list of required materials is provided for each subtest, along with detailed administration instructions. Instructions include when to start and stop timing, how to score, and the allowed reminders and prompts. Numerous scoring examples are also given for each subtest.

All DIBELS subtests are best administered in a quiet location where minimal interruptions can be expected. For individually administered measures (i.e., all but Maze), a table or desk separated from distractions is best. When individually administered measures are conducted in a classroom setting, other students should be engaged in quiet, independent activities. The assessment administrator should sit so that the student's face is easily seen and close

enough to easily point to forms and hear what the student says. No matter how close the student and administrator sit, the scoring booklet should not be visible to the student, which is why we recommend using a clipboard.

General Timing and Scoring Guidelines

DIBELS is a timed measurement system. With the exception of Maze, all DIBELS 8th Edition subtests are 60-second timed measures. Maze is a 3-minute timed measure. In all cases, it is critical to time each administration as accurately as possible. Even small mistakes can result in less reliable, and thus less valid scores, and research has shown that timing mistakes are among the most common (Reed, Cummings, Schaper, Lynn, & Biancarosa, 2018). For DIBELS to be a valid assessment, strict adherence to timing conventions is required.

Scoring for all the 60-second subtests has certain commonalities. When 60 seconds have elapsed, the examiner always places a bracket (i.e., **]**) after the last item completed and says, “Stop.” Also, if a student makes an error, put a slash (i.e., **/**) through the incorrect item. If a student makes an error but Self-correctionss the error within 3 seconds, mark SC over the item.

Discontinue Rules

Each subtest has a specific discontinue rule. An assessment should only be discontinued if the specified conditions have been met, or if the administration is irrevocably interrupted (e.g., a fire drill occurs). See the rules for each subtest for its discontinue criteria.

For some subtests at certain time points, not only is that subtest discontinued, but benchmark assessment is also discontinued altogether. The benchmark discontinue rules are explained where applicable and are also summarized in Table 2.1.

Table 2.1 Benchmark Discontinue Rules

Grade and Season	Benchmark Discontinue Rule	Scoring
Kindergarten, Fall	If PSF is discontinued, do not administer NWF and WRF.	Enter 0 for PSF. Do not enter scores for the remaining subtests: NWF and WRF
Kindergarten, Winter	If NWF is discontinued, do not administer WRF.	Enter 0 for NWF. Do not enter scores for the remaining subtest: WRF
First grade, Fall	If WRF is discontinued, do not administer ORF.	Enter 0 for WRF. Do not enter scores for the remaining subtest: ORF

These additional rules are provided for two purposes. They are intended to save time and to spare students unnecessary frustration. The rules were derived from a national field trial that indicated students who scored 0 for the indicated assessments in the periods specified above were extremely unlikely to get any items correct on the remaining subtests. Nonetheless, examiners have the option of administering the remaining subtests based on professional judgment.

Accounting for Situational and Examiner Error

Should the administration of a DIBELS 8th Edition measure be irrevocably interrupted, or otherwise spoiled, the administration should be discontinued, and no score should be recorded. Situations and errors that spoil an administrations include, but are not limited to, the student refusing to participate, the student being too ill to participate, the administrator forgetting to start the timer or missing the end of the 60-second period, and situational interruptions like fire drills.

In all cases where an administration is spoiled, the student should be reassessed using a progress-monitoring form at another time or on another day depending on the administrator's professional judgment. In general, though, the student should be assessed as soon as possible.

Giving Instructions and Encouragement

DIBELS 8th Edition is a standardized assessment, which means test administrators must adhere to scripted procedures for giving students directions in addition to following the timing rules. Test administrators should only say what is provided in the administration instructions and should speak clearly enough for the student to hear well.

Students should not be given feedback on their performance during or after an assessment. If an examiner wishes to give a student general encouragement in between subtests, praising the student's effort is the best (e.g., "Nice effort! I can see you're working to do your best.>").

For many subtests, practice items are provided. Again, the test administrator should adhere to the scripted instructions. Offering additional practice, corrections, or off-script explanations is not allowed. Because DIBELS is an assessment and is used for instructional decision-making, it is critical to determine a student's performance without undue instruction or intervention. The practice items only serve the purpose of ensuring that students comprehend the task at hand.

If a student clearly does not hear or understand instructions or practice items, the test administrator may repeat these procedures once. If the assessment has already begun, the timer should be kept running.

Articulation and Dialect

DIBELS 8th Edition measures early literacy skills in English. Therefore, students should use the English pronunciation of words. However, it is important to mention that students are not penalized for varied pronunciation due to dialect or articulation. For example, if the student consistently says /th/ for /s/ and pronounces "thee" for "see" when naming the letter "C", credit is given for naming the letter correctly. This is a professional judgment and should be based on the student's responses and any prior knowledge of the student's speech patterns.

Different regions of the country use different dialects of American English. The DIBELS 8th Edition Phoneme Pronunciation Guide (see Appendix C) is particularly helpful with the Phonemic Segmentation and Nonsense Word Fluency subtests. These pronunciation examples may be modified consistent with regional dialects and conventions.

An important update to the DIBELS pronunciation guide is the treatment of r-controlled vowels (e.g., word, far), which are sometimes also called r-colored vowels. Considerable disagreement exists about how many phonemes exist in words with r-controlled vowels and thus in American English (e.g., Bizzocchi, 2017; Fry, 2004; Lockenvitz, Kuecker, & Ball,

2015). Whereas earlier editions treated some as single phonemes and others as two or more phonemes, DIBELS 8th Edition simplifies the treatment of r-controlled vowels by treating them as single phonemes. Nonetheless, test administrators should take local dialects and articulation issues into account when scoring nonsense words or phonemic segmentations involving r-controlled vowels. In some regions in the US, r-controlled vowels are more clearly separated into multiple sounds or phonemes (e.g., “lair” might be pronounced as “layer”). Students using such a pronunciation should not lose points for this practice.

Accommodations

The DIBELS 8th Edition measures are designed to be used unmodified with all students. They have been validated with thousands of students using the DIBELS 8th Edition standardized procedures. Interpretation of student scores is only informative when students have been assessed in this standardized way.

In a very small number of cases, however, several **accommodations** are approved. These accommodations should only be used in situations where they are necessary to obtain an accurate score for a student. In other words, accommodations should only be used if there is evidence that without them, the assessment would be measuring something other than the intended reading-related skill. For example, if a student is hard of hearing and without an accommodation the student would not be able to hear the testing directions, then that would result in the test measuring the student’s hearing abilities rather than reading skills. An accommodation would be appropriate in this case.

DIBELS 8th Edition-approved assessment accommodations involve minor changes to assessment procedures that are unlikely to change the meaning of the results and have been approved either by DIBELS developers or assessment professionals. They should be used only when:

- An accurate score is unlikely to be obtained without the accommodation; and/or
- Specified in a student’s 504 plan or Individualized Education Plan (IEP).

The accommodations approved for DIBELS 8th Edition are listed in Table 2.3. When approved accommodations are used, the examiner should mark an “A” on the front cover of the testing booklet. Scores from tests administered with accommodations can be compared to other DIBELS 8th Edition benchmark scores and norms. *Approved accommodations should **only** be used with students who have a documented need for such supports, such as a 504 plan or IEP.*

Table 2.3 Acceptable Accommodations for DIBELS 8th Edition

Accommodation	LNF	PSF	NWF	WRF	ORF	Maze
Quiet setting for testing	X	X	X	X	X	X
Breaks in between measures	X	X	X	X	X	X
Assistive technology (e.g., hearing aids, assistive listening devices, glasses)	X	X	X	X	X	X
Enlarged student materials	X		X	X	X	X
Colored overlays, filters, or lighting adjustments	X		X	X	X	X
Marker or ruler for tracking	X		X	X	X	X

Anything an assessor does that is not listed in the standardized scoring and administration and is not an approved accommodation falls under the category of a **modification**. Any modification made to the standardized directions, timing or scoring rules renders results that are likely to be meaningfully different than they would have been without the modification. Examples of unapproved accommodations and modifications include: (a) extending the time on a DIBELS probe, (b) repeating practice items, (c) providing different or extra models of the task, (d) adding to or changing administration directions, and (e) offering unapproved prompts and feedback.

When unapproved accommodations or modifications are used, the examiner should mark an “M” on the front cover of the testing booklet. Scores are *not* valid in these cases and should not be entered in a data system or interpreted in relation to DIBELS 8th Edition benchmark goals and norms.

It is important to recognize that there are some students for whom DIBELS is not an appropriate assessment. Students for whom this is true include those:

- With limited verbal language skills,
- With fluency-based speech disorders or oral apraxia, and/or
- For whom reading in English is not an instructional goal (e.g., students learning to read exclusively in a language other than English).

In these cases, other assessments and curricular tools (e.g., end-of-unit tests, individualized progress monitoring materials, other-language reading assessments) are best suited to screening students and monitoring student progress toward goals.

Letter Naming Fluency (LNF)

Applicable grades: Beginning of kindergarten through end of first grade.

Objective: Student names letters for 60 seconds.

Uses: Benchmark and risk assessment.

Materials

- Scoring book
- Student form
- Pen or pencil
- Clipboard
- Timer

Administration

1. Position the clipboard and scoring book so that the student cannot see what you record.
2. Place the student copy of the LNF subtest in front of the student.
3. Say these specific directions:

Here are some letters

(Point to the student form).

Tell me the names of as many letters as you can.

When I say “Begin,” start here,

(Point to the first letter)

and go across the page

(Point).

Point to each letter and tell me the name of that letter.

If you come to a letter you don't know, I'll tell it to you.

Put your finger on the first letter.

Ready?

Begin.

4. Start the timer after saying "Begin."
5. Follow along in the Scoring Booklet. Put a slash (/) through each letter name read incorrectly. See Acceptable Prompts and Scoring Rules for more details.
6. At the end of 60 seconds, place a bracket (]) after the last letter named and say, **"Stop."**

Acceptable prompts

There are two acceptable prompts for LNF: a prompt for when students hesitate and for when they produce letter sounds.

Hesitation Prompt. If the student hesitates for 3 seconds on a letter, score the letter as incorrect, provide the correct letter, point to the next letter, and say:

Keep going.

This prompt may be repeated. For example, if the letters are "p T n" and the student says, "p" then does not say anything for 3 seconds, prompt by saying **"T"**, then point to "n" and say:

Keep going.

Repeat this as many times as needed throughout administration. The maximum time for each letter is 3 seconds.

Letter Sound Prompt. If the student provides the letter sound rather than the letter name, say:

Remember to tell me the letter name, not its sound.

This prompt may be provided *once* during the administration. If the student continues providing letter sounds, mark each letter as incorrect.

Discontinue rules

Discontinue LNF Rule. If the student reads 0 correct letter names within the first line, discontinue LNF, put a bracket after the last letter attempted and record a score of 0.

Discontinue Benchmark Assessments Rule. Benchmark assessment always continues regardless of LNF score.

Scoring rules

LNF provides one score: the number of letters named correctly. Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Do not mark correct responses on the scoring book.
Incorrect responses	Make a slash (/) through each letter named incorrectly.
Self-corrections	If a student makes an error but corrects it within 3 seconds, write “SC” above the letter and score it as correct.

Situation	How to score
-----------	--------------

Letter reversals A letter is incorrect if the student substitutes a different letter for the stimulus letter, even if the substituted letter is similar in appearance. (Note that lowercase l does not appear on LNF forms, and the font used in LNF distinguishes the uppercase I from the lowercase L and number one very well.)

Letters	Student Says	Scoring Procedure	Correct Letters
b T n E	"d...T...n...E"	b T n E	<u>3</u> /4
p S n L	"q...S...m...L"	p S n L	<u>2</u> /4
M I k L	"M...L...k...L"	M I k L	<u>3</u> /4

Letter sounds A letter is incorrect if the student provides the letter-sound for the stimulus letter (e.g., /d/ for "D"). A prompt for providing letter-sounds is allowable only once (see Acceptable Prompts).

Letters	Student Says	Scoring Procedure	Correct Letters
b T n E	"/b/...T...n...E"	b T n E	<u>3</u> /4
p S n L	"p.../s/...m...L"	p S n L	<u>3</u> /4
M I k L	"M...l.../k/...L"	M I k L	<u>3</u> /4

Omissions A letter is incorrect if the student skips the letter. If the student skips an entire line, cross out the line and record a score of 0 for that line

LNF Fidelity of Administration

The observer should judge the full test administration. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner's accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Phonemic Segmentation Fluency (PSF)

Applicable grades: Beginning of kindergarten through end of first grade.

Objective: Student breaks words into phonemes for 60 seconds.

Uses: Benchmark and risk assessment; progress monitoring.

Materials

- Scoring book
- Pen or pencil
- Clipboard
- Timer

Administration

1. Position the clipboard and timer so that the student cannot see what you record.
2. Say these specific directions:

I am going to say a word.

After I say it, you tell me all the sounds in the word.

So, if I say “am,” you would say “/a/ /m/.”

Let’s try one.

(1 second pause)

Tell me the sounds in “it.”

Student response	Examiner response
CORRECT	Very good.
If student says “/i/ /t/”	The sounds in “it” are /i/ /t/.
INCORRECT	The sounds in “it” are /i/ /t/.
Any other response	Your turn. Tell me the sounds in “it.”

OK. Here is your first word.

3. Give the student the first word and start the timer.
4. Follow along in the Scoring Booklet. As the student says the sounds, underline each different, correct, sound segment produced. Put a slash (/) through sounds produced incorrectly. See Acceptable Prompts and Scoring Rules for more details.
5. As soon as the student is finished saying the sounds in the current word, present the next word promptly and clearly.
6. At the end of 60 seconds, stop presenting words and stop the timer. Place a bracket (]) after the last sound provided by the student.

Acceptable prompts

There is only one acceptable prompt for PSF: a prompt for when students hesitate.

Hesitation Prompt. If the student hesitates for 3 seconds, give the next word, and score the word (or remaining sounds in the word if word has been partially segmented) as incorrect by leaving it unmarked (no slashes or underlines). Repeat this prompt as many times as needed throughout administration.

Discontinue rules

Discontinue PSF Rule. If a student has not given any sound segments correctly in the first 5 words, discontinue PSF, put a bracket after the last word attempted and record a score of 0.

Discontinue Benchmark Assessments Rule. *For beginning of kindergarten only*, if student does not get any sounds correct in the first 5 words, discontinue PSF and any

further benchmark assessments (i.e., NWF and WRF) for that time of year. At all other times of year, benchmark assessment continues regardless of PSF score.

Scoring rules

PSF provides one score: the sum of sound segments produced. Students receive 1 point for each different, correct, part of the word. Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Underline the sound segments in the word the student produces that are correctly pronounced.
Incorrect responses	Make a slash (/) through sounds pronounced incorrectly. Circle the item if the student repeats the word correctly, but without segmentation.
Self-corrections	If a student makes an error but corrects it within 3 seconds, write "SC" above the phoneme and score it as correct.

Situation	How to score
Schwa sounds	Schwa sounds (/u/) added to consonants are not counted as errors. Some phonemes cannot be pronounced correctly in isolation without a vowel, and some early learning of sounds includes the schwa. For example, if the word is "track," and the student says "tu...ru...a...ku" they would receive 4 of 4 points.

Word	Student Says	Scoring Procedure	Correct Segments
track	"tu...ru...a...ku"	/t/ /r/ /a/ /k/	<u>4</u> /4
bet	"bu...e...tu"	/b/ /e/ /t/	<u>3</u> /4

Additions

Additions are not counted as errors if they are separated from the other sounds in the word. For example, if the word is “track,” and the student says “t...r...a...ck...s,” they would receive 4 of 4 points.

Word	Student Says	Scoring Procedure	Correct Segments
track	“t...r...a...ck...s”	/t̥/ /r̥/ /ḁ/ /k̥/	<u>4</u> /4
top	“s...t...o...p”	/t̥/ /o̥/ /p/	<u>3</u> /3
top	“st...o...p”	/t̥/ /o̥/ /p/	<u>2</u> /3
top	“s...t...ol...p”	/t̥/ /ə/ /p/	<u>2</u> /3
top	“s...t...o...l...p”	/t̥/ /o̥/ /p/	<u>3</u> /3

**Sound
elongation**

The student may elongate the individual sounds and run them together as long as it is clear he or she is aware of each sound individually. For example, if the student says, “ssssuuuunnnn,” with each phoneme held long enough to make it clear they know the sounds in the word, they would receive credit for 3 phonemes correct. This is a professional judgment and should be based on the student’s responses and prior knowledge of the student’s instruction. When in doubt, no credit is given.

Word	Student Says	Scoring Procedure	Correct Segments
sun	“ssssuuuunnnn”	/s̥/ /u̥/ /n̥/	<u>3</u> /3

Partial segmentation

The student is given credit for each correct sound segment, even if they have not segmented to the phoneme level. Use the underline to indicate the size of the sound segment. For example, if the word is “track,” and the student says “tr...ack,” they would receive 2 of 4 points.

Word	Student Says	Scoring Procedure	Correct Segments
track	“tr...ack”	<u>/t/</u> <u>/r/</u> <u>/a/</u> <u>/k/</u>	<u>2</u> /4
bet	“b...et”	<u>/b/</u> <u>/e/</u> <u>/t/</u>	<u>2</u> /3

Overlapping segmentation

The student receives credit for each different, correct, sound segment of the word. Thus, if the word is “track,” and the student says “tra...ack,” the student would receive 2 of 4 points because /tra/ and /ack/ are both different, correct, sound segments of “track.”

Word	Student Says	Scoring Procedure	Correct Segments
track	“tra...ack”	<u>/t/</u> <u>/r/</u> <u>/a/</u> <u>/k/</u>	<u>2</u> /4
bet	“be...e...et”	<u>/b/</u> <u>/e/</u> <u>/t/</u>	<u>3</u> /3

Mispronounced segment

The student does not receive credit for sound segments that are mispronounced. For example, if the word is “track,” and the student says “t...r...a...gs” they would receive no credit for /gs/ because there is no /g/ or /s/ sound segment in the word “track.”

Word	Student Says	Scoring Procedure	Correct Segments
track	“t...r...a...gs”	<u>/t/</u> <u>/r/</u> <u>/a/</u> <u>/k/</u>	<u>3</u> /4
bet	“p...i...t”	<u>/b/</u> <u>/e/</u> <u>/t/</u>	<u>1</u> /3
bet	“d...e...t”	<u>/b/</u> <u>/e/</u> <u>/t/</u>	<u>2</u> /3

No segmentation

If the student repeats the entire word, no credit is given for any sounds. For example, if the word is “track,” and the student says “track,” circle the entire word and record zero points.

Word	Student Says	Scoring Procedure	Correct Segments
track	“track”	<u>/t/</u> /r/ /a/ /k/	<u>0</u> /4

Spelling

If the student spells the word, no credit is given. For example, if the word is “track,” and the student says “t ...r...a...c...k”, cross out each sound.

Word	Student Says	Scoring Procedure	Correct Segments
track	“t...r...a...c...k”	/t/ /r/ /a/ /k/	<u>0</u> /4

Omissions

A sound is incorrect if the student omits the sound, but the sound is left unmarked.

Word	Student Says	Scoring Procedure	Correct Segments
track	“tr...” (3 seconds)	<u>/t/</u> /r/ /a/ /k/	<u>1</u> /4
bet	“b... t”	/b/ /e/ <u>/t/</u>	<u>2</u> /3

PSF Fidelity of Administration

The observer should judge the full test administration. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner’s accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Nonsense Word Fluency (NWF)

Applicable grades: Beginning of kindergarten through end of third grade.

Objective: Student reads or sounds out nonsense words for 60 seconds.

Uses: Benchmark and risk assessment; progress monitoring.

Materials

- Scoring book
- Student form
- Pen or pencil
- Clipboard
- Timer

Administration

1. Position the clipboard and timer so that the student cannot see what you record.
2. Place the student copy of the NWF practice items in front of the student.
3. Say these specific directions:

Look at this word.

(point to first word on the practice form)

It's a make-believe word.

Watch me read the word: /h/ /a/ /p/, "hap."

(point to each letter, then run your finger fast beneath the whole word)

I can say the sounds of the letters, /h/ /a/ /p/

(point to each letter)

or I can read the whole word "hap."

(run your finger fast beneath the whole word)

Your turn to read a make-believe word.

Read this word the best you can.

(point to the word “lum”)

Make sure you say any sounds you know.

Student response	Examiner response
CORRECT	That’s right.
If student says “lum” or “/l/ /u/ /m/”	The sounds are “/l/ /u/ /m/” or “lum”.
INCORRECT	Remember, you can say the sounds or you can say the whole word.
Any other response	Watch me: the sounds are “/l/ /u/ /m/.” (point to each letter) Or “lum.” (run your finger fast beneath the whole word) Let’s try again. Read this word the best you can. (point to the word “lum”)

(place the student copy of the form in front of the student)

Here are some more make-believe words.

(point to the student form)

Start here

(point to the first nonsense word)

and go across the page

(point across the page)

When I say “Begin,” read the words the best you can.

Point to each letter and tell me the sound or read the whole word.

Put your finger on the first word.

Ready?

Begin.

4. Start the timer after saying “Begin.”
5. Follow along in the Scoring Booklet. As the student says sounds/words, underline each correct sound/word produced. Put a slash (/) through sounds/words produced incorrectly. See Acceptable Prompts and Scoring Rules for more details.
6. At the end of 60 seconds, place a bracket (]) after the last nonsense word for which the student provided sound/word and say, “**Stop.**”

Acceptable prompts

There is only one acceptable prompt for NWF: a prompt for when students hesitate. Execution of the prompt depends on whether a student is initially blending nonsense words or sounding them out. If the student is reading words, the rule applies to words; if the student is sounding words out, the rule applies to sounds.

Hesitation Prompt. If student hesitates for 3 seconds on a sound/word, mark the sound/word as incorrect, point to the next sound/word, and say:

Keep going.

Repeat this as many times as needed throughout administration. The maximum time for each sound/word is 3 seconds.

Discontinue rules

Discontinue NWF Rule. If a student does not get any sounds correct in the first 5 words, discontinue NWF, put a bracket after the last nonsense word attempted and record a score of 0 for both CLS and WRC.

Discontinue Benchmark Assessments Rule. *For middle of kindergarten only*, if student does not get any sounds correct in the first 5 words, discontinue NWF and any further benchmark assessments for that time of year (i.e., WRF). At all other times of year, benchmark assessment continues regardless of NWF score.

Scoring rules

NWF provides two scores: the sum of correct letter sounds (CLS) and the sum of words read or recoded correctly (WRC). Every correct letter sound receives 1 point for CLS, regardless of whether a student blends. Words read correctly, whether sounded out initially or not, receive 1 point each for WRC. Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Underline the letters that the student produces correctly. Underline multiple letters for partially blended words and whole words for fully blended words (with or without sounding out initially).
Incorrect responses	Make a slash (/) through sounds/words produced incorrectly.
Self-corrections	If a student makes an error but corrects it within 3 seconds, write "SC" above the phoneme and score it as correct.

Situation	How to score																				
Sounds followed by word	When a student sounds out a nonsense word and then blends it, underline the individual letters and then the nonsense word as a whole and score a 3 for CLS and a 1 for WRC.																				
	<table><tr><th>Word</th><th>Student Says</th><th>Scoring Procedure</th><th colspan="2">Score</th></tr><tr><td></td><td></td><td></td><th>CLS</th><th>WRC</th></tr><tr><td>rab</td><td>"/r/.../a/...rab"</td><td><u>/r/</u> <u>/a/</u> <u>/b/</u></td><td><u>3</u> /3</td><td><u>1</u> /1</td></tr><tr><td>mot</td><td>"/m/.../o/.../t/...mot"</td><td><u>/m/</u> <u>/o/</u> <u>/t/</u></td><td><u>3</u> /3</td><td><u>1</u> /1</td></tr></table>	Word	Student Says	Scoring Procedure	Score					CLS	WRC	rab	"/r/.../a/...rab"	<u>/r/</u> <u>/a/</u> <u>/b/</u>	<u>3</u> /3	<u>1</u> /1	mot	"/m/.../o/.../t/...mot"	<u>/m/</u> <u>/o/</u> <u>/t/</u>	<u>3</u> /3	<u>1</u> /1
Word	Student Says	Scoring Procedure	Score																		
			CLS	WRC																	
rab	"/r/.../a/...rab"	<u>/r/</u> <u>/a/</u> <u>/b/</u>	<u>3</u> /3	<u>1</u> /1																	
mot	"/m/.../o/.../t/...mot"	<u>/m/</u> <u>/o/</u> <u>/t/</u>	<u>3</u> /3	<u>1</u> /1																	

Repeated sounds

Letter sounds given twice receive credit once. For example, if stimulus word is “rab” and the student says /r/ /a/ /ab/, the student receives only 1 point for the letter sound “a” even though the correct sound was provided twice, and a total CLS score of 3 and a total WRC score of 0.

Word	Student Says	Scoring Procedure	Score	
			CLS	WRC
rab	“r...a...ab”	/r/ / <u>a</u> / /b/	<u>3</u> /3	<u>0</u> /1
mot	“m...o...t...mo...t”	/ <u>m</u> / / <u>o</u> / / <u>t</u> /	<u>3</u> /3	<u>0</u> /1

Partially correct responses

If a word is partially correct, underline the corresponding letters for the sounds produced correctly and word parts for any sounds blended. Put a slash (/) through incorrectly produced letter sounds (to distinguish from omissions; see Omissions scoring rule). For example, if the word is “rab” and the student says “rayb” (with a long /a/), the letters “r” and “b” would be underlined, and the letter “a” would be slashed with a score of 2 for CLS and 0 for WRC.

Word	Student Says	Scoring Procedure	Score	
			CLS	WRC
rab	“r...ay...b”	/r/ / a / /b/	<u>2</u> /3	<u>0</u> /1
nar	“n...er”	/n / a /	<u>1</u> /2	<u>0</u> /1

Sounds out of order

Letter sounds produced out of order are scored as incorrect. For example, if the stimulus word is “mot” and the student says /t/ /o/ /m/, only /o/, the letter sound read correctly, would be underlined with a score of 1 for CLS and 0 for WRC. This is true even if the student using partial or full blending. Blended letter sounds must be correct and in the correct position (beginning, middle, end) to receive credit (see last two examples). If a student reads a nonsense word using blending, letter sounds produced out of order are scored as incorrect.

Word	Student Says	Scoring Procedure	Score	
			CLS	WRC
mot	“t...o...m”	/ m / / <u>o</u> / /t/	<u>1</u> /3	<u>0</u> /1
mot	“to...om...tom”	/ m / / <u>o</u> / /t/	<u>1</u> /3	<u>0</u> /1
mot	“tom”	/ m / / <u>o</u> / /t/	<u>1</u> /3	<u>0</u> /1
mot	“mob”	/ <u>m</u> / / <u>o</u> / /t/	<u>2</u> /3	<u>0</u> /1
ag	“ga”	/a/ / g /	<u>0</u> /2	<u>0</u> /1

Omissions

If a student skips a word or row, skip marking any slash and move to the next word, row, or page with the student.

NWF Fidelity of Administration

The observer should judge the full test administration. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner’s accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Word Reading Fluency (WRF)

Applicable grades: Beginning of kindergarten through end of third grade.

Objective: Student reads sight words for 60 seconds.

Uses: Benchmark and risk assessment; progress monitoring.

Materials

- Scoring book
- Student form
- Pen or pencil
- Clipboard
- Timer

Administration

1. Position the clipboard and timer so that the student cannot see what you record.
2. Place the student copy of the WRF form in front of the student.
3. Say these specific directions:

Please read from this list of words.

(point to the student form)

Start here

(point to the first word)

and go across the page.

(point across the page)

When I say “Begin,” point to each word and read it the best you can.

If you get stuck, I will tell you the word, so you can keep reading.

Put your finger on the first word.

Ready?

Begin.

4. Start the timer when student says first word.
5. Follow along in the Scoring Booklet. As the student provides responses, put a slash (/) through each word read incorrectly. See Acceptable Prompts and Scoring Rules for more details.
6. At the end of 60 seconds, place a bracket (]) after the last word read and say, “**Stop.**”

Acceptable prompts

There is only one acceptable prompt for WRF: a prompt for when students hesitate.

Hesitation Prompt. If student hesitates for 3 seconds on a word, give the correct word, mark the word as incorrect, point to the next word, and say:

Keep going.

Repeat this as many times as needed throughout administration. The maximum time for each word is 3 seconds.

Discontinue rules

Discontinue WRF Rule. If a student does not get any words correct in the first line (5 words), discontinue WRF, put a bracket after the last word attempted and record a score of 0.

Discontinue Benchmark Assessments Rule. *For beginning of first grade only*, if student does not get any sounds correct in the first 5 words: discontinue WRF and any further benchmark assessments for that time of year (i.e., ORF). At all other times of year, benchmark assessment continues regardless of WRF score.

Scoring Rules

WRF provides one score: the sum of words read correctly. Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Do not mark correct items on the scoring book.
Incorrect responses	Put a slash (/) through words produced incorrectly.

Self-corrections

If a student makes an error and corrects it within 3 seconds, write “SC” above the word and score it as correct.

Situation	How to score												
Sounded out words	<p>If a word is sounded out without blending, it is incorrect. If a word is sounded out and then blended, it is correct.</p> <table><tr><th>Words</th><th>Student Says</th><th>Scoring Procedure</th><th>Score</th></tr><tr><td>joy draw cloud</td><td>“/j/ /oy/ draw cloud”</td><td>joy draw cloud</td><td><u>2</u> /3</td></tr><tr><td>joy draw cloud</td><td>“/j/ /oy/ joy draw cloud”</td><td>joy draw cloud</td><td><u>3</u> /3</td></tr></table>	Words	Student Says	Scoring Procedure	Score	joy draw cloud	“/j/ /oy/ draw cloud”	joy draw cloud	<u>2</u> /3	joy draw cloud	“/j/ /oy/ joy draw cloud”	joy draw cloud	<u>3</u> /3
Words	Student Says	Scoring Procedure	Score										
joy draw cloud	“/j/ /oy/ draw cloud”	joy draw cloud	<u>2</u> /3										
joy draw cloud	“/j/ /oy/ joy draw cloud”	joy draw cloud	<u>3</u> /3										
Word order	<p>Words read correctly but in the wrong order are scored as incorrect.</p> <table><tr><th>Words</th><th>Student Says</th><th>Scoring Procedure</th><th>Score</th></tr><tr><td>joy draw cloud</td><td>“joy cloud draw”</td><td>joy draw cloud</td><td><u>1</u> /3</td></tr></table>	Words	Student Says	Scoring Procedure	Score	joy draw cloud	“joy cloud draw”	joy draw cloud	<u>1</u> /3				
Words	Student Says	Scoring Procedure	Score										
joy draw cloud	“joy cloud draw”	joy draw cloud	<u>1</u> /3										
Omissions	<p>A word is incorrect if the student skips the word. If the student skips an entire line, cross out the line and record a score of 0 for that line.</p>												

WRF Fidelity of Administration

The observer should judge the full test administration. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner’s accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Oral Reading Fluency (ORF)

Applicable grades: Beginning of first grade through end of eighth grade.

Objective: Student reads a passage aloud for 60 seconds.

Uses: Benchmark and risk assessment; progress monitoring.

Materials

- Scoring book
- Student form
- Pen or pencil
- Clipboard
- Timer

Administration

1. Position the clipboard and timer so that the student cannot see what you record.
2. Place the student copy of the ORF form in front of the student.
3. Say these specific directions:

Please read this

(point to the 1st word of the 1st paragraph of the passage)

out loud.

If you get stuck, I will tell you the word, so you can keep reading.

When I say “Stop” I may ask you to tell me about what you read, so do your best reading.

Start here

(point to the first word of the passage).

Ready?

Begin.

4. Start the timer when the student says the first word of the passage. Do NOT count the title. If the student fails to say the first word after 3 seconds, tell the student the word and mark it as incorrect, then start the timer.
5. Follow along in the Scoring Booklet. As the student provides responses, put a slash (/) through each word read incorrectly. See Acceptable Prompts and Scoring Rules for more details.
6. At the end of 60 seconds, place a bracket (]) after the last word read and say, “**Stop.**”

Acceptable prompts

There is only one acceptable prompt for ORF: a prompt for when students hesitate.

Hesitation Prompt. If student hesitates for 3 seconds on a word, give the correct word, and mark the word as incorrect. Repeat this as many times as needed throughout administration. The maximum time for each word is 3 seconds.

Discontinue rules

Discontinue ORF Rule. If the student does not read any words correctly in the first line of the passage, discontinue ORF, put a bracket after the last word attempted and record a score of 0.

Discontinue Benchmark Assessments Rule. Benchmark assessment always continues regardless of ORF score.

Scoring rules

ORF provides two scores: the sum of words read correctly and an accuracy percentage. The accuracy percentage is calculated by dividing the sum of words read correctly by the number of total words attempted (including errors) and multiplying by 100:

$$\text{Accuracy} = \frac{\text{words read correctly}}{\text{total words read}} \times 100$$

Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Do not mark correct items on the scoring book.
Incorrect responses	Put a slash (/) through words produced incorrectly.
Self-corrections	If a student makes an error and corrects it within 3 seconds, write "SC" above the word and score it as correct.

Situation	How to score
-----------	--------------

Insertions

Inserted words are ignored and not counted as errors. The student does not get points for inserted words. If the student frequently inserts extra words, it may be worth noting the pattern at the bottom of the scoring page.

Passage	Student Says	Scoring Procedure	Score
I have a dog.	"I have a new dog."	I have a dog.	<u>4</u> /4
The walk was fun.	"The walk was really fun."	The walk was fun.	<u>4</u> /4

Repetitions

Words that are repeated are not scored as incorrect so long as they are read correctly. They are treated as insertions and ignored in scoring.

Passage	Student Says	Scoring Procedure	Score
I have a dog.	"I have a ... I have a dog."	I have a dog.	<u>4</u> /4

Sounded out words

A word is scored as incorrect if it is sounded out correctly but not blended. If it is blended, it is scored as correct.

Passage	Student Says	Scoring Procedure	Score
We like to read.	"We like to rrrr ... eeee ... d read."	We like to read.	<u>4</u> /4
We like to read.	"We like to rrrr ... eeee ... d."	We like to read .	<u>3</u> /4

Abbreviations

Abbreviations should be read in the way you would normally pronounce the abbreviation in conversation. For example, ASAP could be read as "ay ess ay pea" or "ay sap" and Dr. would be read as "doctor."

Passage	Student Says	Scoring Procedure	Score
Tell me ASAP.	"Tell me ay ess ay pea."	Tell me ASAP.	<u>3</u> /3
Tell me ASAP.	"Tell me ay sap."	Tell me ASAP.	<u>3</u> /3
Dr. Jones looked at my teeth.	"Doctor Jones looked at my teeth."	Dr. Jones looked at my teeth.	<u>6</u> /6
Dr. Jones looked at my teeth.	"'D' 'r' Jones looked at my teeth."	Dr. Jones looked at my teeth.	<u>5</u> /6

Mispronounced words

A word is scored as incorrect if it is pronounced incorrectly in the context of the sentence.

Passage	Student Says	Scoring Procedure	Score
We like to read.	"We like to red."	We like to read .	<u>3</u> /4

Word order

All words that are read correctly but in the wrong order are scored as incorrect.

Passage	Student Says	Scoring Procedure	Score
The green park has flowers.	"The park green has flowers."	The green park has flowers.	<u>3</u> /5

Omissions

Omitted words are scored as incorrect. If a student skips an entire row, cross out the row and mark the skipped words incorrect.

ORF Fidelity of Administration

The observer should judge the full test administration. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner's accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Maze

Applicable grades: Beginning of second grade through end of eighth grade.

Objective: Student silently reads a passage for 180 seconds, choosing the best multiple-choice answer for missing words.

Uses: Benchmark and risk assessment; progress monitoring.

Materials

- Maze administration directions and scoring key
- Student worksheets (one per student)
- Pen or pencil (one per student)
- Clipboard
- Timer

Administration

1. Say:

I am going to give you a worksheet. When you get your worksheet, please write your name at the top and put your pencil down.

2. Hand out the Maze student worksheets.
3. Make sure students have written their names down before proceeding.
4. Say these specific directions:

You are going to read a passage with some words missing from it. For each missing word you will see a box with three words in it. Your job is to circle the word you think makes the most sense in the context of the passage. Let's look at the Practice Passage together. Listen as I read.

(Pause)

**Tom goes to a school far from his house. Every morning, he takes
a school**

(Pause)

bus, pen, work

(Pause)

to go to school.

(Pause)

**Let's stop there. Let's circle the word "bus" because I think "bus"
makes the most sense here. Listen to how that sentence sounds
now.**

Every morning, he takes a school bus to go to school.

Now it's your turn. Read the next sentence silently to yourself.

**When you come to a box, read all the words in the box and circle
the word that makes the most sense to you. When you are done,
put your pencil down.**

5. Allow up to 30 seconds for students to complete the example and put their pencils down. If necessary, after 30 seconds say **Put your pencil down.**
6. As soon as all students have their pencils down, say

Good job.

Now listen. In the

(Pause)

afternoon, library, morning,

(Pause)

he also takes a bus home. You should have circled "afternoon"

because “afternoon” makes the most sense.

(Pause)

Listen. In the afternoon, he also takes a bus home.

Okay, when I say “Begin,” turn the page and start reading the passage silently. When you come to a box, read all the words in the box and circle the word that makes the most sense in the passage.

You will stop when you come to a stop sign or I say Stop.

Ready?

Begin.

7. Start the timer.
8. At the end of 3 minutes, stop the timer and say, **Stop. Put your pencils down.**
9. Make sure all students have stopped working and collect all the student worksheets.

Acceptable Prompts

There are two acceptable prompts for Maze: one for when students read aloud and another for when a student stops working.

Student Reading Aloud Prompt. If a student starts reading the passage out loud, say

Please read the passage silently. (Repeat as often as needed.)

Student Stopped Working Prompt. If a student stops working, say

Please keep going until you reach the end of the passage. Just do your best work. (Repeat as often as needed.)

Discontinue rules

There are no discontinue rules for Maze. Every student should be encouraged to try their best until three minutes have passed.

Scoring rules

Maze provides one score that is derived by summing up the number of items answered correctly and subtracting one-half the sum of items answered incorrectly. Worksheets are scored after the assessment has been completed, and students are not present. Use the scoring key to mark answers as correct or incorrect.

1. A response is correct if the student clearly circled or otherwise marked (e.g., underlined or checked) the correct word.
2. Mark a slash (/) through any incorrect responses. Incorrect responses include situations when the wrong answer is circled or otherwise marked, more than one answer is marked, or an item is left blank (only if it occurs before the final item answered).
3. If there are erasure marks, scratched out words, or any other extraneous markings, and the student's final response is obvious, score the item based on the final response.
4. Items left blank after the last response are not slashed or counted as incorrect.
5. Count the number of items answered that are not slashed to obtain the number of items answered correctly. Enter the total next to the word Correct on the student's booklet.
6. Count the number of items marked with a slash. Enter the total next to the word Incorrect on the student's booklet.
7. Calculate the adjusted score (unnecessary for DIBELS Data System and Amplify customers) using the following formula:

$$\text{Maze} = \text{Correct} - \frac{\text{Incorrect}}{2}$$

By definition, this formula will sometimes result in scores with decimal values. **These scores should not be rounded.**

Mark student responses according to the rules in the first table below. The second table provides several examples of common situations and how to score in them.

Correct responses	Do not mark correct items.
Incorrect responses	Put a slash (/) through items answered incorrectly, skipped (before the last valid response), or marked in a confusing manner.
Self-corrections	If a student makes corrections to a response, the answer is counted as correct so long as the final intended answer is both clear and correct.

Situation	How to score
Inconsistent marking	<p>Students sometimes change how they mark the correct answer. So long as the student's intention is clear and correct, changes in marking system are not penalized. In the example below, the student gets 3 items correct and none incorrect.</p> <div> <div> <div>buys</div> <div>food</div> </div> <div> <div>Tom goes a school bus to go to radio . In the afternoon,</div> <div>takes school</div> </div> <div> <div>→ a</div> <div>he also takes few bus home.</div> <div>it</div> </div> </div>

Skipped items

Skipped items are marked incorrect when they are clearly skipped (i.e., a later item is answered), as in the first example below. They are left unmarked and not counted as correct or incorrect if no subsequent item is answered, as in the second example below. In the first example, the student gets 2 correct and 1 incorrect. In the second example, the student gets 1 correct and none incorrect.

buys food
Tom goes a school bus to go to ~~radio~~ . In the afternoon,
takes school

a
he also takes few bus home.
it

buys food
Tom goes a school bus to go to radio . In the afternoon,
takes school

a
he also takes few bus home.
it

Unclear or multiple responses

When more than one choice is marked, and the intended final answer is not clear, the item is scored as incorrect. In the example below, the child gets no items correct and 3 incorrect.

buys food
Tom goes a school bus to go to radio . In the afternoon,
takes school

a
he also takes ~~few~~ bus home.
it

**Multiple marks
with clear
intention**

An item is scored as correct even in the presence of multiple marks if the final intention is clear and correct. In the example below, the child gets 3 items correct and none incorrect.

Tom ~~lays~~ goes a school bus to go to ~~radio~~ food . In the afternoon,
takes school

a
he also takes few bus home.
X

Maze Fidelity of Administration

The observer should judge the full test administration and subsequent scoring and calculations. That includes observing setup and directions, timing and scoring the test in parallel with the examiner, checking the examiner's accuracy in procedures using the fidelity checklist in Appendix C, and deciding if the examiner passes or needs more practice for each procedure listed.

Interpreting DIBELS 8th Edition Scores

This chapter covers the interpretation of DIBELS 8th Edition scores. Topics include the scores available for DIBELS 8 and cautions in interpreting results.

DIBELS 8 Test Scores and their Interpretation

DIBELS 8 offers five types of scores: raw scores, equated scaled scores, percentile ranks, growth zones, and composite scores. These scores offer teachers a wealth of information that can be used in planning instruction and monitoring student growth. Each is discussed in turn. For information about DIBELS 8 benchmark goals, please refer to: <https://dibels.uoregon.edu/docs/DIBELS8thEditionGoals.pdf>.

Raw scores. Raw scores are the most basic score available. They generally represent the number of items a student has answered correctly in one minute, with a few exceptions. Maze provides an adjusted raw score where half the number of incorrect items is subtracted from the total number correct. ORF Accuracy is the proportion of words read correctly in one minute and is derived by dividing the number of words read correctly by the total number of words read, including those that were incorrect.

Raw scores have weaknesses in their interpretation. Despite strenuous efforts to create equivalent forms, differences in difficulty between forms still occurs. While these “form effects” are generally quite mild for many subtests as a result of the constrained item pool (e.g., LNF), they become more apparent in subtests involving connected text (i.e., ORF and Maze). Where form effects are more pronounced, differences in scores over time can be obscured or exaggerated. For example, a student who scores 100 words-correct-per-minute (WCPM) in the beginning of year and 120 WCPM in the middle of year has indeed read the middle of year passage at a faster rate, but whether the difference in 20 WCPM is due to actual growth or the middle of year passage simply being easier to read remains unclear. Because form effects can make interpreting student progress difficult, DIBELS 8 offers several alternative score types, especially equated scaled scores (ESS) for subtests where form effects are most obvious.

Risk classification. Although raw scores are not ideal for tracking growth, they can be utilized for screening purposes. Specifically, we created cut-scores for determining students’ risk using raw scores. To support this use, we have provided three types of cut-scores for classifying students.

The first score, called the risk cut-score, can be used to classify students who are well below benchmark in their performance and *at risk* for reading difficulties, including dyslexia.

The second score, called the benchmark goal, can be used to classify students who are performing at benchmark levels and are at *minimal risk* and on track for meeting grade-level proficiency goals from those who are below benchmark performance levels and thus at some risk for not meeting proficiency goals.

On average, this cut-score identifies 80% or more of students performing below the 40th percentile rank on an external measure of reading ability at the end of the year. Students falling above this cut score are typically in need of core support alone, meaning the general curriculum should serve these children well. Students falling between the risk and benchmark cut-scores are at *some risk* for not meeting proficiency goals compared those who are on track for meeting proficiency goals. These students are in need of *strategic support*.

Finally, we have introduced a third cut-score, which represents an ambitious goal for students, and can be used to classify students who are performing well above benchmark and are at *negligible* risk for not meeting proficiency goals. The ambitious cut-score is designed to identify the students who are least at risk in reading. On average, this cut-score identifies 90% or more of students performing below the 40th percentile rank on an external measure of reading ability at the end of the year. Students who fall at or above the ambitious cut-score have a greater chance of performing above the 40th percentile rank on an external measure of reading ability at the end of the year than do students who fall between the benchmark and ambitious cut-scores. Students falling above this cut-score are very likely in need of *core support* alone, meaning the general curriculum should serve these children well. Students performing well above benchmark may benefit from instruction on more advanced skills.

Equated scaled scores. Equated scaled scores (ESS) account for average differences in the difficulty of forms. By having students take multiple forms at the same time, DIBELS researchers are able to quantify how much the forms differ in difficulty. The results of this analysis enable DIBELS to put different forms onto the same scale. As part of the research design, DIBELS researchers also had students take a single “linking” form at each benchmark period. The incorporation of a linking form means that ESS for equated DIBELS subtests capture growth over time, in addition to removing forms effects.

Presently, two DIBELS subtests have equated benchmark forms: ORF and Maze. Specifically, the ORF WCPM and Maze adjusted raw scores have been equated across benchmark periods. By using ESS, teachers can have confidence that differences between scores from two different benchmarks represent real differences in performance.

ORF ESS are scaled so that 400 is the mean ESS across time for a given grade and the standard deviation is 40. As a result, students scoring 400 can be interpreted as reading at

the average rate for their grade level. Also, increases in ESS between benchmark periods represent real change and can be interpreted relative to the standard deviation of 40.

Maze ESS are scaled so that 100 is the mean ESS across time for a given grade and 10 is the standard deviation. Thus, students scoring 100 can be interpreted as reading with average comprehension for their grade level. Likewise, increases in ESS over time represent real change and can be interpreted relative to the Maze ESS standard deviation of 10.

Percentile ranks. Percentile ranks (also known as percentiles) are a way of expressing student performance relative to the norming sample for DIBELS 8. Percentiles look like percentages and represent the percentage of the norming sample that a given student scores at or above on a given subtest. For example, a student who is at the 60th percentile scored the same as or higher than 60% of the norming sample. Because DIBELS researchers made strong efforts to recruit a nationally representative sample when norming DIBELS 8, percentile ranks have strong generalizability. Zones of Growth. DIBELS 8th Edition also offers scores that can be used to interpret growth relative to the norming sample by defining percentile gains, which are normative data regarding changes in performance over time. Percentile gains facilitate comparisons of an individual student's performance over time relative to the performance over time of other students with a similar starting score. These comparisons provide a more nuanced understanding of student progress than cut-scores or percentile ranks. They are an especially useful tool for evaluating the progress of students who perform below the benchmark level and whose performance over time needs to be monitored more closely.

For DIBELS 8, percentile gains are expressed in terms of Zones of Growth. Each student has a starting Zone that is determined by their beginning-of-year percentile rank. For each starting Zone, five bands of percentile gains, or Zones of Growth, are defined. The five initial Zones and Zones of Growth are defined as:

1. Below the 20th percentile,
2. At or above the 20th percentile but at or below the 40th percentile,
3. At or above the 40th percentile but at or below the 60th percentile,
4. At or above the 60th percentile but at or below the 80th percentile, and
5. At or above the 80th percentile.

Composite scores. DIBELS 8th Edition also provides composite scores as a means of interpreting and reporting student performance across subtests. The approach to creating the composite scores represents a marked improvement over the DIBELS Next approach in

that a confirmatory factor analysis (CFA) was used to determine the optimal weighting of DIBELS subtest raw scores while simultaneously accounting for relations among subtests. Our primary concern was correcting for the fact that NWF and ORF each contribute two scores to the composite. These analyses are described in greater detail in the Validity of DIBELS 8th Edition chapter of this manual. The final CFA models for kindergarten through third grade utilized all available subtests and accounted for the covariance between NWF scores. The final CFA models for fourth through eighth grade utilized all available subtests without accounting for covariances. All solutions were scaled so that 400 represents the mean across time for a given grade with 40 as the standard deviation. Thus, similar to ORF, students scoring 400 can be interpreted as an average reader for their grade level. In addition, increases in composite between benchmark periods can be interpreted relative to the standard deviation of 40.

Cautions in Interpreting DIBELS 8 Scores

Even though DIBELS 8th Edition has undergone rigorous research and development procedures, no test is ever 100% reliable and accurate. Moreover, no single test should drive high-stakes decisions made about individual students. DIBELS 8 is not a diagnostic measure in the sense that it cannot diagnose the root causes of reading problems, although using all the subtests provided within a grade can lead to strong hypotheses. Nonetheless, hypotheses regarding the origins and diagnosis of reading problems should be interpreted with caution and tested through the use of other measures and observations. Beyond this general caution, which applies to any single test, there is one additional caution worthy of mention, namely inter-rater reliability.

The reliability statistics reported in this manual were obtained after teachers were well trained in the administration and scoring of DIBELS 8. Although we obtained excellent inter-rater reliability during the course of DIBELS 8 research, we do not report it in this manual. Inter-rater reliability obtained in a study has no bearing on the use of a measure in schools other than the fact that it suggests high inter-rater reliability is possible to achieve. In other words, the reliability of different raters cannot be assumed and should be established in the specific context in which DIBELS 8 is to be used. In addition to initially training test administrators and assessing inter-rater agreement, DIBELS 8 users should recalibrate (i.e., assess inter-rater agreement after a certain period and retrain as needed) at least once a year.

DIBELS 8th Edition Normative Information

This chapter describes normative information regarding DIBELS 8th Edition, including sample recruitment and selection procedures used in DIBELS 8 research studies, and demographic characteristics of the research sample.

Sample Recruitment and Selection Procedures

We recruited elementary and middle schools from across the US to participate in DIBELS 8 research. Schools were recruited from the pool of DIBELS Data System users, through website postings and email contact, and via connections to colleagues of DIBELS 8 researchers, both within the University of Oregon and across the nation. We communicated information about the project, including participation requirements and incentives, to potential participating schools via a flyer, email, or by phone. Regardless of format, schools received a description of the study and participation options. We asked interested schools for contact and other basic information by one of two methods: using a revised Qualtrics survey (see Attachment 3-4), via a link in the email, on the website, or in the pop-up notice; or over the phone. We then sent an email confirmation of enrollment to the designated contact person.

We recruited schools until we met or exceed our recruitment goals, as determined by a minimum of 200 students per grade in each administration of the Iowa Assessments, and 100 struggling students per grade, or until it was no longer feasible for schools to assess students during the specified benchmark administration windows. Due to differences in school grade level configurations across the U.S., we expected to, and did exceed this amount for some grade levels in order to meet the goal for other grades. For example, for many states in the South, schools run K-4 and 5-8, while in the West and Northeast they run K-5 and 6-8 or K-6 and 7-9. To achieve our minimum for transitional grade levels, we ended up yielding larger sample sizes in the other grade levels.

All children enrolled in participating schools, including students with disabilities and students who are English language learners were included, as long as they received English language reading instruction in general education classrooms and had the response capabilities to participate. At their discretion, schools could opt to not assess students with disabilities who require assessment modifications.

Description of the DIBELS 8 Research Sample

The DIBELS 8th Edition research sample consisted of 4,453 students in grades K – 8 from 29 schools (see Table 1 for distribution of students across grades). Schools ranged in

size from very small ($n = 7$) to large ($n = 666$) and were located throughout the country. All four census regions and six of nine census divisions were represented (U.S. Department of Commerce, 1994). See Table 2 for breakdown of students and schools by census region and division.

Using the National Center for Education Statistics definitions, half of the sample lived in rural areas (22% fringe, 22% distant, and 6% remote); approximately a quarter lived in towns (13% fringe and 9% remote); and one fifth of the sample lived in large, suburban areas. The remainder of the sample lived in cities (3% large and 4% small) (Gevertdt, 2015).

Demographic characteristics of the sample are displayed in Table 4: 48.1% of the sample was female, 50.9% was male. 18.1% of students were Hispanic. The sample included 0.6% Asian students, 14.3% Black/African American students, 0.4% Native Hawaiian or other Pacific Islander, 3.9% American Indian or Alaskan Native, 64.3% white, 3.2% two or more races, and 13.3% unknown or not reported. 6.3% of students were English Learners, and 13.9% were eligible for Special Education services. 57.4% of the students were eligible for the free or reduced lunch program.

Table 1. Number of Students by Grade

Grade	<i>n</i>
K	685
1	782
2	725
3	783
4	408
5	388
6	276
7	216
8	190

Table 2. Number of Students and Schools by Census Region and Division

Region Division	Students <i>n</i>	Schools <i>n</i>
West		
Pacific	1771	15
Mountain	910	6
Midwest		
West North Central	156	1
East North Central	609	4
Northeast		
Middle Atlantic	846	2
New England	0	0
South		
West South Central	0	0
East South Central	0	0
South Atlantic	161	0

Table 3. Number of Students Schools by NCES Locale Classification

Locale Classification	Students <i>n</i>	Schools <i>n</i>
City		
Large	114	1
Midsized	0	0
Small	196	2
Suburb		
Large	913	4
Midsized	0	0
Small	0	0
Town		
Fringe	572	2
Distant	0	0
Remote	397	2
Rural		
Fringe	988	5
Distant	1002	7
Remote	271	2

Table 4. Demographic Characteristics of Sample

	<i>n</i>	Percent
Gender		
Female	2142	48.1
Male	2265	50.9
Unknown/not reported	46	1.0
Ethnicity		
Hispanic	804	18.1
Non-Hispanic	3649	81.9
Race		
Asian	26	0.6
Black/African American	636	14.3
Native Hawaiian/Pacific Islander	16	0.4
American Indian/Alaskan Native	173	3.9
White	2863	64.3
Two or more races	144	3.2
Unknown/not reported	594	13.3
Free/reduced lunch program eligible		
Yes	2557	57.4
No	3979	89.4
Unknown/not reported	195	4.4
Special Education Eligibility		
Yes	619	13.9

Table 4. Demographic Characteristics of Sample

No	3754	84.3
Unknown/not reported	81	1.8

Reliability of DIBELS 8th Edition

Reliability refers to the extent to which a test score is a consistent and stable measure. Reliability is a necessary, but insufficient component of validity. Three forms of test reliability have been examined for DIBELS 8th Edition: concurrent alternate form reliability, delayed alternate form reliability, and test-retest reliability. All three forms can be thought of as estimates of the stability of scores, but both delayed alternate form and test-retest reliability address the stability of scores over time. We have provided individual coefficients in tables by subtest for every correlation calculated. We also provided median coefficients by subtest per grade in most instances, and median coefficients across grades by subtest are always provided. In cases where the number of coefficients contributing to a median was even, the lower of the two coefficients is reported, thereby providing a conservative estimate of reliability. First, concurrent and delayed alternate form reliability evidence is provided, followed by test-retest reliability evidence. We then report standard errors of measurement before providing a summary of reliability evidence.

Alternate Form Reliability

Alternate form reliability describes the relationship between scores produced with different versions of a test. In general, strong correlations are desirable because they imply that different versions of the test are capable of generating similar scores. To obtain excellent alternate form reliability, we used strict item writing and form generation guidelines. Nonetheless, reliability must be tested empirically to establish validity of a measure for almost any purpose.

To calculate alternate form reliability, different versions of each DIBELS subtest were administered at the beginning, middle, and end of each year in kindergarten through eighth grade. Because concurrent alternate form reliability measures are administered in the same sitting, the resulting correlations are expected to be quite strong. In contrast, delayed alternate form reliability measures are administered over longer periods of time and are expected to be more moderate in strength. These expectations are especially the case for measures like DIBELS, which targets precisely the skills that are the subject of instruction. Thus, if students are learning as intended, scores from delayed administrations should be less stable than those from concurrent administrations.

Concurrent alternate form reliability. We studied all DIBELS 8th Edition subtests except LNF for concurrent alternate form reliability. As illustrated in Table 5.1, results of concurrent reliability studies over two years reveal very strong correlations in all but one case, and the latter case still reveals moderately strong reliability. The median reliability of PSF in

kindergarten was .86 and in first grade was .81. The overall median reliability of PSF was also .81.

NWF provides two scores and we examined each for concurrent alternate form reliability. As reported in Tables 5.2 and 5.3, the results of all studies for both types of scores are highly reliable. For correct letter sounds on NWF, the median reliability was .89 or above in all grades, with an overall median reliability of .91. For words recoded correctly on NWF, the median reliability was .86 or above in all grades, and overall median reliability was .90.

Results for WRF are displayed in Table 5.4. Once again, concurrent alternate form reliability is very strong. The median reliability for WRF in all grades was greater than .94. Overall median reliability was .95.

Results for ORF and ORF-ACC are displayed in Tables 5.5 and 5.6. As with NWF scores and WRF, concurrent alternate form reliability is very strong. The median reliability for ORF was .92 or above in all grades, and overall median reliability was .93. For accuracy on ORF, the median reliability ranged from .73 to .92 across grades, and overall median reliability was .79.

Maze concurrent alternate form reliability was somewhat weaker than other DIBELS subtests (see Table 5.7). Correlations ranged from .66 to .93. Medians were not calculated by grade due to the presence of a maximum of two studies per grade, but the overall median for Maze was .71. The lower reliability of this subtest suggests it should not be used alone in making high-stakes decisions about students; however, good educational practice is never to rely on a single test for such decisions.

Table 5.1 Concurrent Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	2017–18	1:PM9	93	.80	.72–.87
		2:PM8	153	.88	.83–.91
	2018–19	3:PM5	119	.80	.72–.85
		3:PM1	105	.86	.80–.90
	Median			.86	.80–.90

Table 5.1 Concurrent Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
First	2018–19	1:PM5	128	.81	.74–.86
		1:PM6	148	.53	.40–.64
		2:PM6	149	.81	.74–.86
	Median			.81	.74–.86
Median				.81	.74–.86

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.2 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:PM2	92	.95	.93-.97
		2:PM2	109	.92	.89-.95
		3:PM2	113	.93	.89-.95
		3:PM7	51	.97	.94-.98
		3:PM8	40	.90	.82-.95
	2018-19	1:PM10	63	.92	.88-.95
	Median			.92	.89-.95
First	2017-18	1:PM1	123	.95	.93-.96
		2:PM1	126	.89	.84-.92
		3:PM1	127	.85	.79-.89
	Median			.89	.84-.92

Table 5.2 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
Second	2017-18	2:PM1	50	.94	.89-.97
		3:PM1	52	.93	.87-.96
	2018-19	1:PM9	73	.91	.87-.95
		1:PM10	227	.86	.82-.89
		1:PM12	138	.92	.89-.94
		2:PM3	135	.88	.84-.91
		2:PM8	69	.94	.90-.96
		2:PM13	255	.90	.88-.92
		2:PM14	81	.91	.86-.94
	Median			.91	.87-.95
Third	2017-18	1:2	45	.75	.59-.86
		1:3	47	.92	.86-.95
	2018-19	1:PM10	161	.87	.83-.90
		1:PM13	66	.87	.79-.92
		1:PM14	205	.92	.89-.94
		2:PM4	164	.90	.87-.93
		2:PM6	69	.94	.91-.96
		2:PM9	225	.90	.87-.92
		2:PM15	68	.90	.85-.94
	Median			.90	.87-.92

Table 5.2 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
Median				.91	.87-.95

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.3 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:PM2	87	.95	.92-.96
		2:PM2	107	.86	.80-.90
		3:PM2	112	.89	.84-.92
		3:PM27	51	.91	.84-.95
		3:PM28	40	.84	.72-.91
	2018-19	1:PM10	49	.83	.72-.90
	Median			.86	.80-.90
First	2017-18	1:PM1	123	.90	.86-.93
		2:PM1	126	.90	.85-.93
		3:PM1	127	.86	.81-.90
	Median			.90	.85-.93

Table 5.3 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
Second	2017-18	2:PM1	50	.97	.95–.98
		3:PM1	52	.93	.88–.96
	2018-19	1:PM9	73	.90	.85–.94
		1:PM10	225	.91	.89–.93
	1:PM12	137	.92	.89–.94	
	2:PM3	65	.88	.81–.92	
	Median		.91	.89–.93	
Third	2017-18	2:1	45	.82	.70–.90
		3:1	47	.95	.90–.97
	2018-19	1:PM10	161	.88	.84–.91
		1:PM13	65	.87	.80–.92
	1:PM14	192	.90	.87–.93	
	2:PM4	74	.90	.85–.94	
	2:PM9	75	.86	.79–.91	
	Median		.88	.84–.91	
Median			.90	.85–.94	

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.4 Concurrent Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	3:PM4	87	.97	.96–.98
	2018-19	1:PM6	127	.96	.95–.97
		2: PM7	50	.96	.94–.98
		2: PM11	179	.87	.83–.90
	Median			.96	.95–.97
First	2017-18	2:PM3	59	.97	.96–.98
		3:PM3	59	.96	.93–.97
	2018-19	1: PM3	201	.96	.95–.97
		1: PM18	103	.95	.92–.96
		2: PM3	236	.97	.97–.98
		2: PM19	69	.97	.95–.98
	Median			.96	.95–.97
Second	2017-18	2:PM7	68	.95	.92–.97
		3:PM4	118	.92	.88–.94
	2018-19	1: PM7	139	.94	.92–.96
		1: PM13	71	.96	.94–.98
		1: PM15	230	.96	.94–.97
		2: PM12	69	.94	.91–.96
		2: PM16	82	.95	.92–.97
		2: PM17	136	.95	.93–.97
		2: PM20	241	.94	.93–.96
	Median			.95	.92–.97

Table 5.4 Concurrent Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
Third	2017-18	1:PM9	48	.93	.87–.96
		2:PM9	146	.96	.94–.97
		3:PM9	50	.95	.91–.97
	2018-19	1:PM9	367	.94	.92–.95
		1:PM19	66	.94	.91–.97
		2:PM12	216	.95	.94–.96
		2:PM14	68	.90	.85–.94
		2:PM17	163	.94	.92–.96
		2:PM20	71	.94	.91–.96
	Median			.94	.92–.96
Median				.95	.92–.97

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
First	2017-18	1:PM5	128	.97	.96–.98
		2:PM7	184	.96	.95–.97
		3:PM8	186	.94	.92–.95
	Median			.96	.95–.97

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Second	2017-18	1: 3	118	.95	.92–.96
		1:PM5	109	.97	.96–.98
		1:PM7	108	.97	.95–.98
		2:PM2	159	.96	.94–.97
		2:PM6	159	.96	.95–.97
		3:PM3	118	.95	.92–.96
	2018-19	1:PM16	365	.95	.93–.96
		2:PM15	375	.96	.96–.97
	Median			.96	.94–.97
Third	2017-18	1:PM3	114	.93	.90–.95
		2:PM1	196	.95	.94–.96
		2:PM2	196	.93	.91–.94
		3:PM5	180	.91	.88–.93
		3:PM12	180	.89	.86–.92
	2018-19	1:PM14	366	.94	.92–.95
		2:PM10	377	.94	.93–.95
	Median			.93	.91–.94

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Fourth	2017-18	1:PM2	146	.93	.91–.95
		1:PM3	146	.94	.91–.95
		1:PM5	147	.94	.92–.96
		2:PM2	145	.81	.74–.86
		2:PM7	144	.87	.83–.91
		2:PM10	145	.85	.80–.89
		3:PM2	144	.92	.89–.94
		3:PM8	143	.88	.84–.91
	2018-19	1:PM12	440	.94	.93–.95
		1:PM20	339	.94	.93–.95
		2:PM18	356	.94	.93–.95
		2:PM20	506	.94	.93–.95
	Median			.93	.91–.95

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Fifth	2017-18	1:2	123	.95	.92–.96
		1:3	131	.91	.87–.93
		1:PM5	133	.92	.89–.94
		1:PM1	132	.92	.89–.94
		1:PM3	133	.92	.89–.94
		2:PM2	181	.93	.91–.95
		2:PM20	58	.94	.89–.96
		3:PM7	131	.93	.90–.95
		3:PM8	131	.92	.89–.95
	2018-19	1:PM14	327	.93	.91–.94
		1:PM15	387	.91	.89–.92
		2:PM16	393	.89	.87–.91
		2:PM19	510	.93	.92–.94
	Median			.92	.89–.94

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	69	.94	.91–.96
		1:PM19	69	.95	.93–.97
		2:3	98	.93	.90–.95
		2:PM1	99	.89	.84–.93
		2:PM2	104	.93	.90–.96
		2:PM3	99	.94	.91–.96
		2:PM4	104	.93	.89–.95
		2:PM11	104	.93	.90–.95
		2:PM16	98	.95	.92–.96
		3:PM15	98	.94	.91–.96
	2018-19	1:PM13	182	.92	.90–.94
		1:PM16	166	.92	.89–.94
		2:PM7	290	.94	.92–.95
		2:PM14	164	.91	.88–.93
	Median			.93	.90–.95

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Seventh	2017-18	1:2	136	.92	.89–.94
		1:3	123	.90	.86–.93
		1:PM8	83	.93	.89–.95
		1:PM9	83	.87	.80–.91
		1:PM11	83	.93	.89–.95
		2:PM1	142	.92	.88–.94
	2018-19	2:PM2	141	.91	.88–.94
		3:PM3	123	.89	.84–.92
		1:PM13	63	.91	.86–.95
		1:PM19	77	.95	.92–.97
		2:PM12	199	.95	.93–.96
		2:PM18	63	.93	.89–.95
	Median			.92	.88–.94

Table 5.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Eighth	2017-18	1:2	110	.92	.89–.95
		1:3	102	.81	.73–.87
		1:PM7	69	.92	.87–.95
		1:PM11	70	.91	.86–.94
		1:PM12	70	.92	.86–.95
		2:PM2	114	.90	.86–.93
		2:PM4	114	.92	.89–.94
		3:PM1	102	.80	.72–.86
		3:PM3	102	.78	.69–.85
	2018-19	1:PM13	66	.95	.92–.97
		1:PM16	74	.95	.92–.97
		2:PM6	134	.92	.89–.94
		2:PM10	61	.94	.91–.97
	Median			.92	.87–.95
Median				.93	.90–.95

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
First	2017-18	1:PM5	128	.92	.89–.94
		2:PM7	184	.90	.87–.93
		2:PM1	75	.92	.86–.95
		3:PM1	44	.93	.87–.96
		3:PM8	186	.91	.88–.93
	Median			.92	.87–.94
Second	2017-18	1:3	118	.77	.69–.84
		1:PM5	109	.91	.87–.94
		1:PM7	108	.89	.85–.93
		2:3	118	.83	.77–.88
		3:PM3	118	.78	.70–.84
	2018-19	1:PM16	363	.85	.82–.88
		2:PM15	66	.96	.95–.98
	Median			.85	.77–.88

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Third	2017-18	1:2	114	.79	.71–.85
		1:PM3	114	.80	.90–.95
		2:PM1	196	.96	.94–.97
		2:PM2	196	.96	.95–.97
		3:PM5	180	.67	.58–.74
		3:PM12	180	.72	.64–.79
	2018-19	1:PM14	366	.79	.75–.82
		2:PM10	148	.81	.74–.86
	Median			.79	.74–.85
Fourth	2017-18	1:PM2	146	.74	.65–.80
		1:PM3	146	.75	.67–.81
		1:PM5	147	.78	.71–.84
		2:PM2	145	.76	.68–.82
		2:PM7	144	.60	.49–.70
		2:PM10	145	.61	.50–.71
		3:PM2	144	.86	.81–.90
		3:PM8	143	.67	.57–.75
	2018-19	1:PM12	440	.89	.87–.91
		2:PM18	203	.88	.84–.91
		2:PM20	247	.93	.89–.93
	Median			.76	.68–.82

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Fifth	2017-18	1:2	123	.97	.95–.98
		1:3	131	.57	.45–.68
		1:PM1	132	.76	.68–.83
		1:PM3	133	.68	.58–.76
		1:PM5	133	.73	.64–.80
		2:PM2	181	.96	.95–.97
		2:PM20	58	.94	.91–.97
		3:PM7	131	.54	.41–.65
		3:PM8	131	.65	.53–.74
	2018-19	1:PM14	327	.75	.70–.80
		1:PM15	387	.85	.82–.87
		2:PM16	70	.57	.38–.71
		2:PM19	82	.69	.56–.79
	Median			.73	.64–.80

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	69	.77	.65–.85
		1:PM19	69	.84	.75–.90
		2:3	98	.78	.69–.85
		2:PM1	99	.68	.56–.78
		2:PM2	104	.90	.87–.94
		2:PM3	99	.64	.50–.74
		2:PM4	104	.83	.76–.88
		2:PM11	104	.77	.68–.84
		2:PM16	98	.76	.67–.84
		3:PM15	98	.81	.73–.87
	2018-19	1:PM13	182	.97	.96–.98
		1:PM16	166	.96	.94–.97
		2:PM7	89	.66	.52–.76
	Median			.78	.69–.85

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Seventh	2017-18	1:2	136	.89	.85–.92
		1:3	123	.93	.90–.95
		1:PM8	83	.89	.83–.93
		1:PM9	83	.90	.85–.94
		1:PM11	83	.89	.83–.93
		2:PM1	142	.87	.82–.90
		2:PM2	141	.83	.78–.88
		3:PM3	123	.87	.81–.91
	2018-19	1:PM13	63	.68	.51–.79
		1:PM19	77	.98	.97–.99
		2:PM12	92	.93	.90–.96
	Median			.89	.83–.93

Table 5.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Eighth	2017-18	1:2	110	.85	.79–.89
		1:3	102	.82	.75–.88
		1:PM7	69	.84	.75–.90
		1:PM11	70	.86	.78–.91
		1:PM12	70	.73	.60–.83
		2:PM2	114	.84	.76–.89
		2:PM4	114	.89	.85–.92
		3:PM1	102	.76	.66–.83
		3:PM3	102	.85	.79–.90
	2018-19	1:PM13	66	.90	.85–.94
		1:PM16	74	.88	.82–.92
		2:PM6	90	.81	.72–.87
	Median			.84	.76–.89
Median				.79	.74–.85

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

Table 5.7 Concurrent Alternate Form Reliability Coefficients for DIBELS 8th Edition MAZE Adjusted Score

Grade	Forms	N	r	CI
Second	1:2	66	.82	.72-.89
	1:3	80	.80	.70-.87
Third	1:2	64	.71	.57-.82
	1:3	64	.66	.49-.77
Fourth	1:2	74	.71	.58-.81
	1:3	71	.71	.58-.81
Fifth	1:2	63	.82	.71-.89
	1:3	47	.71	.53-.83
Sixth	1:2	72	.83	.73-.89
Seventh	1:2	69	.93	.89-.96
Median			.71	.58-.81

Note. Form numbers correspond to benchmark period forms. All coefficients come from the 2018-2019 sample, and data was gathered within a six-week period in the fall of 2018.

Delayed alternate form reliability. We studied all DIBELS 8th Edition subtests except Maze for delayed alternate form reliability. Delayed alternate form reliability was established by correlating benchmark forms; thus, the delay was about three months. As a result, these results were expected to be weaker than concurrent alternate form reliability because three months of instruction intervened between administration of alternate forms in the calculation of delayed alternate form correlations.

Median delayed alternate form reliability for LNF was .83 in kindergarten, .76 in first grade, and .81 overall (see Table 5.8). Median delayed alternate form reliability was lower for PSF (see Table 5.9). PSF median correlations were .49 in kindergarten, .54 in first grade, and .54 across the two grades. For the NWF correct letter sounds score, delayed alternate form reliability was .76 across grades and was .71 in kindergarten, .79 in first grade, .83 in second grade, and .74 in third grade (see Table 5.10). For the NWF words recoded correctly score, delayed alternate form reliability was .72 across grades and was .60 in kindergarten, .69 in first grade,

.81 in second grade, and .78 in third grade (see Table 5.11). WRF coefficients are located in Table 5.12 and demonstrated strong delayed alternate form reliability of .90 overall and .85 in kindergarten, .88 in first grade, .91 in second grade, and .90 in third grade (see Table 5.12). ORF coefficients (see Table 5.13) demonstrated strong delayed alternate form reliability of .89 overall and .92 in first grade, .89 in second grade, .88 in third grade, .82 in fourth grade, .86 in fifth grade, .91 in sixth grade, .86 in seventh grade, and .74 in eighth grade. For the ORF words accuracy score, delayed alternate form reliability was .78 in first grade, .78 in second grade, .75 in third grade, .82 in fourth grade, .86 in fifth grade, .77 in sixth grade, .90 in seventh grade, and .73 in eighth grade. Across grades, median delayed alternate form reliability for ORF accuracy was .78 (see Table 5.14).

Table 5.8 Delayed Alternate Form Reliability for DIBELS 8th Edition Letter Naming Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	2017–18	1:2	150	.86	.82–.90
		1:3	100	.83	.76–.88
		2:3	137	.87	.81–.90
	2018–19	1:2	419	.81	.77–.84
	Median			.83	.76–.88
First	2017–18	1:2	153	.77	.70–.83
		1:3	140	.70	.61–.78
		2:3	183	.76	.70–.82
	2018–19	1:2	413	.83	.80–.86
	Median			.76	.70–.82
Median				.81	.77–.84

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated.

Table 5.9 Delayed Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	2017–18	1:2	94	.49	.32–.63
		1:3	72	.39	.18–.57
		2:3	115	.70	.59–.78
	2018–19	1:2	347	.56	.48–.63
	Median			.49	.32–.63
First	2017–18	1:2	148	.53	.40–.64
		1:3	138	.54	.41–.65
		2:3	181	.63	.53–.71
	2018–19	1:2	411	.64	.58–.69
	Median			.54	.41–.65
Median				.54	.41–.65

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Table 5.10 Delayed Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:2	191	.73	.65-.79
		1:3	180	.63	.53-.71
		2:3	224	.83	.78-.87
	2018-19	1:2	192	.71	.63-.78
	Median			.71	.63-.78
First	2017-18	1:2	268	.80	.75-.84
		1:3	256	.73	.67-.78
		2:3	308	.79	.75-.83
	2018-19	1:2	404	.82	.79-.85
	Median			.79	.75-.83
Second	2017-18	1:2	127	.83	.77-.88
		1:3	113	.84	.77-.88
		2:3	194	.84	.79-.87
	2018-19	1:2	420	.74	.70-.78
	Median			.83	.77-.88
Third	2017-18	2:3	218	.76	.70-.81
	2018-19	1:2	412	.74	.69-.78
	Median			.74	.69-.78
Median				.76	.70-.81

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Table 5.11 Delayed Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency- Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:2	171	.70	.62–.77
		1:3	161	.60	.49–.69
		2:3	219	.77	.71–.82
	2018-19	1:2	77	.45	.25–.61
	Median			.60	.49–.69
First	2017-18	1:2	268	.72	.66–.77
		1:3	256	.67	.59–.73
		2:3	308	.74	.68–.78
	2018-19	1:2	233	.69	.62–.75
	Median			.69	.62–.75
Second	2017-18	1:2	126	.83	.76–.87
		1:3	112	.81	.74–.87
		2:3	194	.83	.78–.87
	2018-19	1:2	105	.72	.61–.80
	Median			.81	.74–.87
Third	2017-18	2:3	218	.78	.72–.83
	2018-19	1:2	173	.83	.78–.87
	Median			.78	.72–.83
Median				.72	.66–.77

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Table 5.12 Delayed Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:2	112	.88	.83–.92
		1:3	86	.85	.77–.90
		2:3	150	.90	.86–.92
	2018-19	1:2	357	.81	.77–.84
	Median			.85	.77–.90
First	2017-18	1:2	153	.91	.88–.93
		1:3	141	.82	.75–.86
		2:3	241	.88	.85–.91
	2018-19	1:2	410	.92	.91–.93
	Median			.88	.85–.91
Second	2017-18	1:2	163	.91	.88–.94
	2018-19	1:2	407	.92	.88–.95
	Median			.91	.88–.94
	2017-18	1:2	198	.90	.88–.93
		1:3	138	.90	.86–.93
Third		2:3	216	.90	.87–.92
	2018-19	1:2	405	.89	.87–.91
	Median			.90	.88–.93
Median				.90	.88–.93

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Table 5.13 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Sample	Forms	N	r	CI
First	2017-18	1:2	161	.94	.91–.95
		1:3	117	.87	.82–.91
		2:3	218	.92	.89–.94
	2018-19	1:2	399	.94	.93–.95
	Median			.92	.89–.94
Second	2017-18	1:2	225	.89	.86–.92
		1:3	116	.85	.79–.89
		2:3	116	.92	.89–.95
	2018-19	1:2	401	.92	.90–.93
	Median			.89	.86–.92
Third	2017-18	1:2	112	.91	.87–.94
		2:3	171	.84	.79–.88
	2018-19	1:2	403	.88	.86–.90
	Median			.88	.86–.90
Fourth	2017-18	1:2	142	.82	.76–.87
	2018-19	1:2	391	.90	.88–.92
	Median			.82	.76–.87
Fifth	2017-18	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	2018-19	1:2	378	.81	.77–.84
	Median			.86	.81–.90

Table 5.13 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency- Words Read Correctly

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	66	.86	.78–.91
		1:3	61	.91	.85–.94
		2:3	98	.91	.87–.94
	2018-19	1:2	173	.89	.86–.92
	Median			.91	.85–.94
Seventh	2017-18	1:2	79	.89	.83–.93
		1:3	65	.84	.74–.90
		2:3	127	.86	.81–.90
	2018-19	1:2	75	.93	.89–.95
	Median			.86	.81–.90
Eighth	2017-18	1:2	67	.92	.89–.95
		1:3	67	.73	.62–.81
		2:3	100	.74	.64–.82
	2018-19	1:2	68	.94	.90–.96
	Median			.74	.64–.82
Median				.89	.83–.92

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Table 5.14 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
First	2017-18	1:2	161	.89	.85–.92
		1:3	117	.71	.61–.79
		2:3	218	.78	.72–.82
	2018-19	1:2	227	.84	.80–.88
	Median			.78	.72–.82
Second	2017-18	1:2	225	.78	.73–.83
		1:3	116	.61	.48–.71
		2:3	116	.80	.72–.86
	2018-19	1:2	100	.83	.76–.88
	Median			.78	.72–.83
Third	2017-18	1:2	112	.75	.66–.82
		2:3	171	.65	.55–.73
	2018-19	1:2	176	.81	.76–.86
	Median			.75	.66–.82
Fourth	2017-18	1:2	142	.82	.76–.87
	2018-19	1:2	240	.82	.77–.86
	Median			.82	.76–.86
Fifth	2017-18	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	2018-19	1:2	76	.51	.32–.66
	Median			.86	.81–.90

Table 5.14 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	66	.89	.83–.93
		1:3	61	.77	.64–.86
		2:3	98	.85	.78–.89
	Median			.77	.78–.89
Seventh	2017-18	1:2	79	.90	.85–.93
		1:3	65	.96	.94–.98
		2:3	127	.83	.77–.88
	Median			.90	.85–.93
Eighth	2017-18	1:2	67	.73	.59–.82
		1:3	67	.67	.51–.78
		2:3	100	.75	.65–.83
	Median			.73	.59–.82
Median				.78	.72–.83

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

Test-Retest Reliability Test-retest reliability describes the correlation between scores on the same test administered at different points in time to the same test-takers. There are no universally accepted standards for judging the acceptability of test-retest reliability coefficients. The ideal degree of test-retest reliability depends on the purpose of the test, the construct it assesses, and the time between test administrations. In the case of DIBELS 8th Edition, we would like to emphasize that very high levels of reliability, especially for component skills like letter naming and phonemic segmentation, are undesirable because these skills develop quite rapidly in the grades in which they are assessed (Paris, 2005). More generally, one should not expect levels of test-retest reliability to be as high as alternate form reliability when the skill measured develops rapidly and time between measurement occasions is sufficient for genuine growth to have occurred.

We studied all DIBELS 8th Edition subtests except PSF and Maze for test-retest reliability. Test-retest reliability was evaluated by administering the same form in two different benchmark periods. Thus, as with delayed alternate form reliability, about three months passed between administrations. As a result of the long delay between administrations, test-retest reliability coefficients are again expected to be lower than concurrent alternate form reliability coefficients due to instructional effects.

For LNF, median test-retest reliability was .82 in kindergarten, .75 in first grade, and .77 overall (see Table 5.15). For the NWF correct letter sounds score, median test-retest reliability was .79 in kindergarten, .81 in first grade, .75 in second grade, and .77 overall; no median is reported for third grade because only one coefficient (.87) was available (see Table 5.16). For the NWF words recoded correctly score, median test-retest reliability was .72 in kindergarten, .78 in first grade, and .72 in second grade and overall; no median is reported for third grade because only one coefficient (.84) was available (see Table 5.17). For WRF, median test-retest reliability was strong, at .92 in kindergarten, .90 in first grade, and .92 overall; no median is reported for second or third grade because only two and one coefficients were available respectively, although all were .88 or above (see Table 5.18). For the ORF words read correctly score, median test-retest reliability was also strong, at .91 across grades; median test-retest reliability by grade was .92 for first, .87 for second, .94 for third, .90 for fourth, .91 for fifth, .91 for sixth, .86 for seventh, and .92 for eighth (see Table 5.19). For the ORF accuracy, median test-retest reliability was adequate at .75 across grades; median test-retest reliability by grade was .74 for first, .75 for second, .80 for third, .75 for fourth, .79 for fifth, .74 for sixth, .90 for seventh, and .83 for eighth (see Table 5.20).

Table 5.15 Test–Retest Reliability Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	122	.82	.76–.87
	1:3	123	.77	.69–.83
	2:3	121	.84	.78–.89
	Median		.82	.76–.87
First	1:2	124	.82	.75–.87
	1:3	123	.67	.56–.76
	2:3	128	.75	.66–.81
	Median		.75	.66–.81
Median			.77	.69–.83

Note. All data comes from the 2017–18 sample and was gathered in the benchmark periods indicated.

Table 5.16 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	92	.84	.77 - .89
	1:3	89	.79	.70 - .86
	2:3	107	.77	.67 - .83
	Median		.79	.70 - .86
First	1:2	120	.81	.74 - .87
	1:3	119	.75	.66 - .82
	2:3	126	.84	.78 - .84
	Median		.81	.74 - .87

Table 5.16 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Benchmark period	N	r	CI
Second	1:2	116	.75	.66 - .82
	1:3	115	.75	.65 - .82
	2:3	165	.88	.83 - .91
	Median		.75	.66 - .82
Third	2:3	158	.87	.82 - .90
Median			.77	.67 - .83

Note. The same form was given in the benchmark periods indicated.

Table 5.17 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	89	.79	.69-.86
	1:3	85	.72	.59-.81
	2:3	106	.69	.58-.78
	Median		.72	.59-.81
First	1:2	120	.78	.70-.84
	1:3	119	.68	.57-.77
	2:3	126	.80	.73-.85
	Median		.78	.70-.84
Second	1:2	116	.72	.62-.80
	1:3	115	.72	.62-.80
	2:3	165	.88	.84-.91
	Median		.72	.62-.80

Table 5.17 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Benchmark period	N	r	CI
Third	2:3	158	.84	.78–.88
Median			.72	.62–.80

Note. The same form was given in the benchmark periods indicated.

Table 5.18 Test-Retest Reliability Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	120	.92	.88–.94
	1:3	120	.88	.84–.92
	2:3	120	.93	.91–.95
	Median		.92	.88–.94
First	1:2	126	.90	.86–.93
	1:3	122	.82	.75–.87
	2:3	128	.92	.89–.94
	Median		.90	.86–.93
Second	2:3	82	.95	.93–.97
Third	1:2	48	.93	.88–.96
	2:3	90	.88	.82–.92
Median			.92	.88–.94

Note. The same form was given in the benchmark periods indicated.

Table 5.19 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Benchmark period	N	r	CI
First	1:2	156	.92	.89–.94
	1:3	123	.88	.83–.91
	2:3	164	.94	.92–.95
	Median		.92	.89–.94
Second	1:2	150	.87	.83–.91
	1:3	116	.85	.79–.90
	2:3	148	.93	.90–.95
	Median		.87	.83–.91
Third	1:2	159	.94	.91–.95
	1:3	110	.92	.89–.95
	2:3	156	.94	.91–.95
	Median		.94	.91–.95
Fourth	1:2	274	.91	.89–.93
	1:3	259	.90	.87–.92
	2:3	316	.88	.86–.90
	Median		.90	.87–.92
Fifth	1:2	229	.91	.89–.93
	1:3	221	.87	.83–.90
	2:3	298	.91	.89–.93
	Median		.91	.89–.93

Table 5.19 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Benchmark period	N	r	CI
Sixth	1:2	169	.91	.89–.94
	1:3	158	.91	.87–.93
	2:3	219	.93	.91–.95
	Median		.91	.89–.94
Seventh	1:2	79	.90	.84–.93
	1:3	65	.86	.78–.91
	2:3	121	.86	.81–.90
	Median		.86	.81–.90
Eighth	1:2	67	.91	.86–.94
	1:3	67	.92	.87–.95
	2:3	96	.93	.90–.95
	Median		.92	.87–.95
Median			.91	.89–.93

Note. The same form was given in the benchmark periods indicated.

Table 5.20 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Benchmark period	N	r	CI
First	1:2	156	.80	.74–.85
	1:3	123	.49	.34–.61
	2:3	164	.74	.67–.81
	Median		.74	.67–.81
Second	1:2	150	.75	.67–.81
	1:3	116	.75	.67–.81
	2:3	148	.83	.78–.88
	Median		.75	.67–.81
Third	1:2	159	.86	.81–.90
	1:3	110	.74	.64–.82
	2:3	156	.80	.73–.85
	Median		.80	.73–.85
Fourth	1:2	274	.83	.79–.87
	1:3	259	.75	.70–.80
	2:3	316	.75	.70–.79
	Median		.75	.70–.80
Fifth	1:2	229	.79	.73–.83
	1:3	221	.79	.73–.83
	2:3	298	.83	.80–.87
	Median		.79	.73–.83

Table 5.20 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Benchmark period	N	r	CI
Sixth	1:2	169	.76	.69–.82
	1:3	158	.60	.49–.69
	2:3	219	.74	.67–.79
	Median		.74	.67–.79
Seventh	1:2	79	.90	.84–.93
	1:3	65	.95	.93–.97
	2:3	121	.84	.78–.89
	Median		.90	.84–.93
Eighth	1:2	67	.84	.75–.90
	1:3	67	.79	.68–.86
	2:3	96	.83	.75–.88
	Median		.83	.75–.88
Median			.75	.70–.81

Note. The same form was given in the benchmark periods indicated.

Standard Error of Measurement Finally, we also estimated the standard error of measurement (SEM) using a classical test theory approach, which multiplies the standard deviation for a measure by the square root of one minus the reliability of the measure. The SEM for each measure in each grade and benchmark period is reported in Table 5.21. In all cases except for LNF, we used the median concurrent alternate form reliability for a grade and the standard deviation (SD) for each benchmark period in these calculations. Because concurrent alternate form reliability was not available for LNF, we used delayed alternate form reliability in this case. By definition, measures with the best reliability have the smallest SEMs relative to each measure's SD.

Table 5.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Benchmark period	N	r	CI
Kindergarten	LNF	8.42	9.06	8.56
	PSF	6.01	6.47	6.29
	NWF-CLS	5.88	6.60	7.40
	NWF-WRC	2.59	3.06	3.52
	WRF	1.58	2.31	2.89
First	LNF	8.64	9.22	8.86
	PSF	7.58	6.26	6.76
	NWF-CLS	10.55	11.57	14.33
	NWF-WRC	3.96	4.64	5.28
	WRF	3.75	4.52	5.06
	ORF	6.06	6.88	8.43
	ORF-ACC	0.10	0.08	0.06
Second	NWF-CLS	12.96	14.51	13.59
	NWF-WRC	4.29	5.27	4.60
	WRF	5.10	5.48	5.90
	ORF	7.84	8.85	9.12
	ORF-ACC	0.05	0.06	0.02
	Maze	2.88	TBD	TBD

Table 5.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Benchmark period	N	r	CI
Third	NWF-CLS	11.68	15.72	17.53
	NWF-WRC	3.96	5.62	5.89
	WRF	5.24	5.69	5.58
	ORF	10.39	10.24	9.59
	ORF-ACC	0.05	0.04	0.02
	Maze	3.74	TBD	TBD
Fourth	ORF	9.63	12.86	10.08
	ORF-ACC	0.03	0.01	0.03
	Maze	4.49	TBD	TBD
Fifth	ORF	11.31	11.01	11.23
	ORF-ACC	0.04	0.05	0.02
	Maze	4.46	TBD	TBD
Sixth	ORF	11.00	9.82	12.87
	ORF-ACC	0.02	0.02	0.02
	Maze	3.46	TBD	TBD
Seventh	ORF	10.16	10.06	9.83
	ORF-ACC	0.02	0.02	0.02
	Maze	3.06	TBD	TBD

Table 5.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Benchmark period	N	r	CI
Eighth	ORF	8.08	9.16	10.23
	ORF-ACC	0.02	0.02	0.01
	Maze	TBD	TBD	TBD

Note. SEMs for Maze middle and end of year benchmarks will be available in summer 2019.

Summary

Taken together, the reliability evidence for DIBELS 8 is strong. The strongest evidence regards concurrent alternate form reliability. Research into the reliability of DIBELS scores is ongoing, and regular addendums to this manual will continue to build the validity argument for DIBELS 8.

Validity of DIBELS 8th Edition

Validity is an argument that hinges on the desired inferences to be made about an individual (Messick, 1995). As such, assuming an adequate reliability of scores, different forms of validity can serve as evidence for different claims. In the sections below, we discuss concurrent and predictive validity.

Concurrent validity is generally seen as a means of validating that the intended construct is being captured by a measure. Concurrent validity is evaluated by correlating each DIBELS 8 subtest with like DIBELS Next subtests, with related external criterion measures, and with the other DIBELS 8 subtests.

Predictive validity can also be seen as a means of validating that the intended construct has been captured, but in addition, it serves as a means of validating the use of a measure for predicting performance at a later period (e.g., often the end of a grade). Predictive validity traditionally includes correlations, but when intended uses of measure include identification of subgroups of students, then screening accuracy is the more valuable evidence that a measure is functioning as intended. Both are evaluated for DIBELS 8.

Concurrent Validity

The more similar the two measures given, the higher the correlation between the scores should be. Thus, when correlating like measures (e.g., DIBELS Next and DIBELS 8 NWF), correlations should be quite strong (i.e., .8 or above). However, when correlating component skills like PSF and LNF with reading achievement scores, correlations ought to be lower. As a result, in the sections that follow, relationships are expected to be strongest between the same subtest for the previous and current editions of DIBELS and weakest for subtests like PSF and LNF with reading achievement measures.

Correlations with DIBELS Next. The equivalency of the previous and current editions of DIBELS was evaluated by correlating like subtests for students who took both DIBELS Next and DIBELS 8 in a given benchmark period. Correlations are reported for the same subtest when it would traditionally be administered on DIBELS Next only.

As shown in Table 6.1, DIBELS Next and DIBELS 8th Edition LNF are strongly related. Similarly strong are the relations between DIBELS Next and DIBELS 8th Edition PSF, although these relations are more variable, which may be due to the different approaches to item ordering of the two editions (see Table 6.2). Of particular note are the correlations between NWF scores on the two editions of DIBELS. Despite substantial changes in the items used and in form construction, correlations for NWF-CLS are quite strong at .73 or above (see

Table 6.3). More interesting is that despite the additional difference in scoring for blending on the two editions (i.e., words recoded correctly in DIBELS 8th Edition and whole words read in DIBELS Next), the correlations are again very strong (see Table 6.4). The strongest correlations are in kindergarten and second grade. The lowest is still quite strong at .66 in fall of first grade, and correlations strengthen over the course of the first grade year. These results suggest that DIBELS Next WWR and DIBELS 8th Edition WRC scores rank students similarly despite differences in scoring method and form composition. Finally, both ORF and ORF-ACC scores in DIBELS 8th Edition also demonstrate strong correlations with the corresponding subtest in DIBELS Next, but ORF is particularly good, with nearly every correlation being .90 or above (see Tables 6.5 and 6.6 respectively). Taken together, this evidence suggests the equivalence of DIBELS Next and DIBELS 8th Edition.

Table 6.1 Correlations between DIBELS 8th Edition and DIBELS Next Letter Naming Fluency

Grade	Period	r	N	CI
Kindergarten	1	.75	153	.67–.81
	2	.87	181	.84–.90
	3	.88	128	.83–.91
First	1	.72	157	.63–.79

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.2 Correlations between DIBELS 8th Edition and DIBELS Next Phoneme Segmentation Fluency

Grade	Period	r	N	CI
Kindergarten	2	.83	28	.65–.92
	3	.75	125	.66–.82
First	1	.49	154	.35–.60
	2	.84	16	.60–.94
	3	.96	16	.88–.99

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.3 Correlations between DIBELS 8th Edition and DIBELS Next Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	r	N	CI
Kindergarten	2	.84	163	.78–.88
	3	.87	139	.82–.91
First	1	.73	184	.66–.79
	2	.81	223	.76–.85
	3	.87	211	.84–.90
Second	1	.83	129	.77–.88

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.4 Correlations between DIBELS 8th Edition and DIBELS Next Nonsense Word Fluency-Words Recoded Correctly

Grade	Period	r	N	CI
Kindergarten	2	.88	130	.84–.92
	3	.84	108	.78–.89
First	1	.66	163	.56–.74
	2	.78	202	.72–.83
	3	.90	190	.86–.92
Second	1	.82	118	.75–.87

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.5 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Period	r	N	CI
First	2	.93	227	.91–.95
	3	.94	188	.92–.96
Second	1	.83	173	.77–.87
	2	.95	192	.93–.96
	3	.97	25	.92–.98
Third	1	.92	100	.89–.95
	2	.90	209	.87–.92
	3	.90	176	.87–.93
Fourth	1	.90	35	.82–.95
	2	.90	37	.81–.95
	3	.92	58	.87–.95
Fifth	1	.95	51	.91–.97
	2	.93	84	.89–.95
	3	.98	25	.94–.99
Sixth	2	.93	40	.88–.96

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.6 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	r	N	CI
First	2	.89	196	.86–.89
	3	.84	163	.78–.88
Second	1	.79	131	.71–.84
	2	.90	171	.87–.92
	3	.98	24	.95–.99
Third	1	.89	73	.83–.93
	2	.76	186	.70–.82
	3	.69	154	.60–.77
Fourth	1	.85	29	.70–.93
	2	.89	26	.77–.95
	3	.88	46	.78–.93
Fifth	1	.90	41	.81–.94
	2	.99	23	.97–.99
Sixth	2	.91	30	.82–.96

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Correlations with external criterion measures. The concurrent validity of DIBELS 8 was also evaluated by correlating its subtests with external criterion measures given in the same benchmark period. These measures included DIBELS Next predominant subtests for the grade and period of the assessment (so long as they were not covered in the previous section), DIBELS Next composite scores, various raw scores from the Comprehensive Test of Phonological Processing (CTOPP), and Total Reading and Word Analysis scores from the Iowa Assessment. Correlations with other DIBELS 8 subtests are reported in the following section.

Correlations for each DIBELS 8 subtest are reported in Tables 6.7-6.14 and are only reported when the study sample size was 40 or greater to ensure a minimum threshold of

precision in the correlation estimate. Where available, concurrent validity with the DIBELS Next composite scores was quite good. Correlations among unrelated DIBELS 8 and DIBELS Next subtests varied in strength as would be predicted based on the similarity of the component skills assessed. Correlations with some of the CTOPP are negative, but negative correlations are sensible for RAN measures because CTOPP scores are times while DIBELS scores are rates (i.e., items per minute), and time and rate have a naturally inverse relationship with each other. CTOPP composite scores were not correlated with DIBELS 8 scores because CTOPP composite scores require using age-level norms, which eliminates some variability in scores and obscures the relationship between performance on specific component skills measured by both CTOPP and DIBELS.

Of particular interest are the moderate to strong correlations between CTOPP rapid naming measures and LNF depending on the CTOPP raw score examined. As might be expected, relations are strongest for rapid letter naming ($r = -.56$ and $-.65$ in kindergarten and first grade respectively). Relations are weaker for other naming scores, suggesting that while LNF can operate as a decent screener for processing speed, it cannot replace more diagnostic measures like CTOPP.

Similarly, PSF is moderately to strongly related to CTOPP phonological awareness raw scores. Relations here are more uniform. These results again suggest the validity of DIBELS as a screener for phonological awareness difficulties, but suggest that it cannot replace more diagnostic measures like CTOPP.

Finally, concurrent relations with the Iowa Assessment scores vary predictably by subtest. PSF showed the weakest relationships with the Iowa total reading and word analysis scores in kindergarten and first grade. LNF was most weakly related to Iowa total reading in kindergarten, but otherwise was moderately strongly correlated with Iowa scores. The rest of the DIBELS 8 subtests displayed moderate to strong relations with external criterion measures regardless of grade.

Table 6.7 Concurrent Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next Composite	.70	128	.60–.78
		CTOPP2 Rapid Digit Naming Raw Score	-.52	124	-.64–.38
		CTOPP2 Rapid Letter Naming Raw Score	-.56	64	-.71–.36
		CTOPP2 Rapid Color Naming Raw Score	-.27	160	-.41–.12
		CTOPP2 Rapid Object Naming Raw Score	-.31	170	-.44–.17
	2	DIBELS Next NWF-CLS	.72	181	.64–.78
		DIBELS Next Composite	.80	156	.74–.85
	3	DIBELS Next NWF-CLS	.78	128	.70–.84
		DIBELS Next Composite	.89	98	.84–.93

Table 6.7 Concurrent Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF-CLS	.49	157	.36–.60
		DIBELS Next Composite	.65	144	.54–.74
		CTOPP2 Rapid Digit Naming Raw Score	-.62	174	-.71–.52
		CTOPP2 Rapid Letter Naming Raw Score	-.65	165	-.73–.56
		CTOPP2 Rapid Color Naming Raw Score	-.47	171	-.58–.34
		CTOPP2 Rapid Object Naming Raw Score	-.48	171	-.59–.36
	2	DIBELS Next ORF-WRC	.61	196	.52–.69
		DIBELS Next Composite	.70	163	.61–.77
	3	DIBELS Next ORF-WRC	.64	188	.55–.72
		DIBELS Next Composite	.63	163	.53–.72
		Iowa Total Reading	.54	117	.40–.66
		Iowa Word Analysis	.46	119	.31–.61

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.8. Concurrent Validity Coefficients for DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.19	99	-.01–.37
		DIBELS Next Composite	.47	75	.27–.63
		CTOPP2 Elision Raw Score	.58	189	.48–.67
		CTOPP2 Blending Raw Score	.45	188	.32–.55
		CTOPP2 Sound Matching Raw Score	.53	187	.41–.62
	3	DIBELS Next NWF-CLS	.32	125	.16–.47
		DIBELS Next Composite	.62	95	.48–.73
First	1	DIBELS Next NWF-CLS	.09	154	-.07–.25
		DIBELS Next Composite	.27	141	.11–.41
		CTOPP2 Elision Raw Score	.40	179	.27–.52
		CTOPP2 Blending Raw Score	.35	179	.21–.47
		CTOPP2 Sound Matching Raw Score	.39	179	.26–.51
	2	DIBELS Next ORF-WRC	.15	194	.01–.28
		DIBELS Next Composite	.17	161	.01–.31
	3	DIBELS Next ORF-WRC	.15	183	.01–.29
		DIBELS Next Composite	.14	163	-.01–.29
		Iowa Total Reading	.12	117	-.06–.30
		Iowa Word Analysis	.14	119	-.05–.31

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.9 Concurrent Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.34	116	.16–.49
		DIBELS Next Composite	.36	92	.26–.52
	2	DIBELS Next Composite	.68	133	.57–.76
	3	DIBELS Next Composite	.82	109	.75–.87
		Iowa Total Reading	.65	113	.53–.75
		Iowa Word Analysis	.43	96	.26–.58
First	1	DIBELS Next Composite	.71	166	.62–.78
	2	DIBELS Next ORF	.79	223	.74–.84
		DIBELS Next Composite	.80	185	.74–.85
	3	DIBELS Next ORF	.78	212	.72–.82
		DIBELS Next Composite	.85	186	.81–.89
		Iowa Total Reading	.65	198	.56–.73
		Iowa Word Analysis	.54	200	.43–.63
Second	1	DIBELS Next ORF	.80	129	.72–.85
		DIBELS Next Composite	.80	115	.72–.85
	2	DIBELS Next ORF	.75	211	.69–.81
		DIBELS Next Composite	.62	107	.49–.72
	3	DIBELS Next ORF	.76	201	.70–.82
		DIBELS Next Composite	.74	112	.65–.82
		Iowa Total Reading	.62	84	.47–.74
		Iowa Word Analysis	.60	51	.39–.75

Table 6.9 Concurrent Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Third	2	DIBELS Next ORF	.72	229	.65–.78
		DIBELS Next Composite	.71	109	.61–.80
	3	DIBELS Next ORF	.70	224	.63–.76
		DIBELS Next Composite	.69	99	.57–.78
		Iowa Total Reading	.50	90	.33–.64
		Iowa Word Analysis	.34	72	.12–.53

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.10 Concurrent Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.60	100	.46–.72
		DIBELS Next Composite	.60	76	.43–.73
	2	DIBELS Next NWF-CLS	.80	160	.53–.85
		DIBELS Next Composite	.66	130	.55–.75
	3	DIBELS Next NWF-CLS	.80	138	.73–.85
		DIBELS Next Composite	.74	108	.65–.82
		Iowa Total Reading	.65	112	.53–.74
		Iowa Word Analysis	.35	95	.16–.52

Table 6.10 Concurrent Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF-CLS	.66	180	.57-.73
		DIBELS Next Composite	.61	163	.50-.70
	2	DIBELS Next NWF-CLS	.78	223	.72-.82
		DIBELS Next ORF	.76	223	.70-.81
		DIBELS Next Composite	.79	185	.73-.84
	3	DIBELS Next NWF-CLS	.87	206	.84-.90
		DIBELS Next ORF	.78	207	.72-.83
		DIBELS Next Composite	.86	186	.81-.89
		Iowa Total Reading	.63	198	.54-.71
		Iowa Word Analysis	.56	200	.45-.65
Second	1	DIBELS Next NWF-CLS	.81	128	.74-.86
		DIBELS Next ORF	.77	128	.69-.84
		DIBELS Next Composite	.79	115	.71-.85
	2	DIBELS Next ORF	.76	211	.70-.81
		DIBELS Next Composite	.70	107	.58-.78
	3	DIBELS Next ORF	.80	198	.74-.84
		DIBELS Next Composite	.74	112	.64-.82
		Iowa Total Reading	.60	84	.45-.72
		Iowa Word Analysis	.62	51	.42-.76

Table 6.10 Concurrent Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	Period	Criterion	r	N	CI
Third	2	DIBELS Next ORF	.71	239	.64–.77
		DIBELS Next Composite	.74	109	.65–.82
	3	DIBELS Next ORF	.71	216	.64–.77
		DIBELS Next Composite	.73	97	.62–.81
		Iowa Total Reading	.51	90	.34–.65
		Iowa Word Analysis	.36	72	.14–.54

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.11 Concurrent Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.63	113	.51–.73
		DIBELS Next Composite	.57	89	.41–.70
	2	DIBELS Next NWF-CLS	.75	181	.68–.81
		DIBELS Next Composite	.63	124	.52–.73
	3	DIBELS Next NWF-CLS	.80	150	.74–.85
		DIBELS Next Composite	.75	92	.64–.82

Table 6.11 Concurrent Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF-CLS	.74	157	.66–.80
		DIBELS Next Composite	.69	144	.60–.77
	2	DIBELS Next ORF-WRC	.91	249	.89–.93
		DIBELS Next Composite	.88	163	.85–.91
	3	DIBELS Next ORF-WRC	.90	239	.87–.92
		DIBELS Next Composite	.88	163	.85–.91
		Iowa Total Reading	.79	117	.71–.85
		Iowa Word Analysis	.67	119	.56–.76
Second	1	DIBELS Next ORF-WRC	.91	166	.87–.93
		DIBELS Next Composite	.91	151	.88–.94
	2	DIBELS Next ORF-WRC	.92	192	.89–.94
		DIBELS Next Composite	.87	138	.83–.91
	3	Iowa Total Reading	.62	87	.47–.74
		Iowa Word Analysis	.60	89	.44–.72

Table 6.11 Concurrent Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Third	1	DIBELS Next ORF-WRC	.83	201	.78–.87
		DIBELS Next Composite	.85	138	.79–.89
	2	DIBELS Next ORF-WRC	.85	228	.81–.88
		DIBELS Next Composite	.85	154	.80–.89
	3	DIBELS Next ORF-WRC	.83	177	.78–.87
		DIBELS Next Composite	.84	97	.77–.89
		Iowa Total Reading	.56	90	.40–.69
		Iowa Word Analysis	.32	72	.09–.51

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.12 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF	.78	166	.72–.84
		DIBELS Next Composite	.75	154	.67–.81
	2	DIBELS Next Composite	.91	196	.88–.93
		DIBELS Next Composite	.91	163	.88–.94
	3	Iowa Total Reading	.82	116	.75–.87
		Iowa Word Analysis	.67	118	.55–.76

Table 6.12 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next Composite	.84	131	.78–.88
	2	DIBELS Next Composite	.87	130	.82–.90
	3	Iowa Total Reading	.71	87	.59–.80
		Iowa Word Analysis	.60	89	.45–.72
Third	1	DIBELS Next Composite	.89	73	.83–.93
	2	DIBELS Next Composite	.83	128	.77–.88
	3	DIBELS Next Composite	.83	96	.75–.88
		Iowa Total Reading	.58	90	.42–.70
		Iowa Word Analysis	.24	72	.01–.45
Fourth	3	Iowa Total Reading	.61	91	.47–.73
Fifth	1	DIBELS Next Composite	.94	41	.90–.97
	3	Iowa Total Reading	.65	59	.48–.78
Sixth	3	Iowa Total Reading	.67	82	.52–.77
Seventh	3	Iowa Total Reading	.54	91	.38–.67
Eighth	3	Iowa Total Reading	.59	77	.42–.72

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.13 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF	.67	166	.58–.75
		DIBELS Next Composite	.76	154	.69–.82
	2	DIBELS Next ORF	.69	227	.61–.75
		DIBELS Next Composite	.78	196	.72–.83
	3	DIBELS Next ORF	.66	188	.58–.74
		DIBELS Next Composite	.76	163	.68–.82
		Iowa Total Reading	.61	116	.48–.71
		Iowa Word Analysis	.60	118	.47–.71
Second	1	DIBELS Next ORF	.62	173	.51–.71
		DIBELS Next Composite	.63	131	.52–.73
	2	DIBELS Next ORF	.62	192	.52–.70
		DIBELS Next Composite	.68	130	.57–.76
	3	Iowa Total Reading	.48	87	.30–.62
Third	1	DIBELS Next ORF	.65	100	.52–.75
		DIBELS Next Composite	.68	73	.53–.78
	2	DIBELS Next ORF	.58	209	.48–.66
		DIBELS Next Composite	.68	128	.58–.76
	3	DIBELS Next ORF	.39	176	.25–.51
		DIBELS Next Composite	.55	96	.39–.67
		Iowa Total Reading	.36	90	.17–.53

Table 6.13 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
Fourth	3	DIBELS Next ORF	.57	58	.37–.72
		Iowa Total Reading	.37	91	.17–.53
Fifth	1	DIBELS Next ORF	.69	51	.52–.81
		DIBELS Next Composite	.73	41	.54–.85
	2	DIBELS Next ORF	.55	84	.38–.68
	3	Iowa Total Reading	.22	59	-.04–.45
Sixth	2	DIBELS Next ORF	.74	40	.56–.85
		Iowa Total Reading	.53	86	.36–.67
	3	Iowa Total Reading	.49	82	.31–.64
Seventh	3	Iowa Total Reading	.29	91	.09–.47
Eighth	3	Iowa Total Reading	.43	77	.22–.59

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.14 Concurrent Validity Coefficients for DIBELS 8th Edition Maze with Fall Administration of Iowa Assessment Total Reading

Grade	Benchmark form	r	N	CI
Second	1	.77	126	.69–.83
	2	.74	94	.63–.82
	3	.69	60	.53–.80
Third	1	.59	123	.46–.70
	2	.66	100	.53–.76
	3	.66	65	.49–.78
Fourth	1	.76	135	.67–.82
	2	.77	97	.67–.84
	3	.67	65	.51–.79
Fifth	1	.84	137	.72–.84
	2	.81	95	.62–.81
	3	.66	69	.31–.66
Sixth	1	.54	59	.33–.70
	2	.65	57	.47–.78
Seventh	1	.77	70	.65–.85
	2	.75	55	.61–.85

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Correlations among DIBELS 8 subtests. One way of establishing that subtests are measuring the same underlying construct, which here is reading, is to correlate them. Relationships should generally be positive and at least moderate, with more closely related subtests (e.g., WRF and ORF) exhibiting the strongest relationships. At the same time, correlations should not be perfect (i.e., 1.0), as that would indicate redundancy between the

subtests. Nonetheless, correlations look at the relationships among all test-takers and can obscure the added value for screening purposes of two subtests that are highly related on average. Thus, even though some subtests may be correlated at .9 or above, differences in predictive validity and screening results can suggest apparent redundancy is not problematic for screening purposes.

Correlations among DIBELS 8 subtests are reported by grade. In kindergarten, correlations are all positive, with the two NWF scores showing the strongest relationship with each other and WRF (see Table 6.15). LNF is most strongly related to NWF-CLS and moderately related to the other subtests. PSF has the weakest relations to the other subtests, as might be expected given that it is administered differently and taps a component skill that does not involve reading. In first grade, all DIBELS 8 subtests are strongly correlated except for PSF (see Table 6.16). Again, PSF shows moderate to weak correlations with the other subtests and the weakest correlations with ORF. From second grade onward, DIBELS 8 subtests administered in these grades are all strongly related (see Tables 6.17-6.19). In these later grades, ORF Accuracy shows the weakest relations, but the correlations are still strong. The two NWF scores demonstrate the strongest relationship in second and third grade, most likely due to the fact that the scores come from the same subtest and that students increasingly read without sounding out in these grades. ORF and WRF are also strongly related in second and third grade. Maze and ORF Accuracy have moderate to strong relationships with ORF but relatively weaker relations with each other.

Table 6.15 Correlations among DIBELS 8th Edition Subtests in Fall for Kindergarten

	NWF-CLS	NWF-WRC	PSF	WRF
LNF				
N	304	268	569	575
r	.61	.29	.42	.40
CI	.53–.68	.18–.40	.35–.48	.33–.47
NWF-CLS				
N		268	280	297
r		.79	.38	.71
CI		.74–.83	.27–.48	.65–.76

Table 6.15 Correlations among DIBELS 8th Edition Subtests in Fall for Kindergarten

	NWF-CLS	NWF-WRC	PSF	WRF
NWF-WRC				
N			253	263
r			.15	.75
CI			.03–.27	.69–.80
PSF				
N				536
r				.26
CI				.17–.33

Table 6.16 Correlations among DIBELS 8th Edition Subtests in Fall for First Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	PSF	WRF
LNF						
N	594	591	592	592	605	604
r	.67	.59	.53	.61	.32	.61
CI	.63–.72	.53–.64	.47–.59	.55–.65	.25–.39	.55–.65
NWF-CLS						
N		591	581	581	591	592
r		.90	.77	.69	.23	.83
CI		.88–.91	.74–.80	.65–.73	.15–.30	.80–.85

Table 6.16 Correlations among DIBELS 8th Edition Subtests in Fall for First Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	PSF	WRF
NWF-WRC						
N			579	579	588	589
r			.71	.61	.21	.78
CI			.67–.75	.56–.66	.13–.19	.75–.81
ORF						
N				592	589	592
r				.77	.08	.91
CI				.73–.80	.00–.16	.90–.93
ORF-ACC						
N					589	592
r					.27	.75
CI					.19–.34	.71–.78
PSF						
N						601
r						.18
CI						.10–.25

Table 6.17 Correlations among DIBELS 8th Edition Subtests in Fall for Second Grade

	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF-CLS					
N	467	463	461	469	469
r	.94	.75	.54	.77	.62
CI	.93–.95	.73–.80	.47–.60	.73–.81	.56–.67
NWF-WRC					
N		461	460	467	468
r		.75	.53	.76	.59
CI		.71–.79	.46–.59	.71–.79	.53–.65
ORF					
N			461	463	463
r			.69	.92	.73
CI			.64–.79	.91–.94	.68–.77
ORF-ACC					
N				461	461
r				.70	.48
CI				.65–.74	.40–.54
WRF					
N					469
r					.71
CI					.67–.76

Table 6.18 Correlations among DIBELS 8th Edition Subtests in Fall for Third Grade

	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF-CLS					
N	440	450	450	451	495
r	.96	.73	.49	.72	.50
CI	.95-.96	.68-.77	.41-.56	.67-.76	.43-.56
NWF-WRC					
N		439	439	440	484
r		.76	.56	.78	.56
CI		.71-.79	.49-.62	.75-.82	.49-.62
ORF					
N			450	450	494
r			.67	.89	.69
CI			.61-.72	.87-.91	.64-.74
ORF-ACC					
N				450	494
r				.70	.54
CI				.65-.74	.47-.60
WRF					
N					495
r					.68
CI					.63-.73

Table 6.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Fourth	ORF		
	N	532	506
	r	.70	.69
	CI	.65-.74	.65-.74
	ORF-ACC		
	N		506
	r		.50
	CI		.43-.56
Fifth	ORF		
	N	447	419
	r	.54	.44
	CI	.47-.60	.36-.51
	ORF-ACC		
	N		419
	r		.39
	CI		.30-.47

Table 6.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Sixth	ORF		
	N	197	191
	r	.59	.69
	CI	.49-.67	.61-.76
	ORF-ACC		
	N		191
	r		.35
	CI		.22-.47
Seventh	ORF		
	N	89	85
	r	.54	.75
	CI	.38-.68	.64-.83
	ORF-ACC		
	N		85
	r		.37
	CI		.17-.54

Table 6.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Eighth	ORF		
	N	82	76
	r	.78	.81
	CI	.67-.85	.71-.87
	ORF-ACC		
	N		76
	r		.53
	CI		.35-.68

Predictive Validity

Another way of establishing the validity of a test is by examining its ability to predict scores on criterion measures taken at a later time. Given the use of DIBELS as a screening and risk prediction tool, this type of validity evidence is arguably the most important. Predictive validity can be evaluated using correlations or through receiver operating characteristic (ROC) curve analyses. DIBELS 8 was evaluated through both methods and results are presented in this section.

Predictive correlations. DIBELS 8th Edition subtests were correlated with end of year administrations of DIBELS Next and the Iowa Assessment. The only exceptions are for NWF-CLS and NWF-WRC in the beginning of third grade and Maze in all grades. DIBELS Next related and predominant measure scores and the DIBELS Next composite score were used as criterion measures, as were Iowa Total Reading and Word Analysis scores. Results are presented in tables by subtest, grade, and the benchmark period in which DIBELS 8 was administered.

Correlations for each DIBELS 8 subtest are reported in Tables 6.20-6.26 and are only reported when the study sample size was 40 or greater to ensure a minimum threshold of precision in the correlation estimate. As with external criterion concurrent validity, where available, predictive validity with the DIBELS Next composite scores was quite good.

Correlations among unrelated DIBELS 8 and DIBELS Next subtests varied in strength as would be predicted based on the similarity of the component skills assessed. Finally, concurrent relations with the Iowa Assessment scores also vary predictably by subtest. PSF showed the weakest relationships with the Iowa total reading and word analysis scores in kindergarten and first grade. LNF was most weakly related to Iowa total reading in kindergarten, but otherwise was moderately strongly correlated with Iowa scores. The rest of the DIBELS 8 subtests displayed moderate to strong relations with external criterion measures regardless of grade.

Table 6.20 Predictive Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.79	149	.72–.84
		DIBELS Next NWF-CLS	.71	149	.62–.78
		DIBELS Next Composite	.78	123	.71–.84
	2	DIBELS Next LNF	.84	183	.79–.88
		DIBELS Next NWF-CLS	.72	183	.64–.78
		DIBELS Next Composite	.82	154	.76–.86
First	1	DIBELS Next ORF-WRC	.63	148	.52–.72
		DIBELS Next Composite	.64	132	.53–.73
		Iowa Total Reading	.57	80	.40–.70
		Iowa Word Analysis	.57	80	.40–.70
	2	DIBELS Next ORF-WRC	.71	191	.63–.77
		DIBELS Next Composite	.70	170	.62–.77
		Iowa Total Reading	.57	115	.43–.68
		Iowa Word Analysis	.52	117	.37–.64

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.21 Predictive Validity Coefficients for DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next PSF	.37	97	.18–.53
		DIBELS Next NWF-CLS	.24	97	.05–.42
		DIBELS Next Composite	.44	72	.23–.61
	2	DIBELS Next PSF	.75	29	.52–.88
		DIBELS Next NWF-CLS	.10	29	-.27–.45
		DIBELS Next Composite	.57	29	.26–.78
First	1	DIBELS Next ORF-WRC	.12	145	.00–.24
		DIBELS Next Composite	.11	129	-.02–.23
		Iowa Total Reading	.12	79	-.05–.28
		Iowa Word Analysis	.02	79	-.14–.19
	2	DIBELS Next ORF-WRC	.21	189	.07–.34
		DIBELS Next Composite	.23	168	.05–.31
		Iowa Total Reading	.29	113	.12–.45
		Iowa Word Analysis	.23	115	.08–.36

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.22 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next NWF-CLS	.45	113	.29–.59
		DIBELS Next Composite	.43	89	.24–.59
		Iowa Total Reading	.54	78	.36–.68
		Iowa Word Analysis	.29	80	.08–.48
	2	DIBELS Next NWF-CLS	.79	157	.72–.84
		DIBELS Next Composite	.72	130	.63–.80
		Iowa Total Reading	.54	111	.39–.66
		Iowa Word Analysis	.38	95	.19–.54
First	1	DIBELS Next NWF-CLS	.79	169	.73–.84
		DIBELS Next ORF	.76	167	.68–.81
		DIBELS Next Composite	.76	154	.69–.82
		Iowa Total Reading	.55	153	.43–.65
		Iowa Word Analysis	.51	153	.38–.62
	3	DIBELS Next NWF-CLS	.82	213	.77–.86
		DIBELS Next ORF	.77	214	.71–.82
		DIBELS Next Composite	.76	194	.70–.82
		Iowa Total Reading	.60	197	.50–.68
		Iowa Word Analysis	.49	199	.38–.59

Table 6.22 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.77	123	.69–.83
		DIBELS Next Composite	.72	97	.61–.80
		Iowa Total Reading	.66	49	.47–.79
	2	DIBELS Next ORF	.75	194	.68–.81
		DIBELS Next Composite	.67	117	.55–.76
		Iowa Total Reading	.56	76	.38–.70
		Iowa Word Analysis	.47	49	.22–.66
Third	3	DIBELS Next ORF	.66	222	.58–.73
		DIBELS Next Composite	.66	107	.54–.76
		Iowa Total Reading	.39	89	.19–.55
		Iowa Word Analysis	.27	72	.05–.48

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.23 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly with End of Year Measures

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next NWF-CLS	.65	97	.52–.75
		DIBELS Next Composite	.71	73	.58–.81
		Iowa Total Reading	.47	75	.27–.63
		Iowa Word Analysis	.18	77	-.05–.39
	2	DIBELS Next NWF-CLS	.74	153	.66–.80
		DIBELS Next Composite	.71	126	.61–.79
		Iowa Total Reading	.55	106	.40–.67
		Iowa Word Analysis	.26	90	.06–.44
First	1	DIBELS Next NWF-CLS	.71	163	.63–.78
		DIBELS Next ORF	.69	163	.59–.76
		DIBELS Next Composite	.70	163	.61–.77
		Iowa Total Reading	.51	153	.39–.62
		Iowa Word Analysis	.50	153	.37–.61
	2	DIBELS Next NWF-CLS	.76	213	.70–.81
		DIBELS Next ORF	.75	214	.68–.80
		DIBELS Next Composite	.75	185	.68–.80
		Iowa Total Reading	.58	197	.47–.66
		Iowa Word Analysis	.51	199	.40–.61

Table 6.23 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly with End of Year Measures

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.76	123	.67–.83
		DIBELS Next Composite	.71	97	.60–.85
		Iowa Total Reading	.64	49	.44–.78
	2	DIBELS Next ORF	.76	194	.69–.81
		DIBELS Next Composite	.70	117	.58–.78
		Iowa Total Reading	.60	76	.43–.73
		Iowa Word Analysis	.56	49	.33–.73
Third	2	DIBELS Next ORF	.66	222	.58–.73
		DIBELS Next Composite	.69	107	.65–.82
		Iowa Total Reading	.41	89	.22–.57
		Iowa Word Analysis	.31	72	.09–.51

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.24 Predictive Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.72	112	.61–.80
		DIBELS Next Composite	.65	86	.50–.76
	2	DIBELS Next NWF-CLS	.71	177	.63–.78
		DIBELS Next Composite	.67	121	.56–.76
First	1	DIBELS Next NWF-CLS	.84	148	.79–.88
		DIBELS Next Composite	.78	132	.71–.84
		Iowa Total Reading	.65	80	.50–.76
		Iowa Word Analysis	.59	80	.43–.72
	2	DIBELS Next ORF-WRC	.87	239	.84–.90
		DIBELS Next Composite	.85	170	.80–.89
		Iowa Total Reading	.74	115	.64–.81
		Iowa Word Analysis	.63	117	.50–.73
Second	1	DIBELS Next ORF-WRC	.88	156	.84–.91
		DIBELS Next Composite	.87	130	.82–.91
		Iowa Total Reading	.78	51	.64–.87
		Iowa Word Analysis	.66	36	.42–.81
	2	DIBELS Next ORF-WRC	.90	189	.87–.92
		DIBELS Next Composite	.87	150	.82–.90
		Iowa Total Reading	.72	77	.59–.81
		Iowa Word Analysis	.69	49	.50–.81

Table 6.24 Predictive Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Third	1	DIBELS Next ORF-WRC	.82	192	.77–.87
		DIBELS Next Composite	.84	131	.78–.88
		Iowa Total Reading	.61	71	.44–.74
		Iowa Word Analysis	.47	53	.23–.66
	2	DIBELS Next ORF-WRC	.82	228	.77–.86
		DIBELS Next Composite	.82	154	.77–.87
		Iowa Total Reading	.59	89	.43–.71
		Iowa Word Analysis	.46	72	.25–.62

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next ORF	.86	156	.82–.90
		DIBELS Next Composite	.82	141	.76–.87
		Iowa Total Reading	.73	59	.58–.83
		Iowa Word Analysis	.60	59	.40–.74
	2	DIBELS Next ORF	.91	223	.88–.93
		DIBELS Next Composite	.88	203	.84–.91
		Iowa Total Reading	.79	115	.71–.85
		Iowa Word Analysis	.69	117	.58–.77

Table 6.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.84	148	.78–.88
		DIBELS Next Composite	.79	121	.71–.85
		Iowa Total Reading	.63	137	.51–.72
		Iowa Word Analysis	.55	129	.41–.66
	2	DIBELS Next ORF	.93	180	.91–.95
		DIBELS Next Composite	.89	140	.85–.92
		Iowa Total Reading	.74	163	.66–.80
		Iowa Word Analysis	.64	137	.53–.73
Third	1	DIBELS Next ORF	.94	84	.91–.96
		DIBELS Next Composite	.93	81	.89–.95
		Iowa Total Reading	.74	61	.61–.84
	2	DIBELS Next ORF	.90	203	.87–.92
		DIBELS Next Composite	.86	126	.80–.90
		Iowa Total Reading	.69	132	.59–.77
		Iowa Word Analysis	.50	96	.33–.64
Fourth	1	DIBELS Next ORF	.91	28	.82–.96
		Iowa Total Reading	.53	55	.30–.70
	2	Iowa Total Reading	.67	55	.49–.79
Fifth	1	Iowa Total Reading	.63	128	.52–.73
	2	DIBELS Next ORF	.91	85	.86–.94
		Iowa Total Reading	.69	99	.57–.78

Table 6.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
Sixth	1	Iowa Total Reading	.50	49	.25–.68
	2	Iowa Total Reading	.65	86	.50–.75
Seventh	1	Iowa Total Reading	.52	59	.31–.69
	2	Iowa Total Reading	.57	101	.42–.69
Eighth	1	Iowa Total Reading	.69	46	.50–.82
	2	Iowa Total Reading	.48	85	.30–.63

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 6.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next ORF	.82	156	.77–.87
		DIBELS Next Composite	.87	141	.82–.90
		Iowa Total Reading	.82	59	.71–.89
		Iowa Word Analysis	.72	59	.56–.82
	2	DIBELS Next ORF	.75	223	.69–.81
		DIBELS Next Composite	.83	203	.78–.87
		Iowa Total Reading	.78	115	.70–.84
		Iowa Word Analysis	.72	117	.62–.80

Table 6.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.62	148	.51–.71
		DIBELS Next Composite	.70	121	.60–.78
		Iowa Total Reading	.61	137	.49–.70
		Iowa Word Analysis	.46	129	.31–.59
	2	DIBELS Next ORF	.61	180	.51–.69
		DIBELS Next Composite	.69	140	.59–.77
		Iowa Total Reading	.51	163	.38–.61
		Iowa Word Analysis	.54	137	.41–.65
Third	1	DIBELS Next ORF	.73	84	.61–.82
		DIBELS Next Composite	.76	81	.65–.84
		Iowa Total Reading	.67	61	.51–.79
	2	DIBELS Next ORF	.56	203	.46–.65
		DIBELS Next Composite	.70	126	.70–.78
		Iowa Total Reading	.53	132	.53–.64
		Iowa Word Analysis	.44	96	.44–.59
Fourth	1	Iowa Total Reading	.44	55	.20–.63
	2	Iowa Total Reading	.37	55	.11–.58
Fifth	1	Iowa Total Reading	.46	128	.31–.59
	2	DIBELS Next ORF	.60	85	.44–.72
		Iowa Total Reading	.41	99	.23–.56

Table 6.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
Sixth	1	Iowa Total Reading	.46	49	.20–.65
	2	Iowa Total Reading	.53	86	.36–.67
Seventh	1	Iowa Total Reading	.29	59	.04–.51
	2	Iowa Total Reading	.36	101	.17–.52
Eighth	1	Iowa Total Reading	.44	46	.17–.65
	2	Iowa Total Reading	.23	85	.02–.42

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Screening accuracy. One of the uses of DIBELS 8th Edition is to identify students who are not on track for meeting reading proficiency standards and those who are at pronounced risk for reading difficulties. To support this use, we have provided two types of cut-scores for classifying students. The first score, called the risk cut-score, can be used to classify students who are at risk for reading difficulties, including dyslexia. The second score, called the benchmark goal, can be used to classify students who are at some risk for not meeting proficiency goals versus those who are on track for meeting proficiency goals.

The cut-scores were calculated using ROC curve analyses. ROC analyses describe the relation between true positive rates (i.e., scores that correctly identify students who were not on track for attaining proficiency) and false positive rates (i.e., scores that indicate a student was not on-track when they really were). In this case, the ROC results describe whether DIBELS 8th Edition scores correctly predicted performance on a criterion measure of reading: DIBELS Next composite score percentile ranks in kindergarten or Iowa Assessment Total Reading percentile ranks in all other grades. ROC analyses yield an area under the curve (AUC) estimate, which describes a test's classification accuracy. An AUC of .5 indicates the test predicts no better than chance. An AUC of 1.0 indicates that a test has perfect predictive power (Habibzadeh, Habibzadeh, & Yadollahie, 2016).

In addition to reporting the AUC for each benchmark subtest, sensitivity and specificity are reported. Sensitivity provides information about how well a subtest's cut-score identifies students who have not met a criterion goal. It is expressed as a proportion, ranging from 0 to 1. The sensitivity value represents the proportion of "truly" at-risk students who are correctly

identified by the screener as being at risk. Specificity, which is also expressed as a proportion, is the counterpart to sensitivity. Specificity represents the proportion of “truly healthy” readers who are accurately identified as not at risk by the screener (i.e., identified as “okay”). Sensitivity can also be interpreted as the probability (likelihood) that a student who meets the criterion goal has been identified as such by the screener.

Although sensitivity and specificity are stable indicators of screening effectiveness regardless of the prevalence of reading difficulties in the population (Pepe, 2003), an important determinant of sensitivity and specificity that does not affect the AUC is how the cut-score is set for a subtest. DIBELS 8 balances sensitivity and specificity because of their complementary roles in a prevention model in education. Specifically, balancing both statistics results in maximizing the proportion of students correctly identified for intervention without under-identifying students correctly identified as not in need of intervention. Thus, wherever possible, recommended cut points for DIBELS 8th Edition subtests were set to maximize sensitivity while maintaining specificity at or above .80. More specifically, for each benchmark, the cut was set at the score with the highest sensitivity among scores with a specificity at or above .80. In cases where the maximum sensitivity value exceeded .90, the cut point selected was the score that minimized the difference between sensitivity and specificity among scores with specificity at or above .80; in other words, maximizing both statistics. For the few benchmarks where no cut scores satisfied these criteria, the cut was set at the score that minimized the difference between sensitivity and specificity.

AUC, sensitivity, and specificity results are reported by grade and within grade by subtest and benchmark period for two cuts on a criterion measure (i.e., 20th percentile rank and 40th percentile rank). For kindergarten, the DIBELS Next composite score in the end of year served as the criterion measure, and in all other grades the criterion measure was the Iowa Assessment Total Reading Score for the end of year. The only exception to this is for NWF in the beginning of third grade and Maze, where the Iowa was administered in the beginning of year. The Iowa Assessment is a published, group-administered, multiple-choice, norm-referenced measure of reading achievement (Welch & Dunbar, 2012).

Regardless of criterion measure, the 20th percentile rank cut is intended for use in identifying students who are well below benchmark, at risk for not meeting end of year learning goals, and in need of intensive intervention. For the LNF, PSF, and NWF subtests, students falling below this cut may also be at risk for reading disabilities, including dyslexia. The 40th percentile cut is intended for use in identifying students who are below benchmark, at some risk or not meeting end of year learning goals, and in need of some support.

In kindergarten, AUCs are uniformly high with the majority falling at .8 or above and sensitivity and specificity are also routinely high (see Table 6.27). AUC, sensitivity, and

specificity statistics are not reported for the NWF-WRC 40th percentile cut-score in beginning and middle of year benchmarks in kindergarten because the cut was identical to the 20th percentile cut. LNF and PSF are the strongest indicators for the beginning of kindergarten, while NWF-CLS and WRF increase in strength over the kindergarten year.

In first grade, for NWF scores, WRF, and ORF scores, AUCs are uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 6.28). They are somewhat lower for LNF and PSF, which is likely due to the change in criterion measure from DIBELS Next Composite Score to the Iowa Assessment Total Reading Score. Whereas DIBELS Next includes letter naming and phonemic awareness component skills in the composite score, the Iowa Total Reading Score does not assess these same component skills. While NWF is a robust predictor, the strongest predictors in first grade are WRF and ORF.

In second grade, AUCs are again uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 6.29). While NWF remains a robust predictor, the strongest predictors in second grade are WRF, ORF, and Maze. However, ORF Accuracy shows a declining trend in its predictive power.

In third grade, AUCs are again uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 6.30). The strongest predictors in third grade are ORF and Maze, although NWF and WRF remain good predictors. ORF Accuracy again shows a declining trend in its predictive power.

In fourth, fifth, and sixth grade, AUCs, sensitivity, and specificity for ORF and Maze remain quite high (see Tables 6.31, 6.32, and 6.33 respectively). ORF Accuracy shows the same declining trend in predictive power, and is particularly pronounced in fifth grade.

In seventh and eighth grade, AUCs, sensitivity, and specificity for ORF remain quite high (see Tables 6.34 and 6.35 respectively). ORF Accuracy shows the same declining trend in predictive power as in other upper elementary grades. Maze ROC analyses are not yet available in seventh and eighth grade, but results will be released during 2019 in an addendum to this manual.

Table 6.27 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	123	.89	.82–.96	.79	.81
		2	154	.92	.88–.96	.89	.81
		3	98	.94	.89–.99	.86	.83
	40th	1	123	.90	.84–.96	.88	.81
		2	154	.93	.89–.96	.86	.83
		3	98	.96	.92–.99	.94	.86
PSF	20th	1	72	.86	.75–.98	.80	.83
		2	29	.98	.92–.99+	.99+	.96
		3	95	.92	.85–.98	.90	.90
	40th	1	72	.79	.67–.91	.70	.76
		2	29	.88	.69–.99+	.83	.91
		3	95	.84	.74–.93	.83	.79
NWF-CLS	20th	1	89	.73	.63–.84	.83	.70
		2	130	.88	.81–.95	.77	.82
		3	109	.90	.84–.96	.82	.84
	40th	1	89	.73	.63–.83	.78	.69
		2	130	.85	.77–.92	.80	.83
		3	109	.93	.88–.98	.84	.87

Table 6.27 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	73	.73	.59–.88	.67	.66
		2	136	.82	.77–.88	.94	.68
		3	108	.86	.79–.93	.77	.82
	40th	1	NA	NA	NA	NA	NA
		2	NA	NA	NA	NA	NA
		3	109	.82	.74–.90	.72	.72
WRF	20th	1	86	.81	.75–.87	.99+	.62
		2	121	.86	.84–.92	.96	.76
		3	92	.95	.90–.99	.88	.83
	40th	1	86	.86	.80–.91	.99+	.71
		2	121	.86	.80–.92	.83	.81
		3	92	.94	.89–.99	.85	.85

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period.

Table 6.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	80	.83	.73–.93	.76	.82
		2	115	.80	.72–.88	.71	.82
		3	117	.77	.68–.86	.69	.68
	40th	1	80	.82	.73–.91	.72	.76
		2	115	.78	.70–.87	.71	.72
		3	117	.78	.69–.86	.69	.73
PSF	20th	1	79	.53	.38–.68	.54	.55
		2	113	.66	.54–.77	.62	.66
		3	117	.58	.46–.69	.62	.60
	40th	1	79	.54	.41–.67	.52	.51
		2	113	.61	.50–.71	.55	.58
		3	117	.54	.43–.65	.57	.56
NWF-CLS	20th	1	153	.83	.76–.90	.77	.75
		2	196	.86	.80–.92	.75	.81
		3	198	.88	.83–.93	.86	.80
	40th	1	153	.76	.72–.87	.73	.71
		2	196	.82	.76–.88	.75	.77
		3	198	.87	.82–.92	.83	.82

Table 6.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	153	.75	.67–.83	.72	.68
		2	196	.82	.76–.88	.75	.74
		3	198	.86	.80–.90	.75	.80
	40th	1	153	.78	.70–.85	.74	.71
		2	196	.81	.75–.87	.73	.71
		3	198	.84	.79–.90	.71	.80
WRF	20th	1	80	.88	.79–.96	.83	.80
		2	115	.89	.82–.95	.86	.84
		3	117	.90	.84–.97	.88	.88
	40th	1	80	.93	.88–.99	.88	.84
		2	115	.96	.92–.99	.92	.88
		3	117	.95	.90–.99+	.92	.90
ORF	20th	1	59	.87	.77–.98	.85	.85
		2	115	.91	.85–.97	.88	.85
		3	116	.89	.83–.96	.85	.89
	40th	1	59	.96	.90–.99+	.90	.97
		2	115	.97	.93–.99+	.92	.92
		3	116	.94	.90–.99	.88	.90

Table 6.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF-ACC	20th	1	59	.87	.77–.98	.85	.87
		2	115	.90	.84–.96	.86	.86
		3	116	.88	.82–.95	.83	.84
	40th	1	59	.95	.89–.99+	.87	.93
		2	115	.95	.92–.99	.89	.92
		3	116	.94	.89–.99	.94	.89

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period.

Table 6.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1	49	.87	.77–.97	.76	.82
		2	76	.81	.71–.91	.76	.77
		3	84	.79	.69–.89	.70	.72
	40th	1	49	.84	.70–.97	.78	.83
		2	76	.82	.69–.95	.86	.80
		3	84	.83	.73–.92	.79	.73

Table 6.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	49	.86	.75–.97	.76	.86
		2	76	.83	.74–.92	.73	.77
		3	84	.77	.66–.87	.70	.75
	40th	1	49	.85	.72–.97	.70	.83
		2	76	.81	.69–.93	.80	.75
		3	84	.79	.67–.90	.69	.77
WRF	20th	1	51	.93	.85–.99+	.91	.93
		2	77	.90	.82–.98	.91	.84
		3	87	.88	.79–.96	.83	.91
	40th	1	51	.90	.82–.99	.82	.83
		2	77	.84	.73–.95	.82	.81
		3	87	.86	.77–.94	.79	.83
ORF	20th	1	116	.89	.81–.98	.87	.82
		2	163	.91	.87–.95	.89	.82
		3	87	.93	.87–.99	.92	.85
	40th	1	116	.86	.79–.93	.77	.81
		2	163	.88	.83–.93	.81	.81
		3	87	.91	.83–.98	.90	.85

Table 6.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF-ACC	20th	1	116	.87	.77–.97	.83	.81
		2	163	.86	.79–.92	.73	.84
		3	87	.74	.58–.92	.67	.75
	40th	1	116	.82	.74–.90	.71	.81
		2	163	.83	.77–.90	.74	.79
		3	87	.73	.62–.84	.72	.67
Maze	20th	1*	119	.91	.86–.96	.88	.86
		2*	87	.95	.91–.99+	.90	.91
		3*	60	.86	.77–.95	.99+	.49
	40th	1*	119	.96	.94–.99	.80	.97
		2*	87	.92	.85–.99	.84	.93
		3*	60	.94	.88–.99	.99+	.45

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.30 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1*	95	.84	.76–.93	.70	.82
		2	89	.71	.57–.85	.70	.74
		3	90	.77	.65–.89	.65	.63
	40th	1*	95	.84	.75–.93	.72	.82
		2	89	.71	.60–.82	.65	.65
		3	90	.75	.64–.85	.73	.63
NWF-WRC	20th	1*	95	.84	.76–.93	.70	.82
		2	89	.71	.57–.85	.70	.65
		3	90	.76	.64–.88	.65	.61
	40th	1*	95	.84	.75–.93	.72	.82
		2	89	.71	.61–.82	.64	.68
		3	90	.73	.68–.83	.65	.61
WRF	20th	1	71	.83	.70–.96	.69	.71
		2	89	.82	.72–.92	.70	.75
		3	90	.80	.70–.90	.70	.67
	40th	1	71	.79	.68–.90	.71	.80
		2	89	.80	.71–.89	.71	.70
		3	90	.79	.70–.88	.71	.71

Table 6.30 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	43	.91	.80–.99+	.99+	.83
		2	132	.86	.78–.94	.78	.86
		3	90	.82	.72–.92	.75	.77
	40th	1	43	.83	.70–.95	.72	.80
		2	132	.84	.77–.90	.74	.82
		3	90	.77	.67–.87	.73	.71
ORF-ACC	20th	1	43	.92	.80–.99+	.99+	.83
		2	132	.84	.73–.94	.83	.77
		3	90	.77	.69–.89	.65	.73
	40th	1	43	.86	.74–.99+	.78	.84
		2	132	.81	.75–.89	.77	.73
		3	90	.70	.62–.82	.77	.50
Maze	20th	1*	123	.91	.86–.97	.88	.80
		2*	100	.92	.86–.97	.89	.73
		3*	65	.90	.82–.98	.90	.74
	40th	1*	123	.81	.73–.89	.71	.80
		2*	100	.83	.74–.91	.77	.79
		3*	65	.81	.68–.93	.96	.29

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.31 ROC Curve Results for DIBELS 8 Fourth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	55	.80	.67–.93	.83	.69
		2	55	.84	.68–.99+	.83	.78
		3	91	.81	.72–.89	.71	.70
	40th	1	55	.75	.61–.88	.77	.73
		2	55	.82	.71–.93	.68	.73
		3	91	.80	.70–.91	.73	.71
ORF-ACC	20th	1	55	.70	.43–.97	.67	.74
		2	55	.81	.66–.98	.67	.76
		3	91	.70	.60–.81	.73	.59
	40th	1	55	.73	.57–.86	.59	.79
		2	55	.41	.46–.76	.68	.49
		3	91	.71	.62–.82	.66	.71
Maze	20th	1*	133	.93	.88–.98	.86	.88
		2*	97	.92	.86–.99	.91	.78
		3*	64	.85	.75–.95	.96	.35
	40th	1*	133	.81	.73–.89	.85	.49
		2*	97	.83	.74–.91	.83	.70
		3*	64	.81	.68–.93	.99+	.18

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.32 ROC Curve Results for DIBELS 8 Fifth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	128	.83	.76–.90	.72	.74
		2	99	.82	.74–.90	.72	.75
		3	59	.86	.70–.99+	.78	.82
	40th	1	128	.81	.73–.88	.72	.71
		2	99	.80	.70–.89	.65	.68
		3	59	.89	.80–.97	.71	.84
ORF-ACC	20th	1	128	.70	.61–.81	.58	.63
		2	99	.69	.60–.81	.69	.60
		3	59	.39	.41–.86	.44	.78
	40th	1	128	.65	.56–.75	.74	.57
		2	99	.71	.61–.82	.62	.74
		3	59	.50	.35–.66	.57	.42
Maze	20th	1*	137	.89	.83–.95	.74	.84
		2*	95	.81	.71–.92	.81	.76
		3*	68	.78	.67–.90	.96	.59
	40th	1*	137	.92	.88–.96	.81	.82
		2*	95	.87	.79–.94	.90	.78
		3*	68	.80	.69–.92	.95	.40

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.33 ROC Curve Results for DIBELS 8 Sixth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	49	.77	.62–.91	.79	.72
		2	86	.82	.73–.91	.78	.80
		3	82	.82	.73–.92	.74	.82
	40th	1	49	.82	.62–.91	.71	.79
		2	86	.82	.73–.91	.70	.86
		3	82	.84	.73–.92	.74	.86
ORF-ACC	20th	1	49	.77	.63–.91	.71	.68
		2	86	.77	.67–.88	.68	.82
		3	82	.71	.59–.82	.58	.82
	40th	1	49	.67	.52–.83	.57	.64
		2	86	.79	.70–.88	.55	.96
		3	82	.80	.72–.88	.69	.81
Maze	20th	1*	59	.86	.69–.99+	.83	.79
		2*	57	.90	.77–.99+	.60	.92
		3*	TBD	TBD	TBD	TBD	TBD
	40th	1*	59	.80	.66–.93	.73	.73
		2*	57	.81	.64–.98	.70	.81
		3*	TBD	TBD	TBD	TBD	TBD

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.34 ROC Curve Results for DIBELS 8 Seventh Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	59	.80	.68–.92	.85	.72
		2	109	.75	.66–.85	.69	.68
		3	99	.78	.70–.88	.71	.71
	40th	1	59	.76	.64–.89	.71	.76
		2	109	.83	.74–.91	.79	.75
		3	99	.79	.70–.89	.72	.72
ORF-ACC	20th	1	59	.79	.69–.92	.67	.84
		2	109	.71	.60–.80	.67	.60
		3	99	.69	.58–.80	.65	.59
	40th	1	59	.73	.65–.89	.53	.86
		2	109	.82	.75–.92	.65	.82
		3	99	.76	.68–.87	.64	.80

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 6.35 ROC Curve Results for DIBELS 8 Eighth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	46	.82	.69–.94	.71	.73
		2	85	.73	.68–.85	.75	.66
		3	77	.74	.63–.86	.64	.70
	40th	1	46	.89	.79–.98	.78	.86
		2	85	.81	.69–.94	.80	.80
		3	77	.89	.80–.99	.87	.81
ORF-ACC	20th	1	46	.79	.65–.92	.63	.73
		2	85	.70	.59–.82	.68	.68
		3	77	.68	.57–.80	.64	.64
	40th	1	46	.82	.70–.97	.81	.71
		2	85	.76	.65–.90	.79	.65
		3	77	.71	.58–.84	.57	.69

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Summary

Taken together, the validity evidence for DIBELS 8 is strong. The strongest evidence regards its primary use, which is as a screener for students at risk for reading difficulties, including reading disabilities like dyslexia. Research into the valid interpretations and uses of DIBELS scores is ongoing, and regular addendums to this manual will continue to build the validity argument for DIBELS 8.

References

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association (AERA); American Psychological Association (APA); National Council on Measurement in Education (NCME). Retrieved from <http://www.apa.org/science/programs/testing/standards.aspx>
- Al Otaiba, S., McDonald Connor, C., Foorman, B., Schatschneider, C., Greulich, L., & Sidler, J. F. (2009). Identifying and intervening with beginning readers who are at-risk for dyslexia: Advances in individualized classroom instruction. *Perspectives on language and literacy*, 35(4), 13–19. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296731/>
- Alonzo, J., & Tindal, G. (2007). *The development of fifth-grade passage reading fluency measures for use in a progress monitoring assessment system (Technical report No. 40)*. Eugene, OR: Behavioral Research and Teaching. Retrieved from <http://www.brtprojects.org/publications/dl/54>
- Baker, D. L., Biancarosa, G., Park, B. J., Bousselot, T., Smith, J. L. M., Baker, S. K., . . . Tindal, G. (2015). Validity of CBM measures of oral reading fluency and reading comprehension on high-stakes reading assessments in grades 7 and 8. *Reading and Writing*, 28, 57–104. doi: 10.1007/s11145-014-9505-4
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445–459. doi: 10.3758/bf03193014
- Biemiller, A. (2010). *Words Worth Teaching: Closing the Vocabulary Gap*. Columbus, OH: McGraw-Hill, SRA.
- Bizzocchi, A. (2017). How many phonemes does the English language have? *International Journal on Studies in English Language and Literature*, 5, 36–46. doi: 10.20431/2347-3134.0510006
- Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, 49(4), 1520–1523. doi: 10.3758/s13428-016-0811-4

- Burns, M. K., Ardoin, S. P., Parker, D. C., Hodgson, J., Klingbeil, D. A., & Scholin, S. E. (2009). Interspersal technique and behavioral momentum for reading word lists. *School Psychology Review*, 38, 428–434.
- Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, 42, 163–176. doi: 10.1177/0022219408326219
- Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Fort Worth, TX: Harcourt Brace.
- Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology*, 47, 55–75. doi: 10.1016/j.jsp.2008.09.004
- Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly*, 26, 231–244. doi: 10.1037/a0025173
- Conoyer, S. J., Lembke, E. S., Hosp, J. L., Espin, C. A., Hosp, M. K., & Poch, A. L. (2017). Getting more from your maze: Examining differences in distractors. *Reading & Writing Quarterly*, 33(2), 141–154. doi: 10.1080/10573569.2016.1142913
- Cummings, K. D., Park, Y., & Bauer Schaper, H. A. (2013). Form effects on DIBELS next Oral Reading Fluency progress-monitoring passages. *Assessment for Effective Intervention*, 38, 91–104. doi: 10.1177/1534508412447010
- Dale, E., & O'Rourke, J. (1981). *The living word vocabulary: A national vocabulary inventory*. Chicago, IL: World Book-Childcraft International.
- Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, 15(3), 358–374.
- Duranovic, M., Senka, S., & Babic-Gavric, B. (2018). Influence of increased letter spacing and font type on the reading ability of dyslexic children. *Annals of Dyslexia*, 68(3), 218–228. doi: <https://doi.org/10.1007/s11881-018-0164-z>
- Dyson, M. C., & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54(4), 585–612. doi: <https://doi.org/10.1006/ijhc.2001.0458>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167–188. doi: 10.1207/s1532799xssr0902_4

- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315–342. doi: 10.1016/j.jsp.2007.06.003
- Fry, E. (2004). Phonics: A large phoneme - grapheme frequency count revised. *Journal of Literacy Research, 36*(1), 85–98. doi: 10.1207/s15548430jlr3601_5
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7–21.
- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*, 195–203. doi: 10.1177/0022219412442150
- Geverdt, D. (2015). Education Demographic and Geographic Estimates Program (EDGE): Locale Boundaries User's Manual (NCES 2016-012). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved January 11, 2019 from <http://nces.ed.gov/pubsearch>.
- Gersten, R. M., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades: A practice guide (Report No. NCE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_reading_pg_021809.pdf
- Good, R., & Kaminski, R. A. (2009). DIBELS next. Dallas, TX: Voyager Sopris Learning.
- Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: The case of tests with continuous results. *Biochemia Medica, 26*(3), 297–307. doi: 10.11613/BM.2016.034
- Hughes, L., & Wilkins, A. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading, 23*(3), 314–324. doi: 10.1111/1467-9817.00126
- International Dyslexia Association. (2019). Universal screening: K–2 reading. Retrieved from <https://dyslexiaida.org/universal-screening-k-2-reading/>
- January, S.-A. A., & Ardoin, S. P. (2012). The impact of context and word type on students' maze task accuracy. *School Psychology Review, 41*, 262–271.

- January, S.-A. A., Ardoin, S. P., Christ, T. J., Eckert, T. L., & White, M. J. (2016). Evaluating the interpretations and use of curriculum-based measurement in reading and word lists for universal screening in first and second grade. *School Psychology Review*, 45(3), 310–326. doi: 10.17105/SPR45-3.310-326
- Jones, M. N., & Mewhort, D. J. K. (2004). Case-sensitive letter and bigram frequency counts from large-scale English corpora. *Behavior Research Methods, Instruments, & Computers*, 36(3), 388–396. doi: 10.3758/bf03195586
- Katzir, T., Hershko, S., & Halamish, V. (2013). The effect of font size on reading comprehension on second and fifth grade children: Bigger is not always better. *PloS one*, 8(9), e74061–e74061. doi: 10.1371/journal.pone.0074061
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge Univ Pr.
- Kučera, H., & Francis, N. W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuster, S. M., van Weerdenburg, M., Gompel, M., & Bosman, A. M. T. (2018). Dyslexie font does not benefit reading in children with or without dyslexia. *Annals of Dyslexia*, 68(1), 25–42. doi: 10.1007/s11881-017-0154-6
- Lockenvitz, S., Kuecker, K., & Ball, M. J. (2015). Evidence for the distinction between “consonantal-/r/” and “vocalic-/r/” in American English. *Clinical Linguistics & Phonetics*, 29(8-10), 613–622. doi: 10.3109/02699206.2015.1047962
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2), 203–208. doi: 10.3758/BF03204766
- Marinus, E., Mostard, M., Segers, E., Schubert, T. M., Madelaine, A., & Wheldall, K. (2016). A special font for people with dyslexia: Does it work and, if so, why? *Dyslexia*, 22(3), 233–244. doi: 10.1002/dys.1527
- Marston, D. B., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. L. Graden, J. Zins & M. E. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137–172). Washington, DC: National Association of School Psychology.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. doi: 10.1037/0003-066X.50.9.741

- Meyer, M. S. (2000). The ability–achievement discrepancy: Does it contribute to an understanding of learning disabilities? *Educational Psychology Review*, 12(3), 315–337. doi: 10.1023/A:1009070006373
- National Institute of Child Health and Human Development. (2000). Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for reading instruction. Bethesda, MD: NICHD Clearinghouse. Retrieved from <http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf>
- Nese, J. F. T., Kamata, A., & Tindal, G. (2017). A two-step sampling weight approach to growth mixture modeling for emergent and developing skills with distributional changes over time. *Journal of School Psychology*, 61, 55–74. doi: 10.1016/j.jsp.2016.12.001
- Norvig, P. (2012). English letter frequency counts Mayzner Revisited or ETAOIN SRHLDCU. Retrieved from <http://norvig.com/mayzner.html>
- O'Brien, B. A., Mansfield, J. S., & Legge, G. E. (2005). The effect of print size on reading speed in dyslexia. *Journal of Research in Reading*, 28(3), 332–349. doi: 10.1111/j.1467-9817.2005.00273.x
- Paris, S., & Hamilton, E. E. (2009). The development of children's reading comprehension. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 32–53). New York: Lawrence Erlbaum.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40, 184–202.
- Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford.
- Petscher, Y., & Kim, Y.-S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, 49, 107–129. doi: 10.1016/j.jsp.2010.09.004
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Reed, D. K., Cummings, K. D., Schaper, A., Lynn, D., & Biancarosa, G. (2018). Accuracy in identifying students' miscues during oral reading: A taxonomy of scorers' mismarkings. *Reading and Writing*, Advanced online publication. doi: 10.1007/s11145-018-9899-5
- Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, 42(4), 546–567. doi: 10.1598/RRQ.42.4.5

- Santi, K. L., Barr, C., Khalaf, S., & Francis, D. J. (2016). Different approaches to equating oral reading fluency passages. In K. D. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications*. (pp. 223–265). New York, NY, US: Springer Science + Business Media.
- Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 17, 229–255. Retrieved from <http://www.jstor.org/stable/747485>
- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.
- Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guildford Press.
- Smith, J. L. M., Cummings, K. D., Nese, J. F. T., Alonzo, J., Fien, H., & Baker, S. K. (2014). The relation of word reading fluency initial level and gains with reading outcomes. *School Psychology Review*, 43, 30–40.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press, National Research Council. Retrieved from <http://files.eric.ed.gov/fulltext/ED416465.pdf>
- University of Iowa. (2015). *Iowa assessments*. Orlando, FL: Houghton Mifflin Harcourt.
- U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census (1994). *Geographic Areas Reference Manual*.
- Welch, C., & Dunbar, S. (2012). *Iowa assessments*. Boston, MA: Houghton Mifflin Harcourt.
- Wery, J. J., & Diliberto, J. A. (2017). The effect of a specialized dyslexia font, opendyslexic, on reading rate and accuracy. *Annals of Dyslexia*, 67(2), 114-127. doi: 10.1007/s11881-016-0127-1
- Wilkins, A., Cleave, R., Grayson, N., & Wilson, L. (2009). Typography for children may be inappropriately designed. *Journal of Research in Reading*, 32(4), 402–412. doi: 10.1111/j.1467-9817.2009.01402.x
- Wise, J., C., Sevcik, R., A., Morris, R., D., Lovett, M., W., Wolf, M., Kuhn, M., . . . Schwanenflugel, P. (2010). The relationship between different measures of Oral Reading Fluency and reading comprehension in second-grade students who evidence different Oral Reading Fluency difficulties. *Language, Speech, and Hearing Services in Schools*, 41(3), 340–348. doi: 10.1044/0161-1461(2009/08-0093)

Appendix A

ORF Benchmark Passage Statistics

Appendix A: ORF Benchmark Passage Statistics

Grade	Period	Title	W	S	FKGL	Lexile	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion
1	1	Lucky Day	172	21	1.5	4-500L	95	72	76	62	42
	2	Jack and Jill	190	22	1.2	4-500L	91	99	53	71	49
	3	Our Pond	169	15	1.8	4-500L	99	50	92	93	44
2	1	Church Pears	225	20	2.4	4-500L	76	59	96	59	9
	2	Pay Phones	201	17	2.9	4-500L	65	50	63	89	80
	3	Puppy Love	201	17	2.7	4-500L	86	71	96	94	44
3	1	Sponges	258	24	4.1	5-600L	57	91	97	53	90
	2	Why We Need Water	202	19	3.6	4-500L	97	73	93	98	73
	3	Trees	201	16	2.8	4-500L	67	75	98	98	58
4	1	The Raft	221	18	4.0	6-700L	49	78	98	27	14
	2	Honesty	292	21	4.2	5-600L	99	42	75	95	64
	3	Sunset at the Beach	250	20	5.0	6-700L	38	85	99	12	83

Grade	Period	Title	W	S	FKGL	Lexile	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion
5	1	Breathing	296	24	5.2	8-900L	63	79	81	91	61
	2	Animal Tools	320	23	5.8	9-1000L	42	78	99	79	97
	3	The Sukkah Next Door	286	18	5.9	9-1000L	79	47	99	51	46
6	1	Helen Keller	299	21	6.5	8-900L	72	76	79	68	72
	2	Slots and Monkeys	257	19	6.5	9-1000L	46	77	87	95	92
	3	Changing Bedtime	324	23	6.4	8-900L	74	73	34	23	95
7	1	Coyotes and Wolves	279	20	7.4	10-1100L	25	60	98	83	68
	2	Fizzy Water	361	29	7.4	9-1000L	27	82	85	77	88
	3	Prize Winning Vegetables	370	22	7.8	10-1100L	53	19	87	52	55
8	1	Crows	353	20	8.5	11-1200L	38	55	93	20	53
	2	Government	368	29	8.7	8-900L	11	86	31	28	70
	3	Digital Music Recording	371	25	8.8	9-1000L	25	91	72	59	99

Note: W = word count; S = sentence count; FKGL = Flesch-Kincaid Grade Level; W, S, and FKGL were drawn from Microsoft Word. The last five columns are Coh-Metrix indices ranging from 1 to 99 where higher values indicate more of a characteristic and less text complexity. Higher values of narrativity indicates passages that are more story-like. Higher values of syntactic simplicity indicate passages that use simpler sentence structures more frequently, although sentences containing dialogue can sometimes inflate these values. Higher values of word concreteness indicate passages that include more concrete rather than abstract words. Higher values of referential cohesion indicate more overlap in words and propositions in a passage. Higher values of deep cohesion indicate more frequent connectives in passages.

Appendix B

Maze Benchmark Passage Statistics

Appendix B: Maze Benchmark Passage Statistics

Grade	Period	Title	W	S	FKGL	Lexile	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion
2	1	Working Together	371	38	2.2	500L-600L	94	83	81	67	79
	2	The New Kid	401	39	2.3	500L-600L	95	77	83	49	92
	3	Summer Reading	403	46	2.6	500L-600L	84	88	93	49	45
3	1	Brush Hogging	380	30	4.1	500L-600L	73	51	98	50	32
	2	The Secret Desert	414	35	3.7	500L-600L	90	54	75	65	79
	3	On the Trail	420	34	3.7	600L-700L	87	84	77	40	57
4	1	Working on Cars	410	27	5.0	800L-900L	85	58	97	30	37
	2	Lucie's Snow	457	30	5.6	800L-900L	79	54	99	35	28
	3	The Hill	497	36	5.7	800L-900L	82	47	83	55	40
5	1	The iPhone	486	32	6.9	900L-1000L	57	70	76	21	86
	2	New Slang	507	37	6.9	800L-900L	52	81	20	17	86
	3	Languages	483	40	7.1	800L-900L	48	68	38	93	63

Grade	Period	Title	W	S	FKGL	Lexile	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion
6	1	How to Make a Woodcut	444	26	7.0	1000L-1100L	36	68	74	75	85
	2	Blizzards of the United States	526	32	6.7	1000L-1100L	27	79	91	15	95
	3	Popcorn	512	34	7.8	1000L-1100L	23	78	90	33	96
7	1	Fireflies	453	32	8.1	900L-1000L	26	78	79	14	78
	2	Archimedes	451	30	8.9	900L-1000L	52	54	84	58	40
	3	The Day the Sky Turned Black	501	25	8.2	1000L-1100L	66	51	96	52	75
8	1	Backstrap Weaving	456	26	8.0	1000L-1100L	33	61	84	49	88
	2	The Circus	538	32	8.6	1000L-1100L	42	40	95	71	85
	3	Space Shuttles	533	35	8.8	1000L-1100L	15	67	77	34	76

Note: W = word count, S = sentence count, FKGL = Flesch-Kincaid Grade Level, W, S, and FKGL were drawn from Microsoft Word. The last five columns are Coh-Metrix indices ranging from 1 to 99, where higher values indicate more of a characteristic and less text complexity. Higher values of narrativity indicate passages that are more story-like. Higher values of syntactic simplicity indicate passages that use simpler sentence structures more frequently, although sentences containing dialogue can sometimes inflate these values. Higher values of word concreteness indicate passages that include more concrete rather than abstract words. Higher values of referential cohesion indicate more overlap in words and propositions in a passage. Higher values of deep cohesion indicate more frequent connectives in passages.

Appendix C

DIBELS 8th Edition Pronunciation Guide

Appendix C: DIBELS 8th Edition Pronunciation Guide

Phoneme	Phoneme Example	Phoneme	Phoneme Example
/a/	bad	/b/	b at
/e/	bed	/d/	d ad
/i/	bid	/f/	f at
/o/	cod, law	/g/	g et
/u/	bu d , “a” in a bout	/h/	h ot
/A/	ba i t	/j/	j am, e dge
/E/	bea d	/k/	ca n , k it, pi ck
/I/	tie	/l/	l ap
/O/	boa t	/m/	m an
/oo/	foo d	/n/	n ap
/uu/	goo d	/p/	p en
/ow/	co w	/r/	r at, w rite
/oy/	poi n t, bo y	/s/	si t , ci ty
/ar/	(1 phoneme) ca r	/t/	t ap
/air/	(1 phoneme) cha i r	/v/	v an
/er/	(1 phoneme) he r , bi r d	/w/	w et
/ear/	(1 phoneme) clea r	/y/	y ak
/or/	(1 phoneme) fo r	/z/	z oo
/oor/	(1 phoneme) pou r	/ch/	ch in
		/sh/	sh ed

Phoneme	Phoneme Example	Phoneme	Phoneme Example
		/SH/	meas <u>ure</u> , beig <u>e</u>
		/th/	thin
		/TH/	then
		/ng/	sing

Note: Both voiced and unvoiced forms of 'th' and 'sh' are acceptable for nonsense words containing these digraphs.

Appendix D

Administration and Scoring Fidelity Checklists

Appendix D: Administration and Scoring Fidelity Checklists

Letter Naming Fluency Fidelity Checklist

Pass	Needs practice	
<input type="checkbox"/>	<input type="checkbox"/>	1. Holds clipboard and timer so student cannot see what is recorded.
<input type="checkbox"/>	<input type="checkbox"/>	2. Places the student copy in front of the student.
<input type="checkbox"/>	<input type="checkbox"/>	3. Places the student copy in front of the student.
<input type="checkbox"/>	<input type="checkbox"/>	4. Starts timer after saying “Begin” .
<input type="checkbox"/>	<input type="checkbox"/>	5. Follows along and marks the scoring book as the student responds.
<input type="checkbox"/>	<input type="checkbox"/>	6. Administers acceptable prompts (i.e., hesitation and letter sound) correctly and when appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	7. Applies scoring rules consistently and correctly.
<input type="checkbox"/>	<input type="checkbox"/>	8. Applies the discontinue rule correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	9. At the end of 60 seconds, puts a bracket (]) after the last letter named and says “Stop” .
<input type="checkbox"/>	<input type="checkbox"/>	10. Accurately determines and records the total number of correct letter names in 60 seconds. Score is within 2 points of the expert examiner.

Phonemic Segmentation Fluency Fidelity Checklist

Pass	Needs practice	
<input type="checkbox"/>	<input type="checkbox"/>	1. Holds clipboard and timer so student cannot see what is recorded.
<input type="checkbox"/>	<input type="checkbox"/>	2. Performs standardized directions verbatim, including the correction procedure, if applicable.
<input type="checkbox"/>	<input type="checkbox"/>	3. Starts timer after presenting the first word.
<input type="checkbox"/>	<input type="checkbox"/>	4. Follows along and marks the scoring book as the student responds.
<input type="checkbox"/>	<input type="checkbox"/>	5. As soon as the student is finished saying the sounds in the current word, presents the next word promptly and clearly.
<input type="checkbox"/>	<input type="checkbox"/>	6. Administers acceptable prompts correctly and when appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	7. Applies scoring rules consistently and correctly.
<input type="checkbox"/>	<input type="checkbox"/>	8. Applies the discontinue rule correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	9. Stops at the end of 60 seconds and puts a bracket (]) after the last response.
<input type="checkbox"/>	<input type="checkbox"/>	10. Accurately determines and records the total number of correctly produced phonemes in 60 seconds. Score is within 2 points of the expert examiner.

Nonsense Word Fluency Fidelity Checklist

Pass	Needs practice	
<input type="checkbox"/>	<input type="checkbox"/>	1. Holds clipboard and timer so student cannot see what is recorded.
<input type="checkbox"/>	<input type="checkbox"/>	2. Places student copy in front of the student.
<input type="checkbox"/>	<input type="checkbox"/>	3. Performs standardized directions verbatim, including the correction procedure when appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	4. Starts timer after saying “Begin” .
<input type="checkbox"/>	<input type="checkbox"/>	5. Follows along and marks the scoring book as the student responds.
<input type="checkbox"/>	<input type="checkbox"/>	6. Administers acceptable prompts correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	7. Administers acceptable prompts correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	8. Applies the discontinue rule correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	9. At the end of 60 seconds, puts a bracket (]) after the last sound provided and says “Stop” .
<input type="checkbox"/>	<input type="checkbox"/>	10. Accurately determines and records the correct letter sounds produced and words read correctly within 60 seconds. Score is within 2 points of the expert examiner.

Word Reading Fluency Fidelity Checklist

Pass	Needs practice	
<input type="checkbox"/>	<input type="checkbox"/>	1. Holds clipboard and timer so student cannot see what is recorded.
<input type="checkbox"/>	<input type="checkbox"/>	2. Places student copy in front of the student.
<input type="checkbox"/>	<input type="checkbox"/>	3. Performs standardized directions verbatim.
<input type="checkbox"/>	<input type="checkbox"/>	4. Starts timer when the student says the first word.
<input type="checkbox"/>	<input type="checkbox"/>	5. Follows along and marks the scoring book as the student responds.
<input type="checkbox"/>	<input type="checkbox"/>	6. Administers acceptable prompts correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	7. Applies scoring rules consistently and correctly.
<input type="checkbox"/>	<input type="checkbox"/>	8. Applies the discontinue rule correctly and when appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	9. At the end of 60 seconds, puts a bracket (]) after the last sound provided and says “Stop” .
<input type="checkbox"/>	<input type="checkbox"/>	10. Accurately determines and records the number of words read correctly. Score is within 2 points of the expert examiner.

Oral Reading Fluency Fidelity Checklist

Pass	Needs practice	
<input type="checkbox"/>	<input type="checkbox"/>	1. Holds clipboard and timer so student cannot see what is recorded.
<input type="checkbox"/>	<input type="checkbox"/>	2. Places student copy in front of the student.
<input type="checkbox"/>	<input type="checkbox"/>	3. Performs standardized directions verbatim.
<input type="checkbox"/>	<input type="checkbox"/>	4. Starts timer when the student says the first word.
<input type="checkbox"/>	<input type="checkbox"/>	5. Follows along and marks the scoring book as the student responds.
<input type="checkbox"/>	<input type="checkbox"/>	6. Administers acceptable prompts correctly, if appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	7. Applies scoring rules consistently and correctly.
<input type="checkbox"/>	<input type="checkbox"/>	8. Applies the discontinue rule correctly and when appropriate.
<input type="checkbox"/>	<input type="checkbox"/>	9. At the end of 60 seconds, puts a bracket (]) after the last sound provided and says “Stop” .
<input type="checkbox"/>	<input type="checkbox"/>	10. Accurately determines and records the number of words read correctly. Score is within 2 points of the expert examiner.

Appendix E

Composite Score Calculation Guide

Appendix E: Composite Score Calculation Guide

To calculate the DIBELS 8 composite score, a student must have been administered all available subtests for their grade. Apply the following steps, in order:

1. For each subtest raw score, multiply the student's raw score by the Weight listed in the table on the next page, rounding the result to the 100ths place.
2. Sum the resulting weighted scores across all applicable subtests.
3. From that sum, subtract the Mean for the appropriate grade from the table on the next page.
4. Divide the result by the standard deviation (SD) for the appropriate grade in the table on the next page.
5. Multiply the result by 40 and round to the ones place.
6. Add the scaling Constant corresponding to the grade and season in which the student was tested from the table on the next page. The result is the composite score.

Note that ORF Accuracy should be represented in these calculations as a proportion of words correct (e.g., .99), rather than percent correct (e.g., 99).

An example calculation is provided after the table on the next page.

Grade	Subtest score	Weight	Mean	SD	Fall constant	Winter constant	Spring constant
Kindergarten	LNF	8.86	1346	682	360	400	440
	PSF	4.13					
	NWF-CLS	14.93					
	NWF-WR	3.56					
	C						
	WRF	5.62					
First	LNF	10.72	4706	2468	360	400	440
	PSF	2.13					
	NWF-CLS	23.13					
	NWF-WR	7.79					
	C						
	WRF	13.51					
	ORF-WR	25.36					
	C						
	ORF-ACC	0.25					
Second	NWF	32.74	7100	3617	360	400	440
	NWF	10.95					
	WRF	21.26					
	ORF-WR	35.36					
	C						
	ORF-ACC	0.15					
	MAZE	4.28					

Grade	Subtest score	Weight	Mean	SD	Fall constant	Winter constant	Spring constant
Third	NWF	40.02	10220	3797	360	400	440
	NWF	11.80					
	WRF	19.83					
	ORF-WR	39.42					
	C						
	ORF-ACC	0.09					
	MAZE	4.79					
Fourth	ORF-WR	36.42	3907	1357	360	400	440
	C						
	ORF-ACC	0.06					
	MAZE	6.29					
Fifth	ORF-WR	31.12	3889	1066	360	400	440
	C						
	ORF-ACC	0.03					
	MAZE	4.58					
Sixth	ORF-WR	40.71	5247	1796	360	400	440
	C						
	ORF-ACC	0.05					
	MAZE	5.03					
Seventh	ORF-WR	40.55	5818	1615	360	400	440
	C						
	ORF-ACC	0.06					
	MAZE	7.34					

Grade	Subtest score	Weight	Mean	SD	Fall constant	Winter constant	Spring constant
Eighth	ORF-WR	37.69	5125	133	360	400	440
	C						
	ORF-ACC	0.03					
	MAZE	6.75					

Example

For a second grade student with fall DIBELS 8 scores of 152 for NWF Correct Letter Sounds (CLS), 48 for NWF Words Read Correctly (WRC), 45 for WRF, 88 for ORF Words Read Correctly (WRC), 99% ORF Accuracy, and 11.5 for Maze Adjusted, we would calculate this student's composite score as follows.

Step 1: Multiply each subtest raw score by the corresponding weight listed in the table.

Subtest score	Raw score	Weight	Weight score
NWF-CLS	152.00	* 32.74	= 4976.48
NWF-WRC	48.00	* 10.95	= 525.60
WRF	45.00	* 21.26	= 956.70
ORF-WRC	88.00	* 35.36	= 3111.68
ORF-ACC	0.99	* 0.15	= 0.15
Maze	11.50	* 0.15	= 49.22

Step 2: Sum the resulting weighted scores across all applicable subtests:

$$4976.48 + 525.6 + 956.7 + 3111.68 + 0.1485 + 49.22 = 9619.83$$

Step 3: Subtract from that sum the mean of the weighted scores for the appropriate grade:

$$9619.83 - 7100 = 2519.82$$

Step 4: Divide that value by the standard deviation for the appropriate grade:

$$2519.82 / 3617 = 0.697$$

Step 5: Multiply that score by 40 and round to the ones place:

$$0.697 * 40 = 28$$

Step 6: Add the scaling constant corresponding to the season in which the student was tested to obtain the final composite score:

$$28 + 360 = 388$$

