Assessing the Utility of DIBELS® Deep:

Phonemic Awareness and Word Reading and Decoding

Kelly A. Powell-Smith

Ruth A. Kaminski

Dynamic Measurement Group

Technical Report No. 15

Recommended Citation:

Powell-Smith, K., & Kaminski, R. A. (2013). *Assessing the Utility of DIBELS® Deep: Phonemic Awareness and Word Reading and Decoding*. (Technical Report No. 15) Eugene, OR: Dynamic Measurement Group.

Author Note

The authors thank the faculty, staff, students, and parents of the participating schools for their effort and support during the course of this study. Correspondence regarding this manuscript should be addressed to Kelly A. Powell-Smith, Dynamic Measurement Group, 859 Willamette Street, Suite 320, Eugene, OR 97401; Email: kpowellsmith@dibels.org.

Abstract

This technical report presents the results of a study conducted as part of the validation of DIBELS® Deep: Phonemic Awareness and DIBELS Deep: Word Reading and Decoding[1]. The DIBELS Deep assessment consists of brief diagnostic reading measures linked to DIBELS®. The purpose of this study was to examine the accuracy of the scope and sequence, determine the relation of DIBELS Deep tasks to each other, and determine the relation between DIBELS Deep tasks and DIBELS benchmark assessments. Therefore, this study was designed to be primarily descriptive.

Approximately 245 students in first through fourth grades (45-65 students per grade) were assessed during the 2006-2007 school year using a research version of DIBELS Deep. Each student was assessed with DIBELS Deep materials at his or her grade level as well as with DIBELS Deep materials above or below grade level according to an assessment schedule provided to each research site. Results indicated that the scope and sequence of skills was generally accurate for each grade level (i.e., measures identified for each grade level). However, section-level data analyses suggested that some sections within measures could be re-ordered. Correlations between DIBELS Deep measures of the same skill construct were moderate-strong to near perfect while those correlations between DIBELS Deep measures of different skill constructs were moderate. Correlations between DIBELS Deep and DIBELS measures of the same skill were moderate to strong.

Feedback on the utility of the measures was obtained via questionnaires given to those personnel who administered and scored the measures (i.e., assessors) as well as

[1] DIBELS Deep Word Reading and Decoding was referred to as DIBELS Deep Alphabetic Principle in the early phases of development.

participating students' teachers. Summarized questionnaire data indicate that both teachers and assessors agreed overall that the measures were useful. Implications for future research and practice are discussed.

Assessing the Utility of DIBELS® Deep: Phonemic Awareness and Word Reading and Decoding

General outcome measures, such as curriculum-based measurement (CBM) and Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are widely used in education for universal screening. General outcome measures have numerous advantages over other types of assessment for screening: they can be administered quickly, with high levels of reliability; they have been shown to accurately identify student need; they are highly predictive of future student performance; and they are often useful for determining response to intervention (RtI) (Batsche et al., 2005).

DIBELS are a set of brief measures of early reading that have been found to be reliable and valid for assessing the acquisition of early literacy skills of children who are learning to read in English (Dynamic Measurement Group, 2008). The DIBELS map on to the five critical areas of early reading instruction as described in recent research reports (e.g., National Reading Panel, 2000) and include indicators from the following skill areas: phonemic awareness, phonics, vocabulary and oral language development, accuracy and fluency with connected text, and comprehension.

DIBELS data are collected routinely by many schools as part of ongoing school improvement efforts in reading. However, diagnostic information with regard to developing specific instructional interventions is not provided explicitly by the current DIBELS measures. To obtain such information, teachers and other educators must develop their own assessments (e.g., teacher-made tests), use lengthy and/or costly diagnostic assessments, or conduct detailed error analysis procedures.

Best practices in the use of DIBELS suggest that the measures be used within a prevention-oriented decision-making model referred to as the Outcomes-Driven Model (Kaminski, Cummings, Powell-Smith, & Good, 2008). Within this model, students are first screened and those who may need instructional support are identified. Second, the need for support is validated. Third, an instructional support plan is developed including the content and goals of instruction. Fourth, progress monitoring occurs to determine whether the support provided is resulting in positive outcomes for the student. At the fifth and final step, outcomes at the child and system level are reviewed. Within the Outcomes-Driven Model, DIBELS® Deep fits within the planning instructional support step.

Purpose and Design Characteristics

The purpose of DIBELS Deep is to provide teachers with brief diagnostic assessments that are cost- and time-efficient from which they may garner specific information for developing instruction that corresponds to the five critical areas of reading instruction described by the National Reading Panel (2000): phonological awareness, phonics, fluency (with text), comprehension and vocabulary. Given this purpose, four primary design characteristics guided the development of DIBELS Deep. These design characteristics were that the measures should be: (1) linked to the DIBELS measures; (2) brief (i.e., take approximately 15 minutes to administer), yet cover the range of skills represented within a skill domain (e.g., phonemic awareness); and (3) used within a prevention-oriented decision-making model.

The fourth and final design characteristic is that the skills assessed by the DIBELS Deep measures and how they are sequenced should correspond to the progression of

skills across grades as well as recognized sequences of instruction (c.f., Carnine, Silbert, Kame'enui, Tarver, & Jungjohann, 2006; Jennings, Caldwell, & Lerner, 2006; National Research Council, 1998; Nippold, 2007; Simmons & Kame'enui, 1999; Wagner, Muse, & Tannenbaum, 2007). While the intended use of these measures is primarily for differentiating instruction for students identified as at-risk for reading difficulties in the elementary grades, the DIBELS Deep measures were also designed to be user-friendly in terms of their time- and cost-effectiveness.

Purpose of the Study

Prior to this Phase 1 validation study, a pilot study was conducted. A technical report (Powell-Smith & Kaminski, 2008) describes the results of that pilot as well as the initial development of the measures. The current study was focused on examining new measures designed to provide more in-depth diagnostic information for teachers in the domains of phonemic awareness and word reading and decoding (including phonics as well as word and sentence reading skills). The research questions posed were primarily related to utility (e.g., feasibility of the revised measure, use for determining instructional content, etc.). The relation between DIBELS benchmark data and DIBELS Deep results was examined as well to determine the validity of DIBELS Deep. The specific research questions examined were:

1. What is the distribution and frequency of DIBELS Deep measures given at each grade level?
2. What is the relationship between the various DIBELS Deep measures?
3. What is the relationship between performance on DIBELS Deep and the DIBELS 6th Edition benchmark measures?

4. What is the relation between sections within each DIBELS Deep measure?
5. Are the items and sections sequenced appropriately?
6. To what extent do teachers find the measures useful?
7. To what extent are assessors satisfied with the measures?

Method

Participants

Schools. Eleven schools in four states participated in this study. The four states were from the West North Central part of the Midwest, the East North Central part of the Midwest, and the Pacific West according the U.S. Census Bureau. Nine of the 11 schools participated in the study during the fall and winter of 2006-2007. Two schools participated in winter and spring of 2007.

Demographic data on each of the schools is found in Table 1. Participating schools represented rural areas as well as midsize cities and larger suburban areas. The schools ranged in the size of the student population served from 182 to 674 students and the grade levels served ranged from pre-K to third grade to sixth grade. Participating schools also had a range of 11% - 53% of students participating in the federal free/reduced price lunch program. All but one of the participating schools was designated as a Title I school. Finally, while the student population across these schools was primarily white, two of the schools had greater ethnic diversity with at least 40% of their student population being non-white.

Students. Student participants included 45-65 students in each grade level (kindergarten to fourth grade). Sites were instructed to select a random sample, stratified across instructional recommendation levels (i.e., benchmark, strategic,

intensive) according to their fall DIBELS benchmark data. In the site with five schools, six students per grade level, stratified across instructional recommendation levels, were selected to participate, resulting in a total of 30 students per grade across the five schools. The final sample consisted of 245 students. Data on the instructional recommendation status for participating students at each grade level for both fall and winter are displayed in Table 2. As shown in the table, the fall sample of students was distributed relatively well across instructional recommendations. Slightly more students were at benchmark, but overall the pattern is consistent with our request for a random stratified sample. In the winter, generally more kindergarten students achieved benchmark level instructional recommendations and few students had intensive instructional recommendations. For second and third grade, the sample had more students at benchmark compared to strategic or intensive levels. This shift in students across instructional recommendation levels from fall to winter may have been a result of the change in student population during the winter testing (e.g., two schools were added to the study). Alternatively, the shift could mean that the schools provided instructional interventions changing the middle-of-year outcomes for these students.

DIBELS scores for participating students by grade level are displayed in Table 3. On average, kindergarten students appear to be performing close to DIBELS 6th Edition benchmark goals on the Initial Sound Fluency (ISF) and Phoneme Segmentation Fluency (PSF) measures. First grade students' mean scores are higher than the winter benchmark goal for DIBELS Oral Reading Fluency (DORF). On average, the second-grade students earned scores on DORF that indicate they are on track to meet the DIBELS 6th edition spring benchmark goal of 90 words read correctly. The mean DORF

score for third-grade students is close to the benchmark goal in the fall, but just below the goal in the winter. The pattern of scores for fourth grade is the opposite of the pattern for third grade, with the mean score on DORF below the benchmark goal in the fall, but above it in the winter.

Teachers and assessors. A total of 31 teachers and 16 assessors completed the usability questionnaires. Between three and 11 teachers at each grade level, K - 4, completed the questionnaires. One of the fourth-grade teachers also was a special education teacher. Assessors who participated in the study served in a variety of roles within their respective school districts, including school psychologist, Title I teacher, and educational assistant.

Measures

The measures used in this study included the DIBELS 6th edition measures appropriate for each grade level's fall, winter, or spring benchmark assessments (depending on the site) as well as the experimental DIBELS Deep Phonemic Awareness and Word Reading and Decoding measures. In what follows, the DIBELS measures are described first, followed by the DIBELS Deep measures.

Initial Sound Fluency (ISF). ISF (Good, Laimon, Kaminski, & Smith, 2003) is a measure of phonemic awareness that assesses a child's skill in recognizing and producing the initial sound or group of sounds in orally presented words. The assessor presents four pictures to the child, names each picture, and then asks the child to identify (i.e., point to or say) the picture that begins with the sound produced orally by the assessor. For example, the assessor says, "This is sink, cat, gloves, and hat. Which picture begins with /s/?" and the students points to the correct picture. The student is

also asked to vocalize the beginning sound of an orally presented word that matches one of the given pictures. The assessor calculates the amount of time taken to identify/produce the correct sound and converts the score into the number of onsets correct in a minute.

ISF is a version of the Onset Recognition Fluency (OnRF) measure (Laimon, 1994) incorporating minimal revisions. Alternate form reliability of OnRF is .72; by repeating the assessment four times, the resulting average is estimated to have a reliability of .91 (Nunnally, 1978). The concurrent criterion-related validity of OnRF with the Woodcock-Johnson Psycho-Educational Battery Readiness Cluster Score is .36.

Letter Naming Fluency (LNF). LNF (Kaminski & Good, 2003) is used as an indicator of risk relating to future literacy development. Students are presented with a page of upper- and lower-case letters arranged in a random order and are asked to name as many letters as they can. Students are told that if they do not know a letter, they will be told the letter. The student is allowed one minute to produce as many letter names as he/she can, and the final score is the number of letters named correctly in one minute.

The one-month alternate form reliability of LNF is .88 in kindergarten. The median criterion-related validity of LNF with the Woodcock-Johnson Psycho-Educational Battery Revised Readiness Cluster standard score is .70 in kindergarten (Good, Kaminski, Shinn, Bratten, Laimon, et al., 2004). The predictive validity of kindergarten LNF with first grade Woodcock Johnson Psycho-Educational Battery-Revised Reading Cluster standard score is .65, and .71 with first grade Curriculum-Based Measurement (CBM) oral reading fluency (ORF) (Good et al., 2004).

Phoneme Segmentation Fluency (PSF). PSF (Good, Kaminski, & Smith, 2003) is a test of phonological awareness. PSF assesses a student's ability to fluently segment three-, four- and five-phoneme words into their individual phonemes. The PSF measure has been found to be a good predictor of later reading achievement (Kaminski & Good, 1996). The assessor administers the PSF task by orally presenting words containing three to five phonemes. It requires the student to verbally produce sound segments for each word, preferably at the individual phoneme level. For example, the assessor says, "sat," and the student says, "/s/ /a/ /t/" to receive the maximum score of three points for the word. After the student responds, the assessor presents the next word, and the number of correct sound segments produced in one minute determines the final score.

The two-week, alternate-form reliability for the PSF measure is .88 (Kaminski & Good, 1996), and the one-month, alternate-form reliability is .79 in May of kindergarten (Good et al., 2004). Concurrent, criterion validity of PSF is .54 with the Woodcock-Johnson Psycho-Educational Battery Readiness Cluster score in spring of kindergarten (Good et al., 2004). The predictive validity of spring-of-kindergarten PSF with (a) winter-of-first-grade DIBELS NWF is .62, (b) spring-of-first-grade Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score is .68, and (c) spring-of-first-grade CBM ORF is .62 (Good et al., 2004).

Nonsense Word Fluency (NWF). NWF (Good & Kaminski, 2003) is a measure of the alphabetic principle, including both letter-sound correspondence and the ability to blend letters into words in which the letters represent their most common sounds. The student is presented an 8.5- by 11-inch sheet of paper with randomly ordered VC and CVC nonsense words (e.g., sig, rav, ov) and asked to verbally produce the individual

sound of each letter or verbally produce, or read, the whole nonsense word. For example, if the stimulus word is “sig,” the student could say /s/ /i/ /g/ or say the word /sig/ to obtain a total of three letter sounds correct. The student is allowed one minute to produce as many letter-sounds as he/she can, and the final score is the number of letter-sounds produced correctly in one minute. Because the measure is fluency based, students receive a higher score if they are phonologically recoding the word and receive a lower score if they are providing letter sounds in isolation.

The one-month, alternate-form reliability for NWF in January of first grade is .83 (Good et al., 2004). The concurrent criterion-validity of DIBELS NWF with the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness Cluster score is .36 in January and .59 in February of first grade (Good et al., 2004). The predictive validity of DIBELS NWF in January of first grade with (a) CBM ORF in May of first grade is .82, (b) CBM ORF in May of second grade is .60, (c) Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score is .66 (Good et al., 2004).

Oral Reading Fluency (ORF). The DIBELS ORF (DORF; Good, Kaminski, & Dill, 2003) measure builds on the work of Stan Deno and colleagues at the University of Minnesota Institute for Research on Learning Disabilities, where Curriculum Based Measurement Reading procedures were developed (Deno, 1985; Shinn, 1989). However, DORF passages are distinguished from other CBM Reading procedures primarily by the set of generic passages that have been developed for benchmark and progress monitoring assessment. Student performance is measured by having students read novel connected text. The student is instructed to read from the passage aloud for one minute while the assessor follows along, marking the errors on the assessor copy.

Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as correct. The number of words read correct within the one-minute time frame is the score.

A series of studies examined the technical adequacy of CBM ORF procedures in general. Test-retest reliabilities for elementary-aged students ranged from .92 to .97; alternate-form reliability of different reading passages drawn from the same level ranged from .89 to .94 (Tindal, Marston, & Deno, 1983). Criterion-related validity data from eight separate studies in the 1980s reported coefficients ranging from .52 to .91 (Good & Jefferson, 1998).

DIBELS Deep Phonemic Awareness 1 and 2. A range of Phonemic Awareness (PA) skills are assessed on these two forms generally beginning with easier skills and becoming progressively more challenging. Deep PA Form 1 (PA1) samples the following skills: blending word parts in compound words, segmenting compound words into their parts, blending syllables, segmenting syllables, blending onset-rime, matching rimes, segmenting onset-rime, saying rhyming words, and recognizing rhyming words. Deep PA Form 2 (PA2) samples the following skills: blending two- and three-phoneme words, recognizing and producing initial sounds, recognizing and producing final sounds, segmenting two-three phoneme words and segmenting a three-phoneme words with blends. Approximately five items per skill area are on each form. Discontinue rules are included so that students are not tested on skills that may be too difficult or frustrating for them. Scores are totaled for each individual section and overall.

DIBELS Deep Word Reading and Decoding Quick Screen. This measure is designed to provide information helpful in determining which DIBELS Deep Word

Reading and Decoding (WRD) forms should be used for further assessment. The measure contains one or two items from across the scope and sequence of phonics skills in grades K-3 (e.g., reading VC words beginning with continuous sounds up through blending words with irregular vowel teams). There are two ways to use this measure:

(1) Use the discontinue rule. If the child misses five consecutive items then discontinue the quick screen. The appropriate entry point for DIBELS Deep Word Reading and Decoding is determined by finding the item number where the discontinue rule was met in a table and finding the corresponding WRD Deep form number to administer.

(2) Administer the entire WRD Deep Quick Screen (WRD QS). Examine the student's response patterns on the WRD QS using a table. For each skill with an incorrect response, administer the corresponding WRD Deep form section(s).

DIBELS® Deep Word Reading and Decoding Forms 1-5. These forms are designed to assess the range of phonics skills children are expected to learn in grades K-3. Deep WRD Form 1 (WRD1) covers kindergarten skills (e.g., letter-sound correspondence, blending VC and CVC words like "at" and "dog"). Deep WRD Forms 2 and 3 (WRD2 and WRD3) cover first grade skills (e.g., blending CVCC, CCVC, CCVCC words, blending words with consonant digraphs, blending one-syllable words with vowel diagraphs and diphthongs, etc.). Deep WRD Form 4 (WRD4) covers second grade skills (e.g., blending two-syllable words with r-controlled vowels, blending words with inflectional endings, blending multisyllabic words, etc.). Deep WRD Form 5 (WRD5) covers third grade skills (e.g., blending two-syllable words with diphthongs, blending

words with irregular vowel teams, blending words with consonant trigraphs). Approximately 5-10 items per skill area are on each form. Each form also contains a section for reading grade level high-frequency (sight) words and short sentences composed of words covered on the form. Discontinue rules are included so that students are not tested on skills that may be too difficult or frustrating for them. Scores are totaled for each individual section and overall.

Teacher questionnaire. This 11-item questionnaire was developed in-house for the purposes of evaluating new DIBELS-related measures and products. The teacher questionnaire includes statements such as, “The measures adequately covered the reading skills in the grade level I teach,” “The measures were a good way to assess students’ reading strengths and weaknesses,” and “Overall, the measures would be beneficial for planning reading instruction.” Teachers are asked to rate each statement on a six point Likert-type scale ranging from strongly disagree to strongly agree.

Assessor questionnaire. This 10-item questionnaire was developed in-house for the purposes of evaluating new DIBELS-related measures and products. The assessor questionnaire includes statements like, “The administration and scoring rules were easy to follow,” “I believe that the number, type, and sequence of practice items were sufficient to ensure that the students understood the task,” and “Overall, the measures would be beneficial for planning reading instruction.” Assessors are asked to rate each statement on a six point Likert-type scale ranging from strongly disagree to strongly agree.

Procedures

Participant recruitment and selection. Research sites were recruited via email invitation, telephone, and for one school site, in person. Invitations to participate in research were sent to sites that had previously indicated an interest in partnering with Dynamic Measurement Group (DMG) in research. All sites recruited were already engaged in DIBELS data collection. Prior to data collection for this study, IRB approval as well as approval from the school districts and schools was obtained. A project description was provided to all participating schools, teachers, and parents of student participants. Teachers and assessors who participated in the study were recruited by a local coordinator at each site. Students whose teachers volunteered to participate were eligible to participate and a project description and consent form/information letter was sent home. All students in general education classrooms receiving English language reading instruction were eligible to participate in this study, including students with disabilities and students who were English language learners, provided they had the response capabilities to participate. Teachers who volunteered to participate were invited to complete the teacher version of the usability questionnaire.

Assessor training. Each of the assessors in the study was trained to administer and score DIBELS Deep PA and WRD measures during a single 90-minute session. During the training, administration and scoring procedures were reviewed as well as the logistics of the study. Time also was allocated for the assessors to ask questions. Most of these trainings occurred via telephone conference, but one of the trainings was conducted onsite.

Data collection. All data collection occurred during the 2006-2007 school year. The data collection schedules for the DIBELS Deep PA and WRD measures are displayed in Figures 1 and 2. All but two schools collected DIBELS Deep data during the fall and winter. The other two schools collected DIBELS Deep data in the winter and spring. DIBELS benchmark assessment data were collected according to each sites' DIBELS benchmark data collection schedule.

Trained assessors completed all of the testing at the school sites. All DIBELS Deep PA and WRD measures were administered individually and were not timed. No student names were recorded on any of the assessment protocols. Student participants were administered the DIBELS Deep measures appropriate for their grade level or lower, depending on student skill according to the data collection schedule and using the decision rules within the measures (see Appendix A). Some latitude was also given for testing students in materials above their grade level if deemed appropriate at the time of testing if the student had not met the discontinue rules for the measures. Finally, teacher and assessor questionnaires were completed after the second round of Deep data collection (e.g., winter or spring depending on the site).

All data were de-identified prior to being sent to DMG. DIBELS Deep assessment score sheets were sent via US mail along with a computer disk containing exported DIBELS 6th Edition benchmark data from the University of Oregon's data system (or, for two sites, an excel spreadsheet created from their own data system). DIBELS Deep data were matched to DIBELS benchmark data via student ID numbers. Finally, in addition to teacher and assessor questionnaire data, the principal investigator kept

anecdotal data (i.e., email and telephone feedback from personnel at the research sites) in an electronic file throughout the project.

Data analysis. Trained DMG data entry personnel entered all DIBELS Deep data into spreadsheets. Section-level data as well as item-level data were entered into separate databases. Reliability checks were conducted on all data entry. Once the DIBELS Deep data entry process was complete, these data were merged with the DIBELS 6th Edition benchmark data from the University of Oregon data system exports provided by each site. Upon completion of the data merging process, data analysis commenced.

Data analysis primarily involved calculating descriptive statistics and correlations. For each measure and for sections within each measure, descriptive statistics were calculated (e.g., mean, standard deviation). Furthermore, the percentage of students earning a correct score on each item within each section was examined and outliers were determined. An item was considered an outlier if the mean percentage of students earning a correct score on that item was more than two standard deviations above or below the mean for the section. Descriptive statistics also were calculated for the items on both teacher and assessor questionnaires. Finally, correlations within DIBELS Deep measures and between DIBELS Deep measures and DIBELS benchmark measures were calculated.

Results

The results of this study are presented for each research question. Only the fall and winter data are reported here due to the small sample size in the spring data set.

Distribution and Frequency of DIBELS Deep Measures

The first research question addressed the frequency and distribution of DIBELS Deep measures given at each grade. This research question was examined by reviewing the distribution of DIBELS Deep measures given at each grade level for fall and winter test administrations. The number of students at each grade level who were given each DIBELS Deep measure in the fall and winter are reported in Table 4. In general, the measures given to students were at the target grade level. Some exceptions to this pattern were noted. In each case, benchmark data examined for those students provided some explanation as to why those students may have been tested on portions of Deep not targeted for their grade level. For example, four kindergarten students in the winter were given WRD2 (a beginning first-grade level measure). Examination of these four students' DIBELS benchmark data indicated that they were performing well above the benchmark for PSF for that time period (18 correct sound segments) with a mean score of 46 sounds segmented correctly. These four students also earned NWF scores above the benchmark (13 correct letter sounds; CLS) with a mean score of 26 CLS. Another exception observed was that WRD2 was given to six fourth-grade students in the fall. These six fourth-grade students' mean DORF score was 67 words read correct (WRC), well below the benchmark (93 WRC) for that time period.

Relationship Between DIBELS Deep Measures

The second research question addressed the relationship between the various DIBELS Deep measures. This research question was examined by reviewing descriptive statistics as well as correlations between DIBELS Deep measures.

Descriptive statistics for the fall test administration of the DIBELS Deep measures for each grade level are reported in Table 5. When examining the data in this table, it is important to keep in mind that the results are most interpretable for measures generally within the target grade and the immediately adjacent grade where the sample size is greater than 10 students. For example, kindergarten student performance on PA2 can be compared to first-grade student performance on PA2. Likewise, we can compare kindergarten student performance on WRD1 with first-grade student performance on WRD1. In making these comparison, one can see higher mean scores for first-grade students on these measure than kindergarten students. However, it would not be appropriate to conclude that fourth-grade students do not earn scores as high as first-grade students on WRD2 based upon the results in this table. Such a comparison is not appropriate because the sample of fourth-grade students is very small (n = 6) and, as stated previously, these students were low performers based on DIBELS benchmark assessment data. In reviewing the data in this table, we also see a higher mean score for first-grade students on WRD1 than on WRD2 and a higher mean score for third-grade students on WRD4 than WRD5. Finally, the mean for fourth-grade students on WRD5 is higher than the mean for third-grade students on WRD5. The patterns in these data for kindergarten, first-, third-, and fourth-grade results are consistent with the notion that the measures are of increasing difficulty. Inconsistent with this general pattern, the mean score for second-grade students was higher on WRD4 than it was for WRD3.

Descriptive statistics for the winter administration of DIBELS Deep are reported in Table 6. Once again, these results are best interpreted for measures at the target grade level and the immediately adjacent grade and when the sample size is greater than 10

students. These results show that the mean for first-grade students on WRD1 was nearly twice that of kindergarten students for that measure. However, the sample of first-grade students was quite small and likely comprised of students who were struggling more in first grade. Despite this, these results indicate that the WRD1 measure is easier for first-grade students than for kindergarteners. The data for first-grade students across WRD2, WRD3, and WRD4 indicate a mean score for WRD3 and WRD4 that is similar, but a mean score for WRD2 that is approximately twice the mean of either WRD3 or WRD4. When examining WRD4 mean scores across first- and second-grade students, we see a much higher mean score and slightly less variability for second-grade students. Also, the mean score for second-grade students on WRD4 is higher than the mean score of second-grade students on WRD5. Finally, the mean for second-grade students on WRD5 is much lower than the mean for third-grade students on that measure. Overall, the results displayed in Table 6 support that the measures are arranged in increasing difficulty.

Correlations between DIBELS Deep measures for the fall and winter test administrations are reported in Tables 7 and 8 respectively. Both sample size and grade level are noted in the table to aid in interpretation. Correlations are based on participants with pairwise complete data and are not reported where the sample size dropped below 20. All the correlations were statistically significant. In the fall, each of these correlations for grade-appropriate measures ranges between .67 and .89, suggesting moderate to strong relationships between the measures. In the winter, the correlations between measures ranged from .58 to .91, again suggesting moderate to strong relationships between the measures. As shown in both tables, the correlations

within a skill area or construct (i.e., between different WRD measures) are stronger than those between constructs (i.e., between PA and WRD measures).

Relation Between DIBELS Deep and DIBELS 6th Edition Measures

To examine the third research question, regarding the relationship between performance on DIBELS Deep and the DIBELS 6th Edition benchmark measures, correlations between both sets of measures were examined. These data are reported in Table 9 by time of year, grade, and measure. Because not all DIBELS measures are given at all grade levels (e.g., PSF is not administered in third grade), there are empty cells in the table. In addition, correlations are based on participants with pairwise complete data for concurrently administered measures only (e.g., fall PA2 and fall PSF, winter WRD1 and winter NWF, etc.). In examining these data, we noted that correlations were low overall for PA1. However, the correlations for PA2 are strong, in particular with the more reliable DIBELS PA measure, PSF. Most of the Deep WRD measures were very strongly correlated with NWF and DORF.

Relation Between Sections within DIBELS Deep Measures

In addition to examining the correlations of DIBELS Deep Measures with DIBELS 6th edition measures, we also examined correlations of DIBELS Deep sections with each of the other sections within each DIBELS Deep measure by grade level and time of administration. These results are provided in Tables 10 - 26. The correlations reported in these tables vary widely. Correlations are not reported for sample sizes below 20.

DIBELS Deep Phonemic Awareness. The correlations between sections on PA1 for fall of kindergarten range from .04 (small) to .69 (moderate-strong), with most

correlations falling in the moderate range. The lowest correlations were obtained for tasks involving onset-rime or rhyming, while the highest were obtained between blending tasks or blending and segmenting tasks (see Table 10).

Correlations for PA 2 (see Table 11) in our kindergarten sample also varied widely in addition to varying depending on the timing of administration (i.e., fall or winter). The correlations ranged from small/trivial (.004) to nearly perfect (.90), but most frequently were moderate in strength. At both time points, the lowest correlation was obtained between production of initial sounds and segmenting two-phoneme words (*r* = .03 & .004, respectively for fall and winter). The highest correlations were obtained between blending two-phoneme and segmenting three-phoneme words (*r* = .72 & .90, respectively for fall and winter), as well as between segmenting two-phoneme and segmenting three-phoneme words in the fall (*r* = .72) and winter (*r* = .82).

Correlations for PA2 (see Table 12) for our first-grade sample in the fall ranged from small (.002) to .74 (strong), with most correlations being small. The lowest correlation was between production of initial sounds and matching final sounds, while the highest correlation was obtained between segmenting three-phoneme words and producing initial sounds.

DIBELS Deep Word Reading and Decoding. The correlations for fall and winter of kindergarten for WRD1 are reported in Table 13. In the fall, correlations range from .51, indicating a moderate relation between letter-sound correspondences and vowel-consonant/consonant-vowel-consonant nonsense words beginning with continuous sounds, to .85, indicating a strong relation between real and nonsense vowel-consonant/consonant-vowel-consonant words beginning with continuous sounds. A

wider range of correlations is observed in the winter data. Correlations for winter range from .19 (small) for the relation between consonant-vowel-consonant nonsense words beginning with stop sounds and pre-primer high-frequency words, to .89 (strong) for vowel-consonant/consonant-vowel-consonant nonsense words beginning with continuous sounds and consonant-vowel-consonant nonsense words beginning with stop sounds. Nearly all correlations were moderate-strong to strong or strong across both time periods.

First-grade correlations for WRD1 are reported in Table 14 for fall administration only. These correlations vary widely and range from .21 (small) to .81 (strong), with most being moderate-strong. The lowest correlation was for letter-sound correspondences with pre-primer high-frequency words and the highest correlation was for the relation between consonant-vowel-consonant real words beginning with stop sounds and sentence reading.

Correlations for WRD2 are reported for first-grade students in Table 15 (fall data) and Table 16 (winter data). In the fall, most correlations were in the moderate-strong to strong range. The lowest correlation was .16 (small) and was between vowel-consonant-consonant real words beginning with continuous sounds and words with "Y" as a vowel. The highest correlation was .89 (strong) and was found between consonant-consonant-vowel-consonant real words (both initial sounds continuous) and consonant-consonant-vowel-consonant real words (one initial sound was a stop sound). The winter data showed a similar range (from small to strong correlations) with the lowest correlation (r = .13) found between nonsense words with double final consonants and consonant-consonant-vowel-consonant-consonant real words. The highest correlation (r

= .85) was between consonant-vowel-consonant-consonant real words (with stop sounds) and consonant-vowel-consonant-consonant nonsense words (with stop sounds). Most of the correlations reported in Table 16 fall in the moderate-strong range.

The correlations between sections on WRD3 for winter of first grade range from .09 (small) to .87 (moderate-strong), with most correlations falling in the moderate to moderate-strong range. The lowest correlation was between one-syllable nonsense words with r-controlled vowels (e.g., purt) and nonsense one-syllable words ending with an "e" (e.g., vete), while the highest correlation was obtained between real words with consonant digraphs and nonsense words with consonant digraphs (see Table 17). Correlations are not reported for several comparisons due to the drop in sample size as the measure progressed (i. e., due to the discontinue rule being met).

The correlations between sections on WRD3 for second-grade students in the fall are displayed in Table 18. More correlations for WRD3 are reported for second-grade students than there were for first-grade students (Table 17), meaning fewer students in second grade discontinued sections on this form than first-grade students. This result supports the notion that WRD3 contains items typically covered in the latter half of first grade. Most correlations between sections of WRD3 for second-grade students for fall data are in the moderate-strong range. However, the correlations vary widely from a very small correlation of .03 between nonsense words with consonant digraphs and one-syllable words with vowel diphthongs to a strong correlation of .89 between one-syllable words with vowel digraphs and first-grade level high-frequency words.

Correlations for the WRD4 sections administered to first-grade students are shown in Table 19 for the winter administration. For many of these correlations, the sample

size is about half the total first-grade sample. Once again, these correlations vary widely and range from .14 (small) to .92 (near perfect), with most being strong or moderate-strong. The lowest correlation was between two-syllable nonsense words with r-controlled vowels and words with contractions as well as two-syllable words with short- and long-vowel patterns and inflectional endings and words with contractions. The highest correlation was between one-syllable words with vowel digraphs and second-grade level high-frequency words.

Correlations for the second-grade student sample administered WRD4 in the fall are in Table 20. Most of these correlations are moderate-strong, but vary widely. The correlations range from .00 (trivial/no relation) to .93 (near perfect). Like the first-grade results, the lowest correlation was found between two-syllable nonsense words with r-controlled vowels and words with possessives. The highest correlation was found between one-syllable words with vowel digraphs and second-grade level high-frequency words.

Correlations for the winter administration of WRD4 to second-grade students are shown in Table 21. Similar to the fall data, most of these correlations are moderate-strong and again, vary widely. The correlations range from .01 (small) to .89 (strong). The lowest correlation was found between words with contractions and words with possessives. The highest correlation was found between one-syllable words with vowel digraphs and words with medial double consonants.

Some third-grade students also were administered WRD4 in the fall. The correlations between WRD4 sections for these data are reported in Table 22. The correlations range from .12 (small) to .84 (strong). The lowest correlation was found

between two-syllable real words with r-controlled vowels and words with contractions. The highest correlation was found sentence reading and two-syllable real words with r-controlled vowels. Overall, the majority of the correlations reported in Table 22 were strong.

The correlations between sections on WRD5 for second-grade students in the winter are displayed in Table 23. Most correlations are moderate-strong. However, the correlations vary from .04 (small) between words with contractions and words with low-frequency vowel teams (oo, ea) to .85 (strong) between compound real words and compound nonsense words.

The correlations for the fall administration of WRD5 to third-grade students are displayed in Table 24. The majority of these correlations are moderate-strong, but range from .00 (trivial/no relation) to .95 (near perfect). The lowest correlation was found between real compound words and two-syllable words with diphthongs. The highest correlation was found between compound nonsense words and third-grade level high-frequency words.

The correlations for the winter administration of WRD5 for third-grade students are displayed in Table 25. The majority of these correlations are moderate-strong, but range from .03 (small) to .84 (strong). The lowest correlation was found between words with contractions and words with consonant trigraphs. The highest correlation was found between compound nonsense words and third-grade high-frequency words.

Correlations for the fall administration of WRD5 to our fourth-grade student sample are shown in Table 26. Most of these correlations are moderate to moderate-strong, but range from .04 (small) to .82 (strong). The lowest correlation was found between

nonsense compound words and words with consonant trigraphs. The highest correlation was found between words with common word parts and words with low-frequency vowel patterns (ou, er, and ow).

Examination of Item- and Section-Level Data

To address the fifth research question regarding appropriate sequencing of sections and items, primarily descriptive analyses were conducted. Each item on each Deep measure was examined to see the percent of correct student responses at the target grade level and at adjacent grades. These descriptive data were examined for both the fall and winter test administrations. Next, item level data were summarized for each section (e.g., mean and standard deviation for the percent of students earning a correct score on each item in each section). These data were examined at target grade levels and at adjacent grades for fall and winter test administrations. This information was used to help verify scope and sequence. In addition, the item data were examined for outliers. An item was considered an outlier if the mean percent students earning a correct score on the item was two standard deviations above or below the mean percent students earning items correct for that section. In what follows, we present the results of these descriptive analyses for each of the DIBELS Deep forms.

Fall kindergarten PA1. The average item percent correct within each section for the fall administration of PA1 in kindergarten students is reported in Figure 3. The percentages shown on the vertical axis of the graph reflect the average percent of students getting items correct across all items within a section. The sample size drops some across sections due to students meeting the discontinue rule. The sample for Section D1 (Saying Rhyming Words) appears to reflect some confusion regarding which

students should have been administered this task. Only those students who met the discontinue rule within the first three sections of the form were supposed to be administered Section D1 (Saying Rhyming Words). The smaller sample size for section D2 (Rhyme Recognition) was due to the fact that students only were administered this section if they did not have at least three correct responses in Section D1 (Saying Rhyming Words).

In examining the pattern across sections, overall the blending tasks were easier than segmenting tasks. In addition, within the segmenting and blending tasks, the type of blending or segmenting task did not discriminate skill level. This means, for example, that, in general, this sample of kindergarten students who exhibited blending skills were successful across blending tasks (i.e., compound words, syllables, and onset-rime).

Kindergarten and first grade PA2. Figure 4 shows the average item percent correct for each section of PA2 for kindergarten and first-grade students during the fall test administration. Several patterns are notable in these data. The sample size drops across sections for kindergarten students, but not first-grade students as tasks progress. These data suggest that the tasks were more challenging for kindergarten students than for first-grade students. In addition, tasks involving identification or production of final sounds were harder than tasks involving identification or production of initial sounds. Again, blending tasks were easier than segmenting tasks. Segmenting words with more phonemes was more difficult than segmenting shorter words.

Kindergarten and first grade WRD1. The average percent correct data for kindergarten and first-grade students' fall administration of WRD1 are shown in Figure 5. Overall, first-grade students have a higher average percent correct than

kindergarteners. Most first-grade students had the opportunity to complete every task (i.e., little drop in sample size). A large drop in sample size for kindergarten is noted after the first task and again after the section on reading consonant-vowel-consonant real words beginning with a stop sound (C1), suggesting that many kindergarteners met the discontinue rule at that point. We also note that the high-frequency words and sentences sections were not attempted with most of the kindergarten sample likely due to either (a) assessor confusion about the requirement to attempt those tasks, or (b) concern regarding student skill and potential frustration with the task given so many kindergarten students struggled with tasks beyond letter/sound correspondence (e.g., note low levels of average percent correct on reading vowel-consonant and consonant-vowel consonant words). Another pattern emerging across these data is that nonsense words were more difficult than real words. However, consonant-vowel-consonant words that began with a stop sound when compared to those that began with a continuous sound did not appear to be more difficult.

In Figure 5, the data for the sentence reading tasks are represented for when sentences were scored dichotomously (i.e., entirely correct or incorrect). This method of scoring does not show much behavior occurring in the kindergarten sample. However, Figure 6 shows alternative scoring of the sentence reading task using a single sentence as an example. In this example, instead of scoring the sentence as correct or incorrect as a whole, the number of words read correct was the scoring metric. Both kindergarten and first grade data are displayed in Figure 6. When responses were scored in this manner, the measure was much more sensitive to showing kindergarten skill and the differences in responding between kindergarten (n = 15) and first-grade (n = 44)

students are more apparent. The figure shows the data for Sentence #5 which contained seven words. For this example, every student read at least 1 word correct, but no kindergartener read more than five words correct. The distribution of kindergarten students is bi-modal, but the distribution of first-grade students is negatively skewed, with most first-grade students reading the entire sentence (all seven words) correctly.

Fall and winter first grade WRD2. The average percent of first-grade students getting items correct for fall and winter WRD2 administrations are shown in Figure 7. More students met the discontinue rule after the first three sections in the fall than in the winter, and students dropped out more continuously in the fall. While general trends in these data can be examined, it is difficult to compare these time points as the sample of students is not the same. For example, these data suggest that some sections did not differentiate skill levels and might be combined in a revised version of the measure (e.g., vowel-consonant-consonant and consonant-vowel-consonant-consonant words beginning with continuous sounds combined with those beginning with stop sounds).

Fall second grade and winter first grade WRD3. The data showing second-grade students in fall and first-grade students in winter on WRD3 is displayed in Figure 8. The WRD3 measure assesses skills students would be expected to learn in the latter half of first grade. In general, second-grade students performed better than first-grade students on these tasks. However, there were five tasks on which the first-grade students, on average, appeared to perform better than second-grade students. These tasks were (1) one-syllable non-real words with l-controlled vowels; (2) VCe and CVCe non-real words (3) words with word parts (word ending, plurals); (4) real words beginning with "qu;" and

(5) non-real words beginning with "qu." Also, words with r-controlled vowels and words with l-controlled vowels were difficult as noted by the big drop in the first-grade sample size after these tasks, suggesting that was the point at which many students met the discontinue criterion. This drop in the first-grade sample means that only the students who had not met the discontinue rule were tested on three of the five tasks where the first-grade sample outperformed the second-grade sample. Thus, the first-grade students being compared to the second-grade students are a restricted, and higher-performing, group of first-grade students.

Fall second and third grade WRD4. The data on the WRD4 measure given to second- and third-grader students during the fall testing are displayed in Figure 9. Without exception, students in third grade obtained a higher average percent correct on these tasks than second-grade students. Sections A and B (one-syllable words with vowel digraphs and one-syllable words with vowel diphthongs) on WRD4 were repeated from WRD3. These sections may have been skipped on WRD4 for students who had been administered WRD3 as the test items were identical, thus the smaller sample size for these two sections. Also, sections F through H2 (contractions, possessives, real compound words, non-real compound words) had very similar average percent correct among the third-grade student sample and appeared to be of about the same difficulty. These sections also seemed to be easier than some sections that proceeded them.

For the sentence reading task, we also compared the number of words read correctly in each sentence across these grades. An example of this comparison for one of the sentences from WRD4 is shown in Figure 10. Once again, it appears that scoring the number of words read correctly in each sentence is more sensitive than

dichotomous scoring (i.e., entirely correct or incorrect). Every student had at least one word correct in the sentence and the majority of students in both grades had all six words correct. However, there was more variability in the second-grade students' scores.

Fall third and fourth grade WRD5. Data on the average percent correct for WRD5 tasks given to third- and fourth-grade students in the fall are shown in Figure 11. The data suggest that tasks were more challenging for third-grade students than for fourth-grade students. Both the average percent correct and sample size drop for the third-grade students more than they do for fourth-grade students. It also appeared that some more challenging tasks were placed in the middle of the form (e.g., section H, words with "ch" pronounced as /k/) and perhaps should be moved toward the end given their difficulty level as observed by the low average percent correct for both grades.

Teacher Feedback

The sixth research question in this study was about teacher satisfaction with the measures. To address this research question, teachers were asked to complete a 10-item consumer satisfaction questionnaire (described previously). Each item on the questionnaire was examined descriptively. Anecdotal data were also examined. The descriptive statistics (i.e., means and standard deviations) are displayed in Table 27. The percent of teacher respondents for each agreement rating with respect to each questionnaire item are shown in Figures 12 - 20. In general, items received an “agree” rating. The sample size was smaller on item #3 because not all teachers teach those skills (e.g., upper grades) and teachers only completed the item if they taught the skills indicated in the item. All anecdotal remarks were examined as well to determine if any

general themes emerged. Two general themes were noted. First, teachers expressed concern that the measures would be too long to give to every student in their classes. Second, remarks suggested that the data obtained from the measures would be useful for targeting instruction. Sample teacher comments included:

"I would consider that for students that I refer to SIT team or when I need more in depth info on students, or questions about processing."

"More ideas on what to do to help. Sometimes data just isn't enough."

"The test was very thorough and really hit where the student's weaknesses were. However, the test did take a great deal of time to take. I don't know if it would be practical for a classroom teacher to use on all students. But, for low-readers, it is incredibly effective."

Assessor Feedback

The seventh and final research question in this study addressed assessor satisfaction with the measures. Data from the 10-item assessor questionnaire were examined descriptively along with anecdotal remarks. These descriptive statistics for each item are show in Table 28. The percent of assessor respondents for each agreement rating with respect to each questionnaire item are shown in Figures 21 - 29. In general, assessors indicated favorable views of the measures with most items receiving an agree rating. Some classroom aids who served as assessors in this study indicated a lack of comfort in responding to items 7 and 9, thus the reason for the smaller sample size for those two items. Anecdotal remarks also were generally positive. One theme that emerged was related to increasing the efficiency in test

administration. For example, providing additional clarity in the directions was suggested. Sample comments from assessors included:

"Having a "cheat sheet" of when to discontinue- sometimes it is hard to find."

"A reminder after discontinue to go on to the Sight Words and Sentence Reading."

"Electronic administration would be helpful. It would be helpful to find the exact spot they need to go next."

Discussion

Summary of Results

The results of this study indicate that the scope and sequence of items, sections and measures is generally accurate. Supporting this conclusion are data on the distribution and frequency of administration (see Table 4), the descriptive statistics for the measures (i.e., means and standard deviations) given at different grade levels (see Tables 5 and 6), and the observed drop-off in sample size for some measures as students progressed through sections of each measure (i.e., students met discontinue rules) (see Figures 3 - 5, Figures 7 - 9, and Figure 11). One general finding from the item and section-level data was that in most cases, non-real words were more difficult (i.e., yielded a lower percent correct) than parallel real word items.

The results of this study provide initial support for the construct validity of the DIBELS Deep measures. We found moderate-strong to near perfect correlations (r = .67 - .91) between DIBELS Deep measures of the same skill and moderate to moderate-strong correlations between DIBELS Deep measures of different skills (r = .41 - .63) (see Tables 7 and 8). The correlations between DIBELS Deep measures and DIBELS 6th Edition measures of the same skill also provide support for the construct validity of

the DIBELS Deep measures. In fact, we found primarily moderate to strong correlations (r = .33 - .77) when examining these relations (see Table 9).

This study also provides overall support for the content within each form being appropriate for that form. Within-form section correlations suggest that most sections were related to other sections within their respective forms. Some exceptions were found (e.g., onset-rime and rhyming sections on PA1) and correlations ranged widely across sections within each form. These results provide some initial support for the construct validity of the measures with most of the section correlations ranging from moderate to strong.

Section- and item-level results provided pertinent information about the scope and sequence of the items and sections on the forms. While in general the scope and sequence was supported, some items within sections appeared to be out of place and some sections within forms appeared to be out of place (e.g., possessives and contractions on WRD4 and WRD5; see Figures 9 and 11). One issue to consider when interpreting these data are changes in sample sizes across sections of each form. For example, a large drop in the kindergarten sample size was noted after the first task on WRD1 and again after the section on reading consonant-vowel-consonant real words beginning with a stop sound (C1) (see Figure 5) suggesting that many kindergarteners met the discontinue rule at that point. However, such results are not surprising given that the assessment was conducted in the fall of kindergarten and many of these students may not have been taught these skills yet.

A similar issue may have impacted the pattern noted between WRD3 and WRD4. The pattern suggests that second-grade students had more consistently correct

responses on sections of WRD4 than on WRD3 (see Figures 8 and 9). However, one hypothesis is that some sections did not function as intended and skewed the results. For example, section C1 (l-controlled vowels) on WRD3 was much more challenging that several tasks that came after it and was also more challenging than section C2 containing parallel items constructed of non-real words. It is unclear why the non-real words were easier than the real word parallel items in section C1. It is possible that some assessors were unclear about how to score the non-real word items. At least one comment regarding this issue was noted in the anecdotal feedback received.

Finally, the data collected in this study also support the utility of DIBELS Deep. Overall teachers agreed that the measures are useful. In addition, assessors expressed satisfaction with the usability of the measures and conveyed constructive feedback about changes to the measures that would increase their utility.

Limitations

As with all research, this study has its limitations. First and foremost, we had a relatively small sample size at each grade level. The smaller sample size limits the external validity of the study and also limited our ability to examine correlations for some sections of the measures. In addition, the study was not designed for all students to attempt all items. This procedural decision limited the study in three ways. First, it prohibited a more formal item analysis from being conducted and did not provide the opportunity to examine the factor structure of the measures. However, this study was not designed to examine the factor structure of the measures. Second, the sample size drop-off impacts the interpretation of correlational data comparing sections within each measure. Third, comparisons between measures administered across more than one

grade level should be done cautiously in cases where the sample drops off dramatically across sections of the measure at any particular grade level. Consider WRD3 given in the fall for second-grade students and the winter for first-grade students. The first-grade students remaining in the sample for the tasks at the end of the form outperformed the second-grade students on some of those latter tasks (see Figure 8). However, it is highly likely that those first-grade students were quite high performing and they represented less than a third of the entire first-grade student sample.

Finally, although the assessors who collected the data in this study were formally trained on the measures, procedural reliability and scoring accuracy data were not collected as part of this study. Future research should collect these data to increase confidence in the results.

Implications for Research and Future Development of DIBELS Deep

Though the scope and sequence of forms, sections and items was generally accurate, some sections did not perform well and others seemed to be slightly out of sequence based upon this study's results. The section correlations within forms provide direction for revision of the measures along with the average percent correct by section data. For example, we took a closer look at the sections of PA1 that had lower correlations (e.g., onset-rime and rhyme tasks). We determined that onset-rime did not appear to add information when considering other PA tasks. Similarly, the rhyming tasks did not seem to add information. Relatedly, the role of rhyming in PA (and, indeed overall reading skill) has been questioned in the research literature (McGuinness, 2005). Consistent with the section correlation findings, PA1 had a lower correlation with WRD1 for kindergarten students (see Table 7) than PA2, suggesting that PA1 scores

were less related to basic word reading and decoding skills found on WRD1 than the scores on PA2. In addition, PA1 had a generally low correlation with DIBELS 6th edition ISF.

Future research with a larger sample and during which every student is administered every item on forms at their grade level with no discontinue rule will help to further refine the ordering of sections on each form as well as the ordering of items within each section. The percent correct per item data will be used to re-order items and replace problematic items in a revised version of these measures. The mean percent correct data for each section also will be used to re-order sections depending on the number of problematic items (e.g., extreme outliers) within those sections. These changes to the measures will be implemented prior to the next round of research, where once again the scope and sequence of sections and items will be examined and then refined.

While the correlational data between DIBELS Deep forms supports the construct validity of the measures, future research should examine the construct validity more directly. One way to accomplish this would be to conduct a factor analytic study of the measures.

As mentioned previously, future research should include an examination of procedural fidelity as well as scoring accuracy (i.e., via shadow scoring or inter-scorer agreement procedures). Accomplishing this goal in future research will allow greater confidence in the results. Finally, future research should consider expanding this work to pilot similar measures in other domains critically related to reading functioning (e.g., comprehension, oral language, etc.).

Implications for Practice

The purpose of DIBELS Deep is to develop a tool for educators to assist with differentiating instruction for students who are struggling with learning to read. Often, these are Tier 2 and Tier 3 students in a system using a Response to Intervention (RtI) approach to service delivery. We believe that DIBELS Deep is a tool that assists educators to put a greater emphasis on the "I" in RtI, and provide information useful for differentiating instruction both within and across instructional tiers. It is also likely that special services personnel (e.g., reading specialist, reading coach, school psychologist) will find the measures useful as they consult with other educators regarding instructional interventions. The information from assessing with DIBELS Deep may be used in consultation with teachers about where and how to make adjustments to instruction for students, in particular, students in Tiers 2 and 3. In addition, the information may assist in the identification of appropriately targeted materials to be used by parent or peer tutors.

Conclusion

This technical report describes a phase 1 validity study conducted on DIBELS Deep, the results of the study and implications for further research and practice related to DIBELS Deep. The design characteristics of DIBELS Deep set it apart from more typical reading diagnostic measures. That is, the measures are designed to be brief and linked explicitly to DIBELS. In addition, the measures provide for dynamic assessment where the information about prompting and teaching strategies used and their effectiveness in the context of the assessment are important, as well as the information on which items a student obtains correct or incorrect scores. Based upon the data collected, we

conclude that DIBELS Deep has the potential to be a useful and practical tool for educators. Some modification of the order of tasks within measures and finalizing discontinue rules will improve ease of use. However, overall the results of this study suggest that DIBELS Deep will be helpful for more specifically targeting instructional needs and planning instruction. Further research is recommended to obtain additional information about the validity and utility of the measures.

References

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., Reschly, D. J., Schrag, J. & Tilly III, W. D. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education, Inc.

Carnine, D. W., Silbert, J., Kame'enui, E. J., Tarver, S. G., & Jungjohann, K. (2006). *Teaching struggling and at-risk readers: A Direct Instruction approach*. Upper Saddle River, NJ: Pearson.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Dynamic Measurement Group (2008). DIBELS 6th Edition Technical Adequacy Information (Tech. Rep. No. 6). Eugene, OR. Available: http://dibels.org/pubs.html.

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on Curriculum-Based Measurement validity. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 61-88). New York: Guilford.

Good, R. H., Kaminski, R. A., & Dill, S. (2002). DIBELS Oral Reading Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement. Available http://dibels.uoregon.edu/.

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., et al. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development.* (Tech. Rep. No. 7). Eugene: University of Oregon.

Jennings, J. H., Caldwell, J., & Lerner, J. W. (2006). *Reading problems: Assessment and teaching strategies* (5th ed.). Boston: Pearson.

Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for formative assessment and evaluation. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-V, (pp. 1181-1204)*. Bethesda, MD: National Association of School Psychologists.

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

Kaminski, R. A., & Good III, R. H. (1998). Assessing early literacy skills in a Problem-Solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 113-142). New York, NY: Guilford Press.

Laimon, D. E. (1994). *The Effects of a Home-Based and Center-Based Intervention on At-Risk Preschool Children's Early Literacy Skills*. Unpublished doctoral dissertation. University of Oregon, Eugene, OR.

McGuinness, D. (2005). *Language development and learning to read.: The scientific study of how language development affects reading skill.* Cambridge, MA: MIT Press.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Reports of the subgroups. Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults (third edition)*. Austin, TX: PRO-ED.

Powell-Smith, K. A., & Kaminski, R. A. (2008). Development of DIBELS Deep: A pilot study. (Technical Report No. 5) Eugene, OR: Dynamic Measurement Group.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.

Simmons, D. & Kame'enui, E. J. (1999) *Curriculum maps: Mapping instruction to achieve instructional priorities in beginning reading, kindergarten-grade 3*. Unpublished manuscript.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

United States Census Bureau (n.d.). http://www.census.gov/geo/www/us_regdiv.pdf. Retrieved July 29, 2008.

Wagner, R. K, Muse, A. E., & Tannenbaum, K. R.(Eds.) (2007). *Vocabulary acquisition: Implications for reading comprehension*. New York: Guilford.

Table 1

School Demographic Characteristics

	School Number										
	1	2	3	4	5	6	7	8	9	10	11
Locale	City: Midsize	Suburb: Large	Town: Fringe	Rural: Distant	City: Midsize	(no data)	Rural: Fringe	Suburb: Large	Town: Fringe	City: Midsize	Town: Distant
Grades Taught	KG - 6	PK - 3	KG - 4	KG - 4	KG - 5	KG - 4	KG - 4	4 - 5	KG - 4	1 - 5	PK - 4
Total Students	376	442	384	194	302	355	438	182	334	586	674
Student/Teacher Ratio	19:1	16:1	14:1	11:1	18:1	13:1	13:1	12:1	15:1	15:1	15:1
Title 1 Eligible	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes[1]	No	Yes	Yes
Free/Reduced Lunch	53%	43%	26%	37%	52%	20%	27%	48%	11%	54%	29%
Percent Female	53%	47%	42%	41%	46%	50%	48%	46%	51%	52%	49%
Student Ethnicity											
Am. Indian	2%	<1%	<1%	0	2%	<1%	2%	0	0	<1%	<1%
Asian	2%	<1%	<1%	2%	2%	2%	2%	2%	<1%	3%	0
Black	<1%	<1%	4%	<1%	27%	2%	4%	2%	<1%	29%	0
Hispanic	10%	6%	4%	2%	11%	8%	8%	6%	1%	11%	<1%
White	84%	92%	83%	96%	59%	87%	84%	90%	95%	56%	99%

[1]School-Wide Title I Program.

Table 2

Number and Percent of Students by DIBELS 6th Edition Instructional Recommendation by Grade and Time of Year

	Intensive		Strategic		Benchmark	
Grade	Fall	Winter	Fall	Winter	Fall	Winter
Kindergarten	12 (24%)	3 (6%)	17 (35%)	21 (40%)	20 (41%)	29 (55%)
First Grade	16 (30%)	16 (24%)	15 (28%)	25 (38%)	23 (43%)	25 (38%)
Second Grade	16 (31%)	18 (29%)	12 (24%)	17 (27%)	23 (45%)	27 (44%)
Third Grade	14 (29%)	17 (29%)	16 (34%)	15 (25%)	17 (36%)	27 (46%)
Fourth Grade	13 (30%)	-	14 (33%)	-	16 (37%)	-

Table 3

Means and Standard Deviations on DIBELS Measures by Grade and Time of Year

	DIBELS Measure									
	LNF		ISF		PSF		NWF		DORF	
Grade	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter
KG	13.3 (11. 9) (n = 49)	31.9 (15.2) (n = 63)	8.7 (7.1) (n = 49)	28.0 (14.6) (n = 63)	-	30.3 (14.8) (n = 63)	-	15.3 (11.2) (n = 63)	-	-
First	31.9 (16.4) (n = 54)	-	-	-	39.6 (12.0) (n = 54)	53.9 (11.9) (n = 66)	25.5 (24.7) (n = 54)	55.6 (31.7) (n = 66)	-	30.5 (33.8) (n = 66)
Second	-	-	-	-	-	-	55.7 (20.1) (n = 33)	-	51.3 (39.1) (n = 51)	73.5 (44.5) (n = 62)
Third	-	-	-	-	-	-	-	-	77.0 (33.8) (n = 47)	91.3 (37.9) (n = 59)
Fourth	-	-	-	-	-	-	-	-	87.8 (32.5) (n = 43)	128.5 (38.7) (n = 6)

Note. KG = Kindergarten. LNF = Letter Naming Fluency, ISF = Initial Sound Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency, and DORF = DIBELS Oral Reading Fluency.

Table 4.

Number of Students Given DIBELS Deep Measures by Grade and Time of Year

	DIBELS Deep Measure													
	PA1		PA2		WRD1		WRD2		WRD3		WRD4		WRD5	
Grade	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter
Kindergarten	47	6	44	58	39	55	0	4	0	0	0	0	0	0
First	1	0	45	1	46	13	49	57	9	56	0	48	0	3
Second	0	0	1	1	7	9	16	16	47	16	42	57	5	53
Third	0	0	0	0	4	0	9	4	12	6	38	9	41	57
Fourth	0	0	0	0	0	0	6	0	6	0	8	0	42	6
All Grades	48	6	90	60	96	77	80	81	74	78	88	114	88	119

Note. PA1 = Phonemic Awareness Form 1, PA2 = Phonemic Awareness Form 2, WRD1 = Word Reading and Decoding Form 1, WRD2 = Word Reading and Decoding Form 2, WRD3 = Word Reading and Decoding Form 3, WRD4 = Word Reading and Decoding Form 4, and WRD5 = Word Reading and Decoding Form 5.

Table 5

Means and Standard Deviations for DIBELS Deep Measures by Grade (Fall)

	DIBELS Deep Measure						
Grade	PA1	PA2	WRD1	WRD2	WRD3	WRD4	WRD5
Kindergarten	30.40 (13.00)*	38.07 (14.22)*	20.38 (19.49)	-	-	-	-
	(n = 47)	(n = 44)	(n = 39)				
First	43 (na)	51.29 (3.80)	91.61 (25.07)	82.34 (66.52)*	70.78 (48.43) [a]	-	-
	(n = 1)	(n = 45)	(n = 46)	(n = 49)	(n = 9)		
Second	-	-	72.86 (25.12) [a]	79.31 (52.87)	78.81 (48.06)	83.45 (44.69)*	64.60 (46.52) [a]
			(n = 7)	(n = 16)	(n = 47)	(n = 42)	(n = 5)
Third	-	-	60.75 (45.63) [a]	66.56 (52.05) [a]	61.83 (36.57)	101.11 (39.05)	84.44 (35.50)*
			(n = 4)	(n = 9)	(n =12)	(n = 38)	(n = 41)
Fourth	-	-	-	50.17 (43.18) [a]	32.67 (24.82) [a]	42.38 (46.28) [a]	102.12 (28.50)
				(n = 6)	(n = 6)	(n = 8)	(n = 42)

Note. Standard deviations are noted in parentheses. PA1 = Phonemic Awareness Form 1 (maximum possible score = 60), PA2 = Phonemic Awareness Form 2 (maximum possible score = 55), WRD1 = Word Reading and Decoding Form 1 (maximum possible score = 119), WRD2 = Word Reading and Decoding Form 2 (182), WRD3 = Word Reading and Decoding Form 3 (maximum possible score = 144), WRD4 = Word Reading and Decoding Form 4 (maximum possible score = 132), and WRD5 = Word Reading and Decoding Form 5 (maximum possible score = 131).

* Target grade level and time frame

[a] very small sample (n < 10)

Table 6

Means and Standard Deviations for DIBELS Deep Measures by Grade (Winter)

	DIBELS Deep Measure						
Grade	PA1	PA2	WRD1	WRD2	WRD3	WRD4	WRD5
Kindergarten	18.67 (12.21) [a]	43.55 (10.03)*	39.49 (27.25)*	10.75 (13.02) [a]	-	-	-
	(n = 6)	(n = 58)	(n = 55)	(n = 4)			
First	-	51 (na)	77.15 (19.07)	105.93 (58.86)	49.46 (44.55)*	52.02 (47.59)	84.00 (37.51) [a]
		(n = 1)	(n = 13)	(n = 57)	(n = 56)	(n = 48)	(n = 3)
Second	-	-	86.33 (9.72) [a]	56.94 (38.32)	44.16 (34.46)	83.77 (45.19)*	66.53 (37.96)
			(n = 9)	(n = 16)	(n = 19)	(n = 57)	(n = 53)
Third	-	-	-	113.75 (31.03) [a]	41.33 (30.38) [a]	81.89 (47.42) [a]	98.08 (28.60)*
				(n = 4)	(n = 6)	(n = 9)	(n = 57)
Fourth	-	-	-	-	-	-	115.67 (11.67) [a]
							(n = 6)

Note. Standard deviations are noted in parentheses. PA1 = Phonemic Awareness Probe 1 (maximum possible score = 60), PA2 = Phonemic Awareness Probe 2 (maximum possible score = 55), WRD1 = Word Reading and Decoding Probe 1 (maximum possible score = 119), WRD2 = Word Reading and Decoding Probe 2 (182), WRD3 = Word Reading and Decoding Probe 3 (maximum possible score = 144), WRD4 = Word Reading and Decoding Probe 4 (maximum possible score = 132), and WRD5 = Word Reading and Decoding Probe 5 (maximum possible score = 131).

* Target grade level and time frame

[a] very small sample (n < 10)

Table 7

Correlations Between DIBELS Deep Measures Given in the Fall

Variable	PA1	PA2	WRD1	WRD2	WRD3	WRD4	WRD5
PA1	-	.72 (n = 44 KG[a])	.41 (n = 38 KG[a])	.	.	.	.
PA2		-	.61 (n = 38 KG[a])	.63 (n = 44 1st)	.	.	.
			.46 (n = 44 1st)				
WRD1			-	.89 (n = 43 1st)	.	.	.
WRD2				-	.	.	.
WRD3					-	.84 (n = 41 2nd)	.
WRD4						-	.67 (n = 36 3rd)
WRD5							-

Note. All correlations are statistically significant, $p < .05$. Data are not reported in cases where $n < 20$. PA1 = Phonemic Awareness Form 1, PA2 = Phonemic Awareness Form 2, WRD1 = Word Reading and Decoding Form 1, WRD2 = Word Reading and Decoding Form 2, WRD3 = Word Reading and Decoding Form 3, WRD4 = Word Reading and Decoding Form 4, and WRD5 = Word Reading and Decoding Form 5.

[a]KG = Kindergarten

Table 8

Correlations Between DIBELS Deep Measures Given in the Winter

Variable	PA1	PA2	WRD1	WRD2	WRD3	WRD4	WRD5
PA1	-	.	.	.	.	.	.
PA2		-	.58 (n = 54 KG[a])	.	.	.	.
WRD1			-	.	.	.	.
					.79 (n = 52 1st)		
WRD2				-		.84 (n = 46 1st)	.
WRD3					-	.91 (n = 46 1st)	.
WRD4						-	.85 (n = 53 2nd)
WRD5							-

Note. All correlations are statistically significant, $p < .05$. Data are not reported in cases where $n < 20$. PA1 = Phonemic Awareness Form 1, PA2 = Phonemic Awareness Form 2, WRD1 = Word Reading and Decoding Form 1, WRD2 = Word Reading and Decoding Form 2, WRD3 = Word Reading and Decoding Form 3, WRD4 = Word Reading and Decoding Form 4, and WRD5 = Word Reading and Decoding Form 5.

[a]KG = Kindergarten

Table 9

Correlations Between DIBELS Deep Measures and DIBELS 6th Edition Benchmark Data

Measure	LNF		ISF		PSF		NWF		DORF	
	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter
PA1										
Kindergarten	.20 (47)	.	.23 (47)	.	.	.	.	.	.	.
PA2										
Kindergarten	.47* (44)	.36* (57)	.40* (44)	.44* (57)	.	.61* (57)	.	.44* (57)	.	.
First Grade	.36* (45)	.	.	.	.44* (45)	.	.29 (45)	.	.	.
WRD1										
Kindergarten	.61* (39)	.75* (54)	.43* (39)	.49* (54)	.	.24 (54)	.	.76* (54)	.	.
First Grade	.42* (46)	.	.	.	.18 (46)	.	.50* (46)	.	.	.
WRD2										
First Grade	.49* (49)	.	.	.	.32* (49)	.27* (57)	.55* (49)	.66* (57)	.	.62* (57)
WRD3										
First Grade	.	.	.	.	.	.08 (56)	.	.76* (56)	.	.77* (56)
Second Grade	.	.	.	.	.	.	.33 (30)	.	.61* (47)	.
WRD4										
Second Grade	.	.	.	.	.	.	.47* (27)	.	.70* (42)	.67* (57)
Third Grade	.	.	.	.	.	.	.	.	.48* (38)	.
WRD5										
Third Grade	.	.	.	.	.	.	.	.	.66* (41)	.64* (57)
Fourth Grade	.	.	.	.	.	.	.	.	.56* (42)	.

Note. Correlations reflect measures given at the same time of year. Numbers in parentheses indicate sample size. Data are not reported in cases where n < 20, or where one of the measures was not appropriate for student grade or time of year. PA1 = Phonemic Awareness Form 1, PA2 = Phonemic Awareness Form 2, AP1 = Word Reading and Decoding Form 1, WRD2 = Word Reading and Decoding Form 2, WRD3 = Word Reading and Decoding Form 3, WRD4 = Word Reading and Decoding Form 4, and WRD5 = Word Reading and Decoding Form 5. LNF = Letter Naming Fluency, ISF = Initial Sound Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency, and DORF = DIBELS Oral Reading Fluency. * $p < .05$

Table 10

Correlations Between DIBELS Deep Phonemic Awareness Form 1 Sections in Kindergarten (Fall)

Section	B1	C1	A2	B2
A1	.47*(46)	.17(42)	.60*(46)	.37(42)
B1	__	-.05(42)	.40*(46)	.24(42)
C1		__	.04(42)	.15(42)
A2			__	.69*(42)
	C2	C1a	D1	D2
A1	.42*(42)	-	.36(28)	-
B1	.39(42)	-	.32(28)	-
C1	.13(43)	-	.40(25)	-
A2	.42*(42)	-	0.26(28)	-
B2	.33*(42)	-	0.35(25)	-
C2	__	-	.40*(25)	-
C1a		__	-	-
D1			__	-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. Data are not reported in cases where n < 20. Data are reported for fall only due to small sample sizes in the winter. A1 = Blending Compound Words, B1 = Blending Syllables, C1 = Blending Onset-Rime, A2 = Segmenting Compound Words, B2 = Segmenting Syllables, C2 = Segmenting Onset-Rime, C1a = Matching Rimes, D1 = Saying Rhyming Words, D2 = Recognizing Rhyming Words.
*p < .05

Table 11

Correlations Between DIBELS Deep Phonemic Awareness Form 2 Sections in Kindergarten (Fall and Winter)

Section	A2		B1		B2		C1	
	F	W	F	W	F	W	F	W
A1	.72*(44)	.90*(58)	.64*(44)	.27*(58)	.51*(41)	.16(58)	.22(41)	.57*(58)
A2	__	__	.67*(44)	.26*(58)	.37*(41)	.17(58)	.38*(41)	.59*(58)
B1			__	__	.18(41)	.30*(58)	.30(41)	.34*(58)
B2					__	__	.35*(41)	.19(58)
	C2		D1		D2		D3	
	F	W	F	W	F	W	F	W
A1	.23(41)	.52*(58)	.36*(40)	.65*(57)	.38*(39)	.64*(56)	.26(34)	.18(52)
A2	.45*(41)	.59*(58)	.33*(40)	.68*(57)	.35*(39)	.65*(56)	.41*(34)	.28*(52)
B1	.34*(41)	.38*(58)	.25(40)	.33*(57)	.28(39)	.28*(56)	.22(34)	.19(52)
B2	.34*(41)	.04(58)	.03(40)	.004(57)	.34*(39)	.06(56)	.36*(34)	.23(52)
C1	.71*(41)	.68*(58)	.25(40)	.58*(57)	.45*(39)	.34*(56)	.27(34)	.42*(52)
C2	__	__	.45*(40)	.60*(57)	.64*(39)	.44*(56)	.49*(34)	.26(52)
D1			__	__	.72*(39)	.82*(56)	.33(34)	.42*(52)
D2					__	__	.50*(34)	.34*(52)
D3							__	__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. F = Fall, W = Winter, A1 = Blending Two-Phoneme Words, A2 = Blending Three-Phoneme Words, B1 = Matching Initial Sounds, B2 = Production of Initial Sounds, C1 = Matching Final Sounds, C2 = Production of Final Sounds, D1 = Segmenting Two-Phoneme Words, D2 = Segmenting Three-Phoneme Words, D3 = Segmenting Four-Phoneme Words with Blends.
*$p < .05$

Table 12

Correlations Between DIBELS Deep Phonemic Awareness Form 2 Sections in First Grade (Fall, N = 45)

Section	A2	B1	B2	C1
A1	-.07	-.03	-.03	.08
A2	__	.14	.62*	.25
B1		__	.48*	.12
B2			__	.002
	C2	D1	D2	D3
A1	.02	-.06	-.05	.18
A2	.47*	.26	.59*	.47*
B1	.48*	-.08	.20	.14
B2	.48*	-.08	.74*	.37*
C1	.48*	-.07	-.06	.19
C2	__	.12	.51*	.42*
D1		__	.20	.11
D2			__	.36*
D3				__

Note. Correlations reported are based upon subjects with pairwise complete data. A1 = Blending Two-Phoneme Words, A2 = Blending Three-Phoneme Words, B1 = Matching Initial Sounds, B2 = Production of Initial Sounds, C1 = Matching Final Sounds, C2 = Production of Final Sounds, D1 = Segmenting Two-Phoneme Words, D2 = Segmenting Three-Phoneme Words, D3 = Segmenting Four-Phoneme Words with Blends.
*$p < .05$

Table 13

Correlations Between DIBELS Deep Word Reading and Decoding Form 1 Sections in Kindergarten (Fall and Winter)

Section	**B1**		**B2**		**C1**	
	F	W	F	W	F	W
A	.58*(27)	.69*(52)	.51*(25)	.64*(50)	.55*(23)	.67*(46)
B1	__	__	.85*(25)	.81*(51)	.84*(23)	.88*(47)
B2			__	__	.82*(23)	.83*(47)
	C2		**D**		**E**	
	F	W	F	W	F	W
A	-	.58*(25)	-	.52*(50)	-	.51*(40)
B1	-	.71*(25)	-	.59*(49)	-	.64*(40)
B2	-	.89*(25)	-	.52*(47)	-	.68*(40)
C1	-	.73*(25)	-	.59*(44)	-	.67*(40)
C2	__	__	-	.19(25)	-	.61*(25)
D			__	__	-	.70*(41)
E					__	__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. Data are not reported in cases where n < 20. A = Letter-Sound Correspondences, B1 = VC and CVC Words Beginning with Continuous Sounds (real), B2 = VC and CVC Words Beginning with Continuous Sounds (nonsense), C1 = CVC Words Beginning with Stop Sounds (real), C2 = CVC words Beginning with Stop Sounds (nonsense), D = Pre-Primer High-Frequency Words, E = Sentence Reading.
*p < .05

Table 14

Correlations Between DIBELS Deep Word Reading and Decoding Form 1 Sections in First Grade (Fall)

Section	B1	B2	C1	C2	D	E
A	.68*(45)	.44*(44)	.29(44)	.31(43)	.21(44)	.34*(44)
B1	__	.63*(45)	.48*(44)	.52*(43)	.54*(44)	.56*(44)
B2		__	.63*(44)	.73*(43)	.41*(44)	.58*(44)
C1			__	.76*(43)	.60*(44)	.81*(44)
C2				__	.51*(43)	.70*(43)
D					__	.74*(44)
E						__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A = Letter-Sound Correspondences, B1 = VC and CVC Words Beginning with Continuous Sounds (real), B2 = VC and CVC Words Beginning with Continuous Sounds (nonsense), C1 = CVC Words Beginning with Stop Sounds (real), C2 = CVC Words Beginning with Stop Sounds (nonsense), D = Pre-Primer High-Frequency Words, E = Sentence Reading.
*$p < .05$

Table 15

Correlations Between DIBELS Deep Word Reading and Decoding Form 2 Sections in First Grade (Fall)

Section	A2	B1	B2	C1	C2	D1	D2	E1
A1	.86*(45)	.82*(44)	.67*(34)	.74*(31)	.61*(30)	.67*(28)	.47*(27)	.78*(27)
A2	__	.87*(44)	.82*(34)	.73*(31)	.60*(30)	.69*(28)	.56*(27)	.70*(27)
B1		__	.72*(35)	.72*(32)	.52*(31)	.77*(28)	.60*(27)	.62*(27)
B2			__	.72*(32)	.67*(31)	.69*(28)	.76*(27)	.66*(27)
C1				__	.48*(32)	.89*(28)	.68*(27)	.44*(28)
C2					__	.45*(28)	.41*(27)	.61*(28)
D1						__	.78*(28)	.56*(27)
D2							__	.50*(27)
	E2	F	G1	G2	H1	H2	I	J
A1	.66*(26)	.16(26)	.50*(26)	.54*(25)	.71*(25)	.69*(24)	.63*(37)	.57*(35)
A2	.76*(26)	.24(26)	.53*(26)	.78*(25)	.69*(25)	.74*(24)	.71*(36)	.56*(35)
B1	.85*(26)	.26(26)	.52*(27)	.64*(26)	.60*(26)	.74*(25)	.58*(37)	.70*(36)
B2	.65*(26)	.31(26)	.55*(27)	.74*(26)	.82*(26)	.45*(25)	.63*(31)	.62*(30)
C1	.37(27)	.28(26)	.47*(27)	.44*(26)	.36(26)	.45*(25)	.42*(30)	.44*(30)
C2	.54*(27)	.26(26)	.61*(27)	.54*(26)	.73*(26)	.45*(25)	.41*(30)	.41*(30)
D1	.54*(26)	.46*(26)	.70*(27)	.72*(26)	.55*(26)	.65*(25)	.46*(27)	.57*(27)
D2	.50*(26)	.52*(26)	.47*(27)	.71*(26)	.59*(26)	.62*(25)	.59*(27)	.75*(27)
E1	.62*(27)	.28(26)	.53*(26)	.55*(25)	.70*(25)	.66*(24)	.44*(27)	.42*(27)
E2	__	.22(26)	.31(26)	.79*(25)	.65*(25)	.71*(24)	.44*(26)	.47*(26)
F		__	.43*(26)	.51*(25)	.23(25)	.39(24)	.66*(25)	.63*(25)
G1			__	.64*(27)	.78*(27)	.39(26)	.31(27)	.48*(27)

G2	__	.72*(27)	.66*(26)	.46*(27)	.56*(27)
H1		__	.58*(26)	.18(27)	.46*(27)
H2			__	.29(26)	.60*(26)
I				__	.80*(26)
J					__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = VCC & CVCC Words Beginning with Continuous Sounds (real), A2 = VCC & CVCC Words Beginning with Continuous Sounds (nonsense), B1 = CVCC Words Beginning with Stop Sounds (real), B2 = CVCC Words Beginning with Stop Sounds (nonsense), C1 = CCVC Words with Both Initial Sounds Continuous (real), C2 = CCVC with Both Initial Sounds Continuous (nonsense), D1 = CCVC with One Initial Stop Sound (real), D2 = CCVC with One Initial Stop Sound (nonsense), E1 = Words with Double Final Consonants (real), E2 = Words with Double Final Consonants (nonsense), F = Reading Words with Y as a Vowel, G1 = CCVCC Words (real), G2 = CCVCC Words (nonsense), H1 = CCCVC and CCCVCC Words (real), CCCVC and CCCVCC Words (nonsense), I = Primer High-Frequency Words, J = Sentence Reading. *$p < .05$

Table 16

Correlations Between DIBELS Deep Word Reading and Decoding Form 2 Sections in First Grade (Winter)

Section	A2	B1	B2	C1	C2	D1	D2	E1
A1	.76*(57)	.79*(57)	.74*(52)	.72*(46)	.75*(46)	.59*(44)	.69*(43)	.63*(42)
A2	__	.73*(57)	.77*(52)	.56*(46)	.56*(46)	.69*(44)	.72*(43)	.54*(42)
B1		__	.85*(52)	.67*(46)	.67*(46)	.62*(44)	.76*(43)	.56*(42)
B2			__	.63*(46)	.64*(46)	.68*(44)	.82*(43)	.55*(42)
C1				__	.75*(46)	.66*(44)	.51*(43)	.48*(42)
C2					__	.60*(44)	.66*(43)	.63*(42)
D1						__	.74*(43)	.55*(42)
D2							__	.78*(42)
	E2	F	G1	G2	H1	H2	I	J
A1	.42*(42)	.46*(40)	.26(38)	.42*(38)	.55*(38)	.44*(37)	.58*(55)	.74*(55)
A2	.49*(42)	.51*(40)	.50*(38)	.56*(38)	.65*(38)	.63*(37)	.62*(55)	.76*(55)
B1	.29(42)	.54*(40)	.32(38)	.48*(38)	.49*(38)	.56*(37)	.65*(55)	.68*(55)
B2	.44*(42)	.67*(40)	.29(38)	.62*(38)	.53*(38)	.68*(37)	.79*(50)	.74*(50)
C1	.17(42)	.64*(40)	.34*(38)	.51*(38)	.55*(38)	.55*(37)	.64*(45)	.65*(45)
C2	.39*(42)	.55*(40)	.46*(38)	.51*(38)	.48*(38)	.43*(37)	.67*(45)	.69*(45)
D1	.32*(42)	.52*(40)	.29(38)	.73*(38)	.78*(38)	.47*(37)	.59*(44)	.58*(44)
D2	.61*(42)	.48*(40)	.36*(38)	.72*(38)	.71*(38)	.69*(37)	.64*(43)	.52*(43)
E1	66*(42)	.48*(40)	.31(38)	.54*(38)	.64*(38)	.56*(37)	.51*(42)	.43*(42)
E2	__	.23(40)	.13(38)	.37*(38)	.27(38)	.52*(37)	.56*(42)	.43*(42)
F		__	.24(38)	.41*(38)	.42*(38)	.61*(37)	.56*(40)	.68*(40)

G1	__	.51*(38)	.33*(38)	.16(37)	.37*(38)	.37*(38)
G2		__	.61*(38)	.57*(37)	.64*(38)	.48*(38)
H1			__	.56*(37)	.53*(38)	.35*(38)
H2				__	.68*(37)	.60*(37)
I					__	.75*(55)
J						__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = VCC & CVCC Words Beginning with Continuous Sounds (real), A2 = VCC & CVCC Words Beginning with Continuous Sounds (nonsense), B1 = CVCC Words Beginning with Stop Sounds (real), B2 = CVCC Words Beginning with Stop Sounds (nonsense), C1 = CCVC Words with Both Initial Sounds Continuous (real), C2 = CCVC with Both Initial Sounds Continuous (nonsense), D1 = CCVC with One Initial Stop Sound (real), D2 = CCVC with One Initial Stop Sound (nonsense), E1 = Words with Double Final Consonants (real), E2 = Words with Double Final Consonants (nonsense), F = Reading Words with Y as a Vowel, G1 = CCVCC Words (real), G2 = CCVCC Words (nonsense), H1 = CCCVC and CCCVCC Words (real), CCCVC and CCCVCC Words (nonsense), I = Primer High-Frequency Words, J = Sentence Reading. *$p < .05$.

Table 17

Correlations Between DIBELS Deep Word Reading and Decoding Form 3 Sections in First Grade (Winter)

Section	A2	A2	B1	B2	C1	C2	D1	D2	E	F
A1	___	.87*(55)	.63*(55)	.55*(45)	.40*(40)	.15(23)	.49*(21)	.37(20)	-	-
A2		___	.59*(55)	.58*(45)	.36*(40)	.13(23)	.42(21)	.40(20)	-	-
B1			___	.76*(45)	.63*(40)	.45*(23)	.36(21)	22(20)	-	-
B2				___	.62*(40)	.45*(23)	.34(21)	.09(20)	-	-
C1					___	.68*(23)	.48*(21)	.31(20)	-	-
C2						___	.41(21)	.41(21)	-	-
D1							___	.60*(20)	-	-
D2								___	-	-
E									___	-

	G	H	I1	I2	J	K	L1	L2	M	N
A1	-	-	-	-	-	-	-	-	.68*(49)	.57*(48)
A2	-	-	-	-	-	-	-	-	.67*(49)	.48*(48)
B1	-	-	-	-	-	-	-	-	.66*(49)	.66*(48)
B2	-	-	-	-	-	-	-	-	.52*(39)	.50*(39)
C1	-	-	-	-	-	-	-	-	.34(34)	.62*(34)
C2	-	-	-	-	-	-	-	-	.42(21)	.66*(21)
D1	-	-	-	-	-	-	-	-	-	-
D2	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-
G	___	-	-	-	-	-	-	-	-	-
H		___	-	-	-	-	-	-	-	-
I1			___	-	-	-	-	-	-	-

I2	___	-	-	-	-	-	-
J		___	-	-	-	-	-
K			___	-	-	-	-
L1				___	-	-	-
L2					___	-	-
M						___	.80*(49)
N							___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. Data are not reported in cases where n < 20. A1 = Words with Consonant Digraphs (real), A2 = Words with Consonant Digraphs (nonsense), B1 = One-Syllable Words R-Controlled Vowels (real), B2 = One-Syllable Words R-Controlled Vowels (nonsense), C1 = One-Syllable Words L-Controlled Vowels (real), C2 = One-Syllable Words L-Controlled Vowels (nonsense), D1 = VCe and CVCe Words (real), D2 = VCe and CVCe Words (nonsense), E = Words with Hard & Soft C, F = Words with Hard & Soft G, G = Words with Word Parts (word endings, plurals), H = Contractions, I1 = Compound Words (real), I2 = Compound Words (nonsense), J = One-syllable Words with Vowel Digraphs, K = One-Syllable Words with Vowel Diphthongs, L1 = Words Beginning with “qu” (real), L2 = Words Beginning with “qu” (nonsense), M = First-Grade High-Frequency Words, N = Sentence Reading. *p < .05.

Table 18

Correlations Between DIBELS Deep Word Reading and Decoding Form 3 Sections in Second Grade (Fall)

Section	A1	A2	B1	B2	C1	C2	D1	D2	E	F
A1	___	.76*(43)	.64*(43)	.67*(41)	.45*(39)	.57*(32)	.83*(29)	.57*(28)	.38*(27)	.37(26)
A2		___	.58*(43)	.64*(41)	.48*(39)	.50*(32)	.49*(29)	.47*(28)	.24(27)	.41*(26)
B1			___	.73*(41)	.59*(40)	.53*(33)	.62*(30)	.61*(29)	.67*(28)	.58*(27)
B2				___	.55*(40)	.63*(33)	.54*(30)	.48*(29)	.61*(28)	.59*(27)
C1					___	.67*(33)	.43*(30)	.58*(29)	.60*(28)	.72*(27)
C2						___	.60*(30)	.50*(29)	.49*(28)	.63*(27)
D1							___	.62*(30)	.45*(29)	.29(28)
D2								___	.33(29)	.23(28)
E									___	.56*(28)
	G	H	I1	I2	J	K	L1	L2	M	N
A1	.45*(26)	.37(26)	.58*(26)	.44*(26)	.63*(25)	.58*(25)	.46*(26)	.47*(26)	.86*(41)	.75*(41)
A2	.50*(26)	.14(26)	.08(26)	.31(26)	.06(24)	.03(24)	.49*(26)	.45*(26)	.70*(41)	.65*(41)
B1	.63*(26)	.38(26)	.78*(26)	.63*(26)	.77*(25)	.74*(25)	.63*(27)	.65*(27)	.74*(41)	.78*(41)
B2	.62*(26)	.59*(26)	.40*(26)	.50*(26)	.57*(25)	.64*(25)	.67*(25)	.61*(27)	.66*(39)	.68*(39)
C1	.57*(26)	.12(26)	.37(26)	.61*(26)	.52*(25)	.62*(25)	.59*(27)	.61*(27)	.47*(37)	.61*(37)
C2	.46*(26)	.51*(26)	.39*(26)	.43*(26)	.48*(25)	.64*(25)	.60*(27)	.72*(27)	.51*(32)	.61*(32)
D1	.36(26)	.50*(26)	.71*(26)	.64*(26)	.72*(25)	.74*(25)	.56*(27)	.43*(27)	.86*(29)	.80*(29)
D2	.20(26)	.11(26)	.43*(26)	.51*(26)	.46*(25)	.36(25)	.54*(27)	.53*(27)	.63*(28)	.49*(28)
E	.61*(26)	.45*(26)	.63*(26)	.66*(26)	.61*(25)	.75*(25)	.45*(27)	.41*(27)	.77*(27)	.79*(27)
F	.51*(26)	.12(26)	.39(26)	.55*(26)	.55*(25)	.65*(25)	.43*(27)	.45*(27)	.54*(26)	.62*(26)
G	___	.40*(26)	.55*(26)	.66*(26)	.52*(24)	.62*(24)	.48*(26)	.37(26)	.61*(26)	.62*(26)
H		___	.46*(26)	.31(26)	.40(24)	.58*(24)	.46*(26)	.47*(26)	.52*(26)	.35(26)
I1			___	.79*(26)	.83*(24)	.80*(24)	.43*(26)	.43*(26)	.88*(26)	.77*(26)

I2	___	.73*(24)	.70*(24)	.63*(26)	.39(26)	.76*(26)	.68*(26)
J		___	.84*(25)	.51*(25)	.50*(25)	.89*(24)	.77*(24)
K			___	.51(25)	.46*(25)	.87*(24)	.84*(24)
L1				___	.69*(27)	.50*(26)	.27(26)
L2					___	.53*(26)	.31(26)
M						___	.85*(41)
N							___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = Words with Consonant Digraphs (real), A2 = Words with Consonant Digraphs (nonsense), B1 = One-Syllable Words R-Controlled Vowels (real), B2 = One-Syllable Words R-Controlled Vowels (nonsense), C1 = One-Syllable Words L-Controlled Vowels (real), C2 = One-Syllable Words L-Controlled Vowels (nonsense), D1 = VCe and CVCe Words (real), D2 = VCe and CVCe Words (nonsense), E = Words with Hard & Soft C, F = Words with Hard & Soft G, G = Words with Word Parts (word endings, plurals), H = Contractions, I1 = Compound Words (real), I2 = Compound Words (nonsense), J = One-syllable Words with Vowel Digraphs, K = One-Syllable Words with Vowel Diphthongs, L1 = Words Beginning with “qu” (real), L2 = Words Beginning with “qu” (nonsense), M = First-Grade High-Frequency Words, N = Sentence Reading. *$p < .05$.

Table 19

Correlations Between DIBELS Deep Word Reading and Decoding Form 4 Sections in First Grade (Winter)

Section	A	B	C	D	E1	E2	F	G
A	__	.87*(46)	.86*(46)	.74*(33)	.83*(26)	.44*(26)	.59*(24)	.60*(22)
B		__	.86*(46)	.75*(33)	.75*(26)	.41*(26)	.64*(24)	.68*(22)
C			__	.66*(34)	.72*(27)	.32(27)	.40(24)	.42(22)
D				__	.59*(27)	.51*(27)	.42*(24)	.30(22)
E1					__	.62*(27)	.45*(24)	.65*(22)
E2						__	.14(24)	.47*(22)
F							__	.17(22)
	H1	H2	I	J	K	L	M	N
A	.79*(22)	.71*(22)	.56*(21)	.57*(21)	.75*(20)	.56*(20)	.92*(46)	.83*(44)
B	.71*(22)	.88*(22)	.70*(21)	.61*(21)	.85*(20)	.65*(20)	.89*(46)	.84*(44)
C	.38(22)	.54*(22)	.51*(21)	.64*(21)	.47*(20)	.50*(20)	.81*(47)	.75*(45)
D	.42(22)	.49*(22)	.59*(21)	.49*(21)	.79*(20)	.62*(20)	.64*(34)	.76*(34)
E1	.70*(22)	.56*(22)	.63*(21)	.64*(21)	.64*(20)	.74*(20)	.78*(27)	.78*(27)
E2	.46*(22)	.30(22)	.54*(21)	.64*(21)	.27(20)	.46*(20)	.47*(27)	.58*(27)
F	.46*(22)	.58*(22)	.14(21)	.28(21)	.51*(20)	.30(20)	.60*(24)	.60*(24)
G	.72*(22)	.62*(22)	.75*(21)	.49*(21)	.58*(20)	.71*(20)	.85*(22)	.68*(22)
H1	__	.80*(23)	.63*(22)	.54*(21)	.74*(20)	.70*(20)	.75*(23)	.74*(23)
H2		__	.69*(22)	.57*(21)	.74*(20)	.55*(20)	.79*(23)	.77*(23)
I			__	.83*(21)	.64*(20)	.72*(20)	.80*(22)	.70*(22)
J				__	.57*(20)	.68*(20)	.64*(21)	.67*(21)
K					__	.81*(20)	.82*(20)	.88*(20)

L	__	.75*(20)	.79*(20)
M		__	.80*(46)
N			__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A = One-Syllable Words with Vowel Digraphs, B = One-Syllable Words with Vowel Diphthongs, C = Words with Medial Double Consonants, D = Words with Consonant Digraphs (ck, gh, ph, & wr), E1 = Two-Syllable Words with R-Controlled Vowels (real), E2 = Two-Syllable Words with R-Controlled Vowels (nonsense), F = Contractions, G = Possessives, H1 = Compound Words (real), H2 = Compound Words (nonsense), I = Two-Syllable Words with Short & Long Vowel Patterns and Inflectional Endings, J = Two-Syllable Words with Vowel Digraphs and Inflectional Endings, K = Words with Common Word Parts, L = Multisyllabic Words, M = Second-Grade High-Frequency Words, N = Sentence Reading. *$p < .05$.

Table 20

Correlations Between DIBELS Deep Word Reading and Decoding Form 4 Sections in Second Grade (Fall)

Section	A	B	C	D	E1	E2	F	G
A	__	.91*(37)	.86*(37)	.88*(33)	.81*(32)	.58*(29)	.48*(26)	.02(26)
B		__	.81*(37)	.83*(33)	.77*(32)	.69*(29)	.35(26)	.09(26)
C			__	.75*(35)	.73*(35)	.58*(32)	.57*(28)	-.04(28)
D				__	.86*(35)	.67*(31)	.47*(28)	.16(28)
E1					__	.67*(32)	.48*(28)	.19(28)
E2						__	.51*(28)	.00(28)
F							__	.28(28)
	H1	H2	I	J	K	L	M	N
A	.79*(26)	.36(26)	.34(26)	.55*(26)	.72*(26)	.31(25)	.93*(34)	.79*(34)
B	.58*(26)	.40*(26)	.38(26)	.67*(26)	.74*(26)	.67*(25)	.89*(34)	.80*(34)
C	.41*(28)	.14(28)	.20(28)	.35(28)	.66*(29)	.51*(27)	.78*(34)	.73*(38)
D	.62*(28)	.35(28)	.64*(28)	.62*(28)	.66*(29)	.54*(27)	.74*(33)	.84*(33)
E1	.53*(28)	.30(28)	.47*(28)	.43*(28)	.69*(29)	.44*(27)	.69*(34)	.71*(34)
E2	.31(28)	.42*(28)	.53*(28)	.71*(28)	.69*(28)	.84*(27)	.50*(32)	.75*(32)
F	.35(28)	.17(28)	.51*(28)	.55*(28)	.57*(28)	.37(27)	.56*(28)	.59*(28)
G	.10(28)	-.07(28)	.37(28)	.20(28)	.24(28)	-.06(27)	.12(28)	.17(28)
H1	__	.58*(28)	.29(28)	.50*(28)	.70*(28)	.22(27)	.74*(28)	.59*(28)
H2		__	.32(28)	.33(28)	.52*(28)	.30(27)	.47*(28)	.36(28)
I			__	.45*(28)	.38*(28)	.39*(27)	.51*(28)	.62*(28)
J				__	.62*(28)	.59*(27)	.65*(28)	.72*(28)
K					__	.67*(27)	.78*(30)	.75*(30)

L	__	.57*(27)	.62*(27)
M		__	.71*(39)
N			__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A = One-Syllable Words with Vowel Digraphs, B = One-Syllable Words with Vowel Diphthongs, C = Words with Medial Double Consonants, D = Words with Consonant Digraphs (ck, gh, ph, & wr), E1 = Two-Syllable Words with R-Controlled Vowels (real), E2 = Two-Syllable Words with R-Controlled Vowels (nonsense), F = Contractions, G = Possessives, H1 = Compound Words (real), H2 = Compound Words (nonsense), I = Two-Syllable Words with Short & Long Vowel Patterns and Inflectional Endings, J = Two-Syllable Words with Vowel Digraphs and Inflectional Endings, K = Words with Common Word Parts, L = Multisyllabic Words, M = Second-Grade High-Frequency Words, N = Sentence Reading. *$p < .05$.

Table 21

Correlations Between DIBELS Deep Word Reading and Decoding Form 4 Sections in Second Grade (Winter)

Section	A	B	C	D	E1	E2	F	G
A	__	.87*(54)	.89*(54)	.74*(49)	.78*(47)	.64*(41)	.37(39)	.47*(38)
B		__	.81*(54)	.77*(49)	.61*(47)	.51*(41)	.19(39)	.33*(38)
C			__	.73*(49)	.72*(47)	.64*(31)	.30(39)	.45*(38)
D				__	.70*(48)	.58*(42)	.33*(40)	.46*(39)
E1					__	.61*(42)	.55*(40)	.20(39)
E2						__	.52*(40)	.15(39)
F							__	.01(39)
	H1	H2	I	J	K	L	M	N
A	.72*(37)	.55*(37)	.44*(37)	.61*(37)	.67*(37)	.67*(37)	.86*(52)	.85*(52)
B	.79*(37)	.62*(37)	.45*(37)	.55*(37)	.64*(37)	.69*(37)	.78*(52)	.82*(52)
C	.66*(37)	.41*(37)	.58*(37)	.42*(37)	.75*(37)	.80*(37)	.86*(52)	.85*(52)
D	.70*(38)	.67*(38)	.57*(38)	.57*(38)	.68*(38)	.77*(38)	.61*(49)	.81*(49)
E1	.43*(38)	.28(38)	.42*(38)	.41*(38)	.67*(39)	.67*(38)	.65*(47)	.82*(47)
E2	.33(38)	.16(38)	.72*(38)	.73*(38)	.65*(38)	.64*(38)	.49*(42)	.77*(42)
F	.30(38)	.11(38)	.37*(38)	.40*(38)	.26(38)	.28(38)	.48*(40)	.52*(40)
G	.53*(38)	.24(38)	.24(38)	.12(38)	.33*(38)	.32(38)	.58*(39)	.37*(39)
H1	__	.68*(38)	.31(38)	.33(38)	.57*(38)	.60*(38)	.81*(38)	.69*(38)
H2		__	.24(38)	.30(38)	.42*(38)	.43*(38)	.34*(38)	.47*(38)
I			__	.71*(38)	.71*(38)	.54*(38)	.39*(38)	.66*(38)
J				__	.70*(38)	.54*(38)	.41*(38)	.64*(38)
K					__	.83*(38)	.60*(38)	.81*(38)

L	__	.70*(38)	.82*(38)
M		__	.76*(53)
N			__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A = One-Syllable Words with Vowel Digraphs, B = One-Syllable Words with Vowel Diphthongs, C = Words with Medial Double Consonants, D = Words with Consonant Digraphs (ck, gh, ph, & wr), E1 = Two-Syllable Words with R-Controlled Vowels (real), E2 = Two-Syllable Words with R-Controlled Vowels (nonsense), F = Contractions, G = Possessives, H1 = Compound Words (real), H2 = Compound Words (nonsense), I = Two-Syllable Words with Short & Long Vowel Patterns and Inflectional Endings, J = Two-Syllable Words with Vowel Digraphs and Inflectional Endings, K = Words with Common Word Parts, L = Multisyllabic Words, M = Second-Grade High-Frequency Words, N = Sentence Reading. *$p < .05$.

Table 22

Correlations Between DIBELS Deep Word Reading and Decoding Form 4 Sections in Third Grade (Fall)

Section	A	B	C	D	E1	E2	F	G
A	__	.63*(33)	.53*(33)	.83*(33)	.67*(33)	.64*(31)	-.12(29)	.
B		__	.72*(34)	.70*(34)	.65*(34)	.65*(32)	.21(30)	.
C			__	.73*(34)	.72*(34)	.74*(32)	.57*(30)	.
D				__	.81*(35)	.76*(32)	.21(30)	.
E1					__	.80*(32)	.12(30)	.
E2						__	.46*(30)	.
F							__	.

	H1	H2	I	J	K	L	M	N
A	.	.39*(29)	.49*(28)	.38(29)	.37*(29)	.43*(29)	.81*(31)	.61*(31)
B	.36(30)	.53*(30)	.68*(29)	.58*(30)	.44*(30)	.64*(30)	.61*(32)	.69*(32)
C	.23(30)	.66*(30)	.76*(29)	.62*(30)	.46*(30)	.58*(30)	.69*(32)	.80*(32)
D	.32(30)	.55*(30)	.77*(29)	.60*(30)	.73*(33)	.74*(30)	.81*(33)	.82*(33)
E1	.43*(30)	.59*(30)	.78*(29)	.65*(30)	.75*(30)	.81*(30)	.80*(32)	.84*(32)
E2	.32(30)	.70*(30)	.74*(29)	.58*(30)	.63*(30)	.66*(30)	.82*(30)	.82*(30)
F	.20(30)	.64*(30)	.41*(29)	.14(30)	-.07(30)	.20(30)	.48*(29)	.44*(29)
G	.	.	.	.	.	.	.	.
H1	__	.77*(30)	.31(29)	.50*(30)	.29(30)	.71*(30)	.57*(29)	.50*(29)
H2		__	.62*(29)	.59*(30)	.33(30)	.77*(30)	.82*(29)	.76*(29)
I			__	.60*(29)	.61*(29)	.71*(29)	.77*(28)	.77*(28)
J				__	.63*(30)	.77*(30)	.63*(29)	.79*(29)
K					__	.68*(30)	.58*(30)	.68*(30)

L	__	.73*(29)	.83*(29)
M		__	.78*(33)
N			__

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A = One-Syllable Words with Vowel Digraphs, B = One-Syllable Words with Vowel Diphthongs, C = Words with Medial Double Consonants, D = Words with Consonant Digraphs (ck, gh, ph, & wr), E1 = Two-Syllable Words with R-Controlled Vowels (real), E2 = Two-Syllable Words with R-Controlled Vowels (nonsense), F = Contractions, G = Possessives, H1 = Compound Words (real), H2 = Compound Words (nonsense), I = Two-Syllable Words with Short & Long Vowel Patterns and Inflectional Endings, J = Two-Syllable Words with Vowel Digraphs and Inflectional Endings, K = Words with Common Word Parts, L = Multisyllabic Words, M = Second-Grade High-Frequency Words, N = Sentence Reading. *$p < .05$.

Table 23

Correlations Between DIBELS Deep Word Reading and Decoding Form 5 Sections in Second Grade (Winter)

Section	A2	B	C	D	E	F	G	H	I
A1	.85*(52)	.69*(52)	.62*(40)	.52*(38)	.71*(36)	.67*(35)	.38*(35)	.56*(33)	.56*(33)
A2	___	.62*(52)	.44*(40)	.50*(38)	.82*(36)	.52*(35)	.22(35)	.24(36)	.38(33)
B		___	.71*(40)	.29(38)	.34*(36)	.51*(35)	.66*(35)	.65*(36)	.51*(33)
C			___	.43*(38)	.44*(36)	.59*(35)	.63*(35)	.57*(36)	.72*(33)
D				___	.62*(36)	.35*(35)	.40*(35)	.27(35)	.34(33)
E					___	.67*(35)	.31(35)	.39*(35)	.52*(33)
F						___	.45*(35)	.61*(35)	.51*(33)
G							___	.53*(35)	.54*(33)
H								___	.53*(33)
	J	K	L	M	N	O	P	Q	R
A1	.62*(29)	.60*(29)	.55*(29)	.55*(27)	.48*(25)	.61*(24)	.41(20)	.81*(50)	.65*(49)
A2	.61*(29)	.39*(29)	.41*(29)	.48*(27)	.45*(25)	.40(24)	.31(20)	.82*(50)	.59*(49)
B	.47*(29)	.78*(29)	.60*(29)	.63*(27)	.58*(25)	.48*(24)	.55*(20)	.63*(50)	.78*(49)
C	.63*(29)	.76*(29)	.46*(29)	.57*(27)	.63*(25)	.62*(24)	.57*(20)	.46*(38)	.68*(38)
D	.22(29)	.47*(29)	.49*(29)	.33(27)	.22(24)	.04(23)	-	.65*(36)	.37*(36)
E	.52*(29)	.44*(29)	.51*(29)	.46*(27)	.38(24)	.51*(23)	-	.76*(34)	.48*(34)
F	.75*(29)	.51*(29)	.46*(29)	.57*(27)	.75*(24)	.66*(23)	-	.73*(33)	.63*(33)
G	.29(29)	.53*(29)	.74*(29)	.46*(27)	.47*(24)	.29(23)	-	.27(33)	.57*(33)
H	.45*(29)	.50*(29)	.49*(29)	.55*(27)	.59*(25)	.56*(24)	.57*(20)	.45*(35)	.76*(35)
I	.58*(29)	.52*(29)	.20(29)	.50*(27)	.45*(24)	.63*(23)	-	.52*(32)	.58*(32)
J	___	.61*(29)	.40*(29)	.56*(27)	.40(24)	.52*(23)	-	.78*(28)	.55*(28)
K		___	.59*(29)	.65*(27)	.60*(24)	.45*(23)	-	.67*(28)	.67*(28)
L			___	.59*(27)	.43*(24)	.20(23)	-	.50*(28)	.61*(28)

M	___	.71*(24)	.77*(23)	-	.60*(26)	.71*(26)
N		___	.71*(24)	.55*(20)	.60*(24)	.70*(24)
O			___	.56*(20)	.60*(23)	.58*(23)
P				___	.64*(20)	.70*(20)
Q					___	.51*(50)
R						___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. Data are not reported in cases where n < 20. A1 = Compound Words (real); A2 = Compound Words (nonsense); B = Words with Common Word Parts; C = Multi-Syllabic Words; D = Contractions; E = Possessives; F = Variant Plurals; G = CCCVC, CCCVCC, & CCCVCCC Words with Inflectional Endings; H = Words with "ch" Pronounced as /k/; I = Words with Low-Frequency Vowel Patterns (ou, er, ow); J = Words with Silent Letters; K = Two-Syllable Words with Diphthongs; L = Words with Consonant Trigraphs; M = Words with ive, ous, ious Endings; N = Words with Low-Frequency Vowel Teams (ie, ai, ei); O = Words with Low-Frequency Vowel Teams (oo, ea); P = Words with Low-Frequency Vowel Teams (ea, eau); Q = Third-Grade High-Frequency Words, R = Sentence Reading. *p < .05.

Table 24

Correlations Between DIBELS Deep Word Reading and Decoding Form 5 Sections in Third Grade (Fall)

Section	A2	B	C	D	E	F	G	H	I
A1	.82*(37)	.82*(37)	.59*(31)	.25(31)	.37*(31)	.70*(31)	.57*(31)	.41*(31)	.43*(29)
A2	___	.76*(37)	.70*(31)	.16(31)	.59*(31)	.73*(31)	.37*(31)	.41*(31)	.62*(29)
B		___	.	.	.55*(31)	.52*(31)	.66*(31)	.45*(33)	.52*(29)
C			___	.	.54*(31)	.69*(31)	.51*(31)	.52*(32)	.71*(29)
D				___	.20(31)	.26(31)	.26(31)	.29(31)	.14(29)
E					___	.51*(31)	.48*(31)	.19(31)	.50*(29)
F						___	.56*(31)	.50*(31)	.59*(29)
G							___	.57*(31)	.59*(29)
H								___	.57*(29)
	J	K	L	M	N	O	P	Q	R
A1	.02(28)	.00(28)	-.03(28)	.26(28)	.46*(27)	.45*(27)	.45*(26)	.81*(34)	.68*(34)
A2	.18(28)	.15(28)	.09(28)	.34(28)	.24(27)	.16(27)	.14(26)	.95*(34)	.60*(34)
B	.19(29)	.48*(28)	.42*(28)	.59*(28)	.31(28)	.13(27)	.56*(27)	.74*(36)	.69*(36)
C	.36(29)	.49*(28)	.43*(28)	.62*(28)	.39*(28)	.12(27)	.51*(28)	.67*(31)	.60*(31)
D	.10(28)	-.06(28)	-.07(28)	.18(28)	.29(27)	.29(27)	.53*(26)	-.18(29)	.30(29)
E	.28(28)	.24(28)	.25(28)	.47*(28)	-.02(27)	-.03(27)	.31(26)	.59*(29)	.25(29)
F	.22(28)	.06(28)	.21(28)	.36*(28)	.47*(27)	.45*(27)	.46*(26)	.65*(29)	.45*(29)
G	.42*(28)	.36(28)	.50*(28)	.50*(28)	.60*(27)	.39*(27)	.66*(26)	.51*(29)	.58*(29)
H	.17(29)	.27(28)	.30(28)	.37(28)	.60*(28)	.50*(27)	.57*(28)	.32(32)	.69*(32)
I	.48*(28)	.25(28)	.58*(28)	.61*(28)	.70*(27)	.38*(27)	.62*(26)	.50*(28)	.56*(28)
J	___	.62*(28)	.49*(28)	.54*(28)	.57*(28)	.27(27)	.30(27)	.22(28)	.25(28)
K		___	.48*(28)	.52*(28)	.48*(27)	.16(27)	.21(26)	.39*(27)	.44*(27)
L			___	.57*(28)	.45*(27)	.29(27)	.44*(26)	.43*(27)	.06(27)

M	___	.51*(27)	.54*(27)	.58*(26)	.43*(27)	.38(27)
N		___	.61*(27)	.51*(27)	.38*(27)	.61*(27)
O			___	.29(26)	.42*(26)	.34(26)
P				___	.23(29)	.51*(29)
Q					___	.54*(38)
R						___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = Compound Words (real); A2 = Compound Words (nonsense); B = Words with Common Word Parts; C = Multi-Syllabic Words; D = Contractions; E = Possessives; F = Variant Plurals; G = CCCVC, CCCVCC, & CCCVCCC Words with Inflectional Endings; H = Words with "ch" Pronounced as /k/; I = Words with Low-Frequency Vowel Patterns (ou, er, ow); J = Words with Silent Letters; K = Two-Syllable Words with Diphthongs; L = Words with Consonant Trigraphs; M = Words with ive, ous, ious Endings; N = Words with Low-Frequency Vowel Teams (ie, ai, ei); O = Words with Low-Frequency Vowel Teams (oo, ea); P = Words with Low-Frequency Vowel Teams (ea, eau); Q = Third-Grade High-Frequency Words, R = Sentence Reading. *$p < .05$.

Table 25

Correlations Between DIBELS Deep Word Reading and Decoding Form 5 Sections in Third Grade (Winter)

Section	A2	B	C	D	E	F	G	H	I
A1	.76*(53)	.72*(53)	.71*(52)	.28*(52)	.64*(52)	.50*(51)	.48*(51)	.48*(51)	.63*(51)
A2	___	.66*(53)	.65*(52)	.38*(52)	.67*(52)	.51*(51)	.33*(51)	.46*(51)	.65*(51)
B		___	.78*(52)	.39*(52)	.66*(52)	.55*(51)	.65*(51)	.57*(51)	.56*(51)
C			___	.30*(52)	.68*(52)	.57*(51)	.64*(51)	.61*(51)	.61*(51)
D				___	.41*(52)	.26(51)	.26(51)	.22(51)	.36*(51)
E					___	.34*(51)	.55*(51)	.35*(51)	.50*(51)
F						___	.40*(51)	.49*(51)	.51*(51)
G							___	.59*(51)	.54*(51)
H								___	.53*(53)
	J	K	L	M	N	O	P	Q	R
A1	.44*(49)	.45*(49)	.49*(49)	.44*(49)	.53*(47)	.46*(45)	.35*(43)	.57*(52)	.60*(51)
A2	.43*(49)	.36*(49)	.36*(49)	.36*(49)	.39*(47)	.28(45)	.29(43)	.84*(52)	.53*(51)
B	.63*(49)	.59*(49)	.62*(49)	.75*(49)	.69*(47)	.56*(45)	.52*(43)	.63*(52)	.81*(51)
C	.61*(49)	.58*(49)	.67*(49)	.64*(49)	.61*(47)	.55*(45)	.57*(43)	.61*(52)	.75*(51)
D	.11(49)	.21(49)	.03(49)	.26(49)	.41*(47)	.48*(45)	.38*(43)	.36*(52)	.36*(51)
E	.31*(49)	.48*(49)	.21(49)	.37*(49)	.44*(47)	.55*(45)	.31*(43)	.66*(52)	.54*(51)
F	.47*(49)	.42*(49)	.50*(49)	.42*(49)	.40*(47)	.31*(45)	.32*(43)	.61*(51)	.64*(51)
G	.59*(49)	.59*(49)	.65*(49)	.70*(49)	.62*(47)	.76*(45)	.67*(43)	.34*(51)	.71*(51)
H	.47*(51)	.61*(51)	.67*(51)	.66*(51)	.65*(50)	.60*(47)	.65*(46)	.41*(54)	.71*(54)
I	.29*(51)	.20(51)	.33*(51)	.52*(51)	.62*(49)	.41*(47)	.41*(45)	.50*(53)	.65*(53)
J	___	.61*(51)	.67*(51)	.69*(51)	.59*(40)	.55*(47)	.45*(46)	.52*(52)	.60*(52)
K		___	.69*(51)	.58*(51)	.65*(49)	.72*(47)	.55*(45)	.35*(51)	.69*(51)
L			___	.63*(51)	.63*(49)	.68*(47)	.65*(45)	.36*(51)	.68*(51)

M	___	.66*(49)	.63*(47)	.51*(45)	.36*(51)	.70*(51)
N		___	.73*(47)	.65*(46)	.40*(50)	.79*(50)
O			___	.70*(45)	.22(47)	.66*(47)
P				___	.26(47)	.56*(47)
Q					___	.56*(55)
R						___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = Compound Words (real); A2 = Compound Words (nonsense); B = Words with Common Word Parts; C = Multi-Syllabic Words; D = Contractions; E = Possessives; F = Variant Plurals; G = CCCVC, CCCVCC, & CCCVCCC Words with Inflectional Endings; H = Words with "ch" Pronounced as /k/; I = Words with Low-Frequency Vowel Patterns (ou, er, ow); J = Words with Silent Letters; K = Two-Syllable Words with Diphthongs; L = Words with Consonant Trigraphs; M = Words with ive, ous, ious Endings; N = Words with Low-Frequency Vowel Teams (ie, ai, ei); O = Words with Low-Frequency Vowel Teams (oo, ea); P = Words with Low-Frequency Vowel Teams (ea, eau); Q = Third-Grade High-Frequency Words, R = Sentence Reading. *$p < .05$.

Table 26

Correlations Between DIBELS Deep Word Reading and Decoding Form 5 Sections in Fourth Grade (Fall)

Section	A2	B	C	D	E	F	G	H	I
A1	.76*(38)	.65*(37)	.33*(36)	.06(36)	.05(36)	.46*(37)	.18(37)	.30(36)	.47*(36)
A2	___	.48*(37)	-.08(36)	-.11(36)	-.08(36)	.14(37)	.24(37)	.08(37)	.09(36)
B		___	.80*(36)	.39*(36)	.34*(36)	.68*(36)	.59*(36)	.65*(36)	.82*(36)
C			___	.31(36)	.36*(36)	.57*(36)	.67*(36)	.58*(36)	.66*(36)
D				___	.36*(36)	.25(36)	.47*(36)	.21(36)	.33(36)
E					___	.10(36)	.37*(36)	.25(36)	.25(36)
F						___	.44*(37)	.59*(36)	.62*(36)
G							___	.62*(36)	.51*(36)
H								___	.54*(37)
	J	K	L	M	N	O	P	Q	R
A1	.21(35)	.40*(35)	.33(35)	.44*(35)	.20(35)	.05(34)	.28(35)	.26(37)	.26(37)
A2	.15(35)	.19(35)	.04(35)	.24(35)	.21(35)	.08(34)	-.10(35)	-.14(37)	.08(37)
B	.49*(35)	.80*(35)	.61*(35)	.75*(35)	.66*(35)	.62*(34)	.48*(34)	.78*(36)	.64*(36)
C	.31(35)	.58*(35)	.48*(35)	.37*(35)	.44*(35)	.31(34)	.48*(34)	.77*(36)	.51*(36)
D	.45*(35)	.42*(35)	.41*(35)	.27(35)	.15(35)	.14(34)	.34(34)	.22(36)	.33*(36)
E	.19(35)	.26(35)	.44*(35)	.06(35)	.13(35)	.19(34)	.54*(34)	.24(36)	.45*(36)
F	.43*(35)	.58*(35)	.34*(35)	.58*(35)	.57*(35)	.30(34)	.23(35)	.55*(37)	.54*(37)
G	.48*(35)	.55*(35)	.57*(35)	.52*(35)	.53*(35)	.67*(34)	.49*(35)	.47*(37)	.73*(37)
H	.47*(36)	.51*(36)	.65*(37)	.57*(36)	.64*(36)	.60*(35)	.58*(36)	.66*(38)	.73*(38)
I	.48*(36)	.68*(36)	.46*(36)	.73*(36)	.46*(36)	.52*(35)	.43*(35)	.52*(37)	.61*(37)
J	___	.43*(36)	.43*(36)	.47*(36)	.35*(36)	.37*(35)	.54*(35)	.30(36)	.57*(36)
K		___	.70*(36)	.70*(36)	.65*(36)	.55*(35)	.28(35)	.52*(36)	.57*(36)
L			___	.56*(36)	.58*(36)	.53*(35)	.61*(36)	.49*(37)	.61*(37)

M	___	.66*(36)	.59*(35)	.33(35)	.38*(36)	.72*(36)
N		___	.53*(35)	.24(35)	.59*(36)	.63*(36)
O			___	.45*(35)	.42*(35)	.66*(35)
P				___	.30(38)	.55*(38)
Q					___	.56*(41)
R						___

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. A1 = Compound Words (real); A2 = Compound Words (nonsense); B = Words with Common Word Parts; C = Multi-Syllabic Words; D = Contractions; E = Possessives; F = Variant Plurals; G = CCCVC, CCCVCC, & CCCVCCC Words with Inflectional Endings; H = Words with "ch" Pronounced as /k/; I = Words with Low-Frequency Vowel Patterns (ou, er, ow); J = Words with Silent Letters; K = Two-Syllable Words with Diphthongs; L = Words with Consonant Trigraphs; M = Words with ive, ous, ious Endings; N = Words with Low-Frequency Vowel Teams (ie, ai, ei); O = Words with Low-Frequency Vowel Teams (oo, ea); P = Words with Low-Frequency Vowel Teams (ea, eau); Q = Third-Grade High-Frequency Words, R = Sentence Reading. *$p < .05$.

Table 27

Teacher Usability Questionnaire Ratings

Item	N	Mean Rating (SD)
1. The measures adequately covered the reading skills in the grade level I teach.	31	4.6 (1.3)
2. Most teachers would find the measures appropriate for assessing reading difficulties.	31	4.7 (1.1)
3. I believe the measures would be helpful in planning instruction for phonemic awareness.	19	5.2 (0.63)
4. I believe the measures would be helpful in planning instruction for phonics (alphabetic principle).	26	5.2 (0.65)
5. I would suggest the use of the measures to other teachers.	30	4.7 (1.3)
6. I would be willing to use the measures in my classroom.	31	4.8 (1.3)
7. I liked the procedures used for the measures.	31	4.7 (0.77)
8. The measures were a good way to assess students' reading strengths and weaknesses.	30	4.8 (1.17)
9. Overall, the measures would be beneficial for planning reading instruction.	31	4.7 (1.0)

Note. 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, 6 = Strongly Agree

Table 28

Assessor Usability Questionnaire Ratings

Item	N	Mean Rating (SD)
1. The administration and scoring rules were easy to follow.	16	3.9 (0.99)
2. The materials were organized appropriately for efficient administration of the measures.	16	4.3 (1.0)
3. I believe that the number, type, and sequence of the items were sufficient to ensure that the students understood the task.	16	5.0 (0.51)
4. I believe that the tasks were appropriate to the age/grade level of the students I tested.	16	4.9 (0.57)
5. All items included within the measure were appropriate (e.g., all words seemed at the appropriate level, passages were of equivalent difficulty).	16	4.4 (0.89)
6. I believe that the scores obtained from the measure accurately reflect students' skill level.	16	4.9 (0.89)
7. I would suggest the use of the measures to others.	12	4.7 (0.98)
8. The measures were a good way to assess students' reading strengths and weaknesses.	14	4.8 (1.1)
9. Overall, the measures would be beneficial for planning reading instruction.	12	4.8 (1.3)

Note. 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, 6 = Strongly Agree

	Kindergarten		First Grade		Second Grade		Third Grade		Fourth Grade
Measure/Month	Nov/Dec	Jan/Feb	Nov/Dec	Jan/Feb	Nov/Dec	Jan/Feb	Nov/Dec	Jan/Feb	Nov/Dec
DEEP PA1	***X***								
DEEP PA2	***X***	***X***	***X***						
DEEP WRD SCREEN	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***
DEEP WRD1	***X***	***X***	***X***						
DEEP WRD2			***X***	***X***					
DEEP WRD3				***X***	***X***				
DEEP WRD4				***X***	***X***	***X***	***X***		
DEEP WRD5						***X***	***X***	***X***	***X***
Teacher & Examiner Questionnaires		***X***		***X***		***X***		***X***	***X***

Key: ***X*** = Deep measure **scheduled** to be given
= Deep measure **may be** given dependent on student skill level
Shaded cells = Deep measure not given

Figure 1. Fall and Winter DIBELS® Deep Measures & Questionnaires Data Collection Schedule by Grade Level

	Kindergarten		First Grade		Second Grade		Third Grade	
Measure/Month	Jan/Feb	Spring	Jan/Feb	Spring	Jan/Feb	Spring	Jan/Feb	Spring
DEEP PA1								
DEEP PA2	***X***							
DEEP WRD SCREEN	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***
DEEP WRD1	***X***	***X***						
DEEP WRD2			***X***	***X***				
DEEP WRD3			***X***	***X***				
DEEP WRD4			***X***	***X***	***X***	***X***		
DEEP WRD5					***X***	***X***	***X***	***X***
Teacher & Examiner Questionnaires		***X***		***X***		***X***		***X***

Key: ***X*** = Deep measure **scheduled** to be given
= Deep measure **may be** given dependent on student skill level
Shaded cells = Deep measure not given

Figure 2. Winter and Spring DIBELS® Deep Measures & Questionnaires Data Collection Schedule by Grade Level

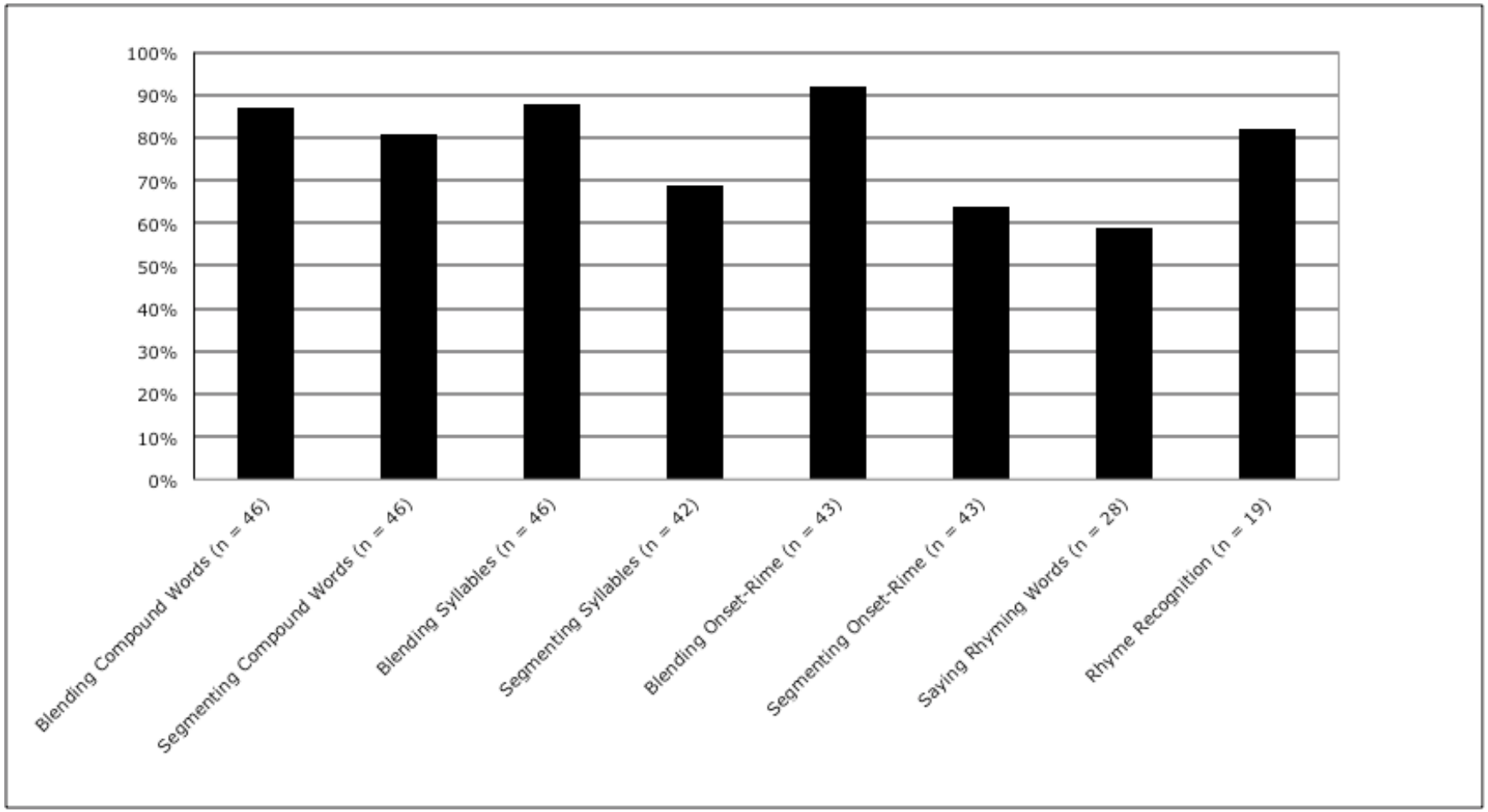

Figure 3. Average Item Percent Correct by Section: Phonemic Awareness Form 1 Fall Administration (Kindergarten)

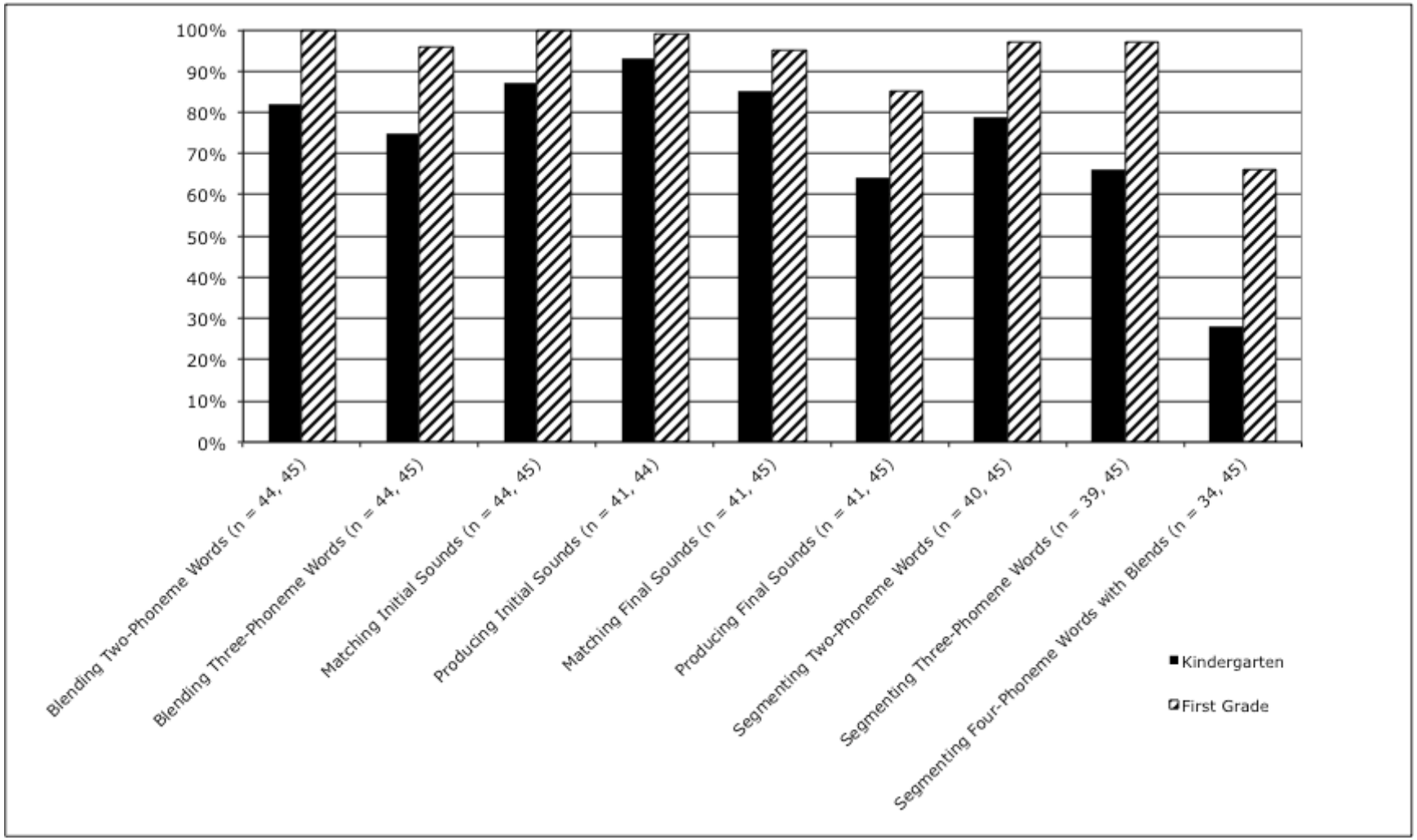

Figure 4. Average Item Percent Correct by Section: Phonemic Awareness Form 2 Fall Administration

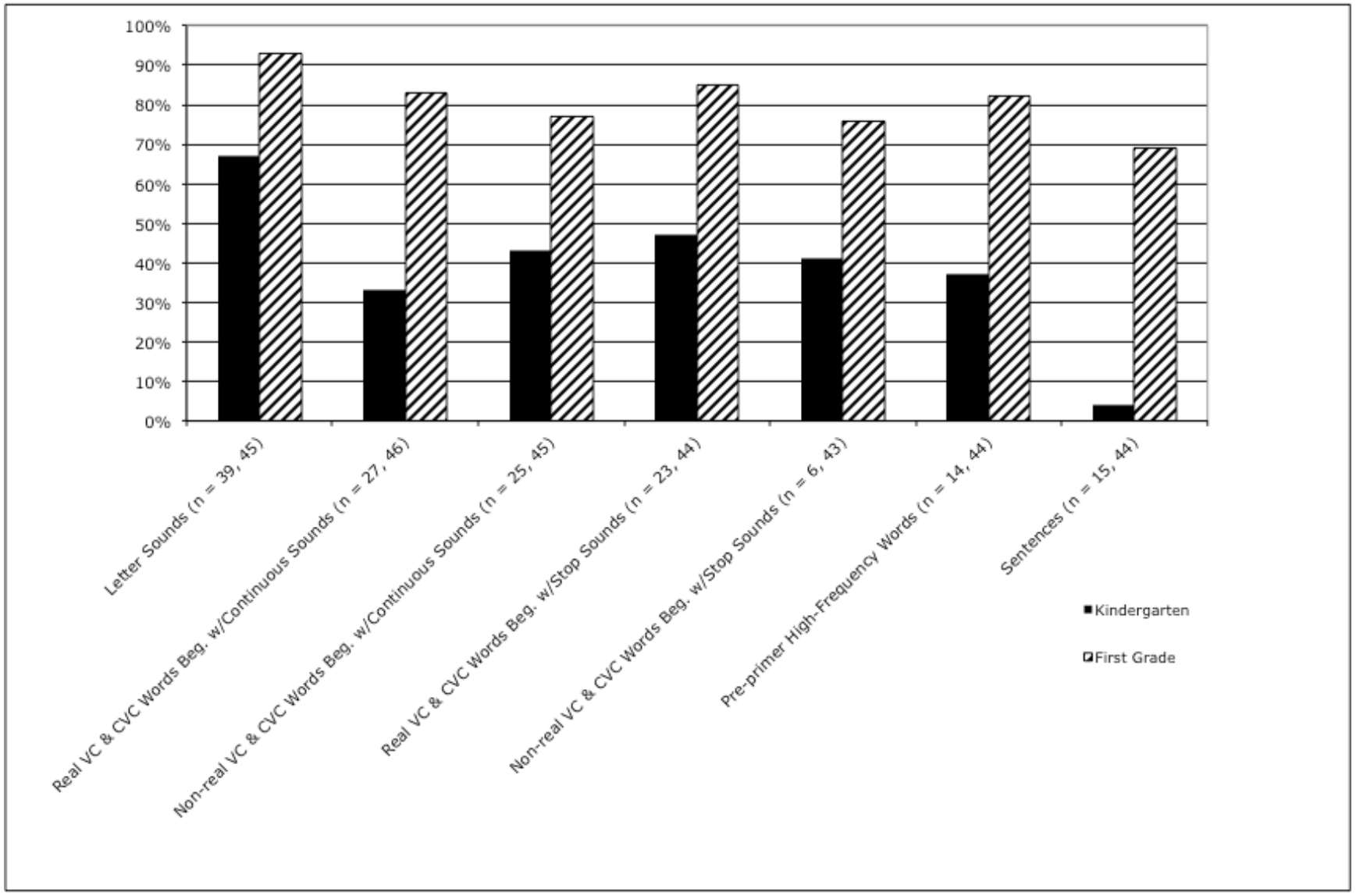

Figure 5. Average Item Percent Correct by Section: Word Reading and Decoding Form 1 Fall Administration

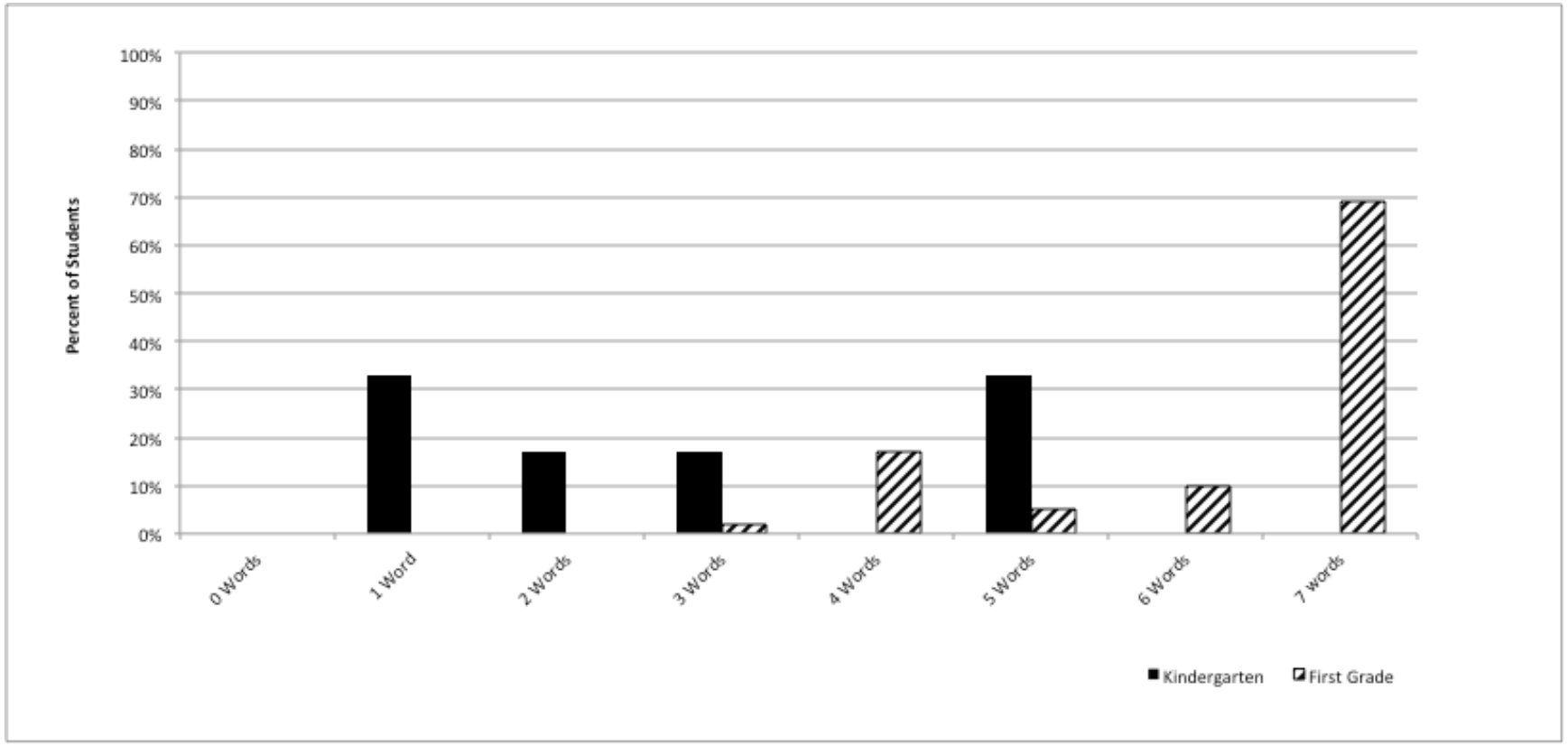

Figure 6. Percent of Students Earning Different Words Correct Scores on Word Reading and Decoding Form 1 Sentence #5

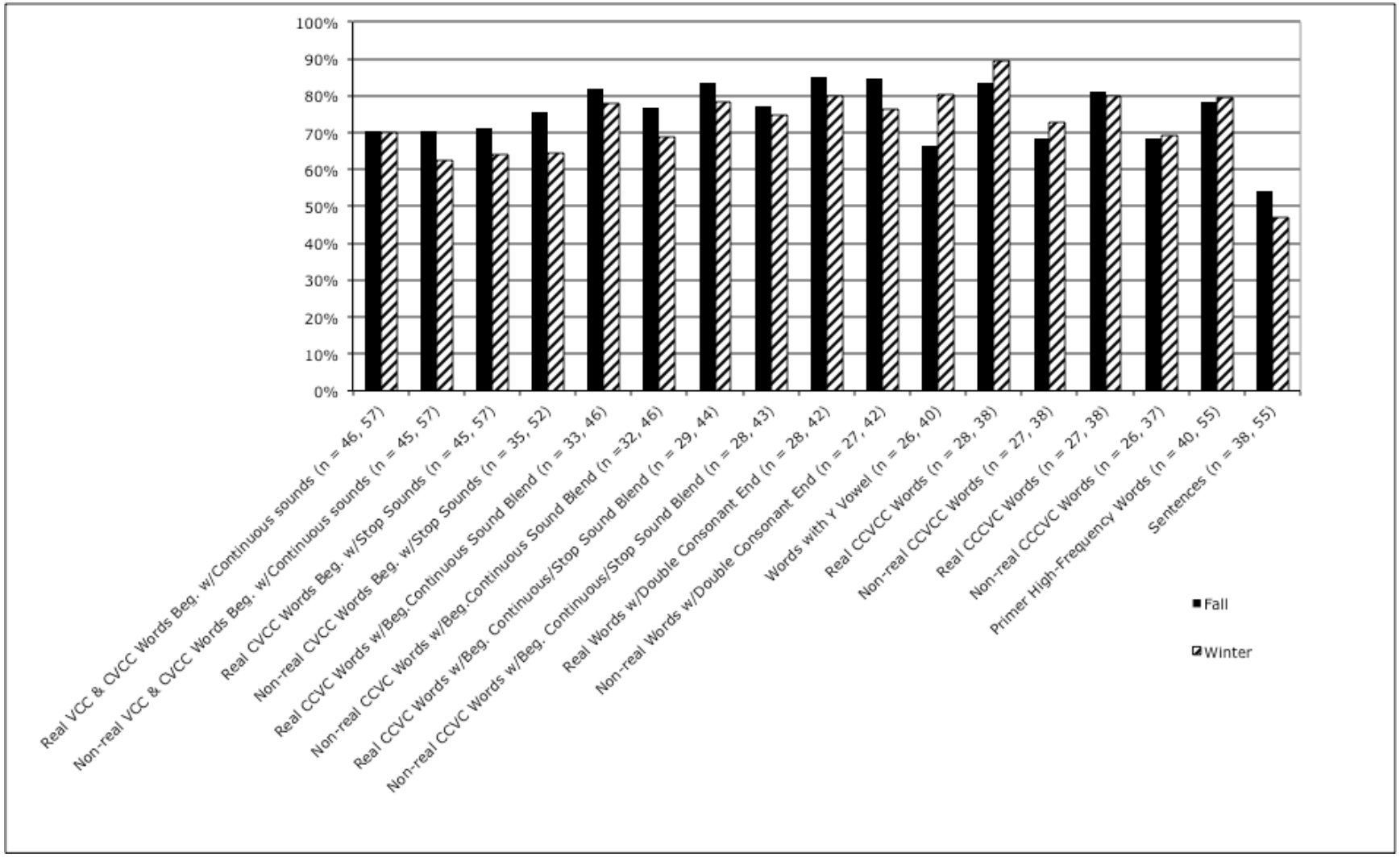

Figure 7. Average Item Percent Correct by Section: Word Reading and Decoding Form 2 First Grade

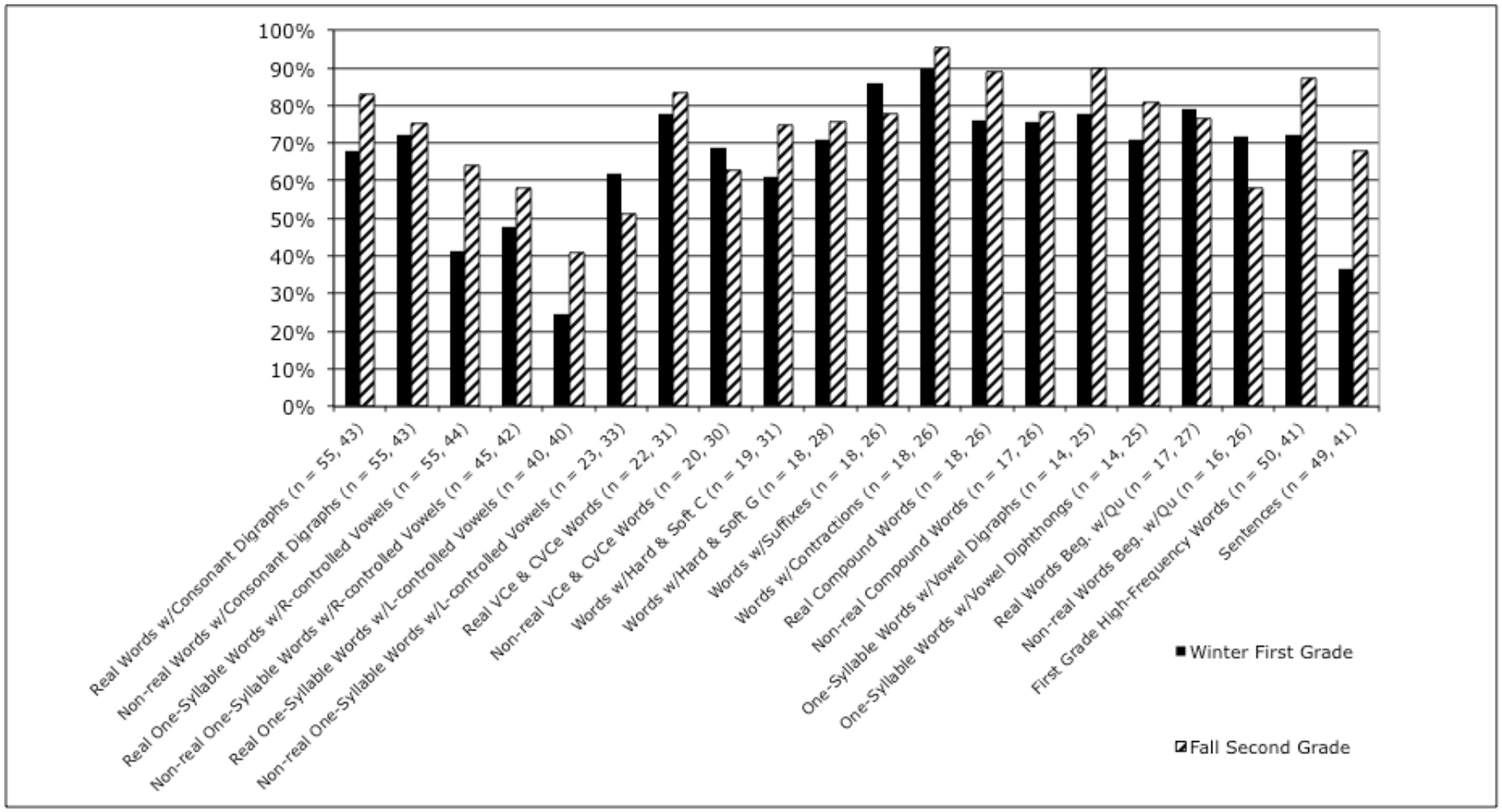

Figure 8. Average Item Percent Correct by Section: Word Reading and Decoding Form 3 Winter Grade 1 & Fall Grade 2

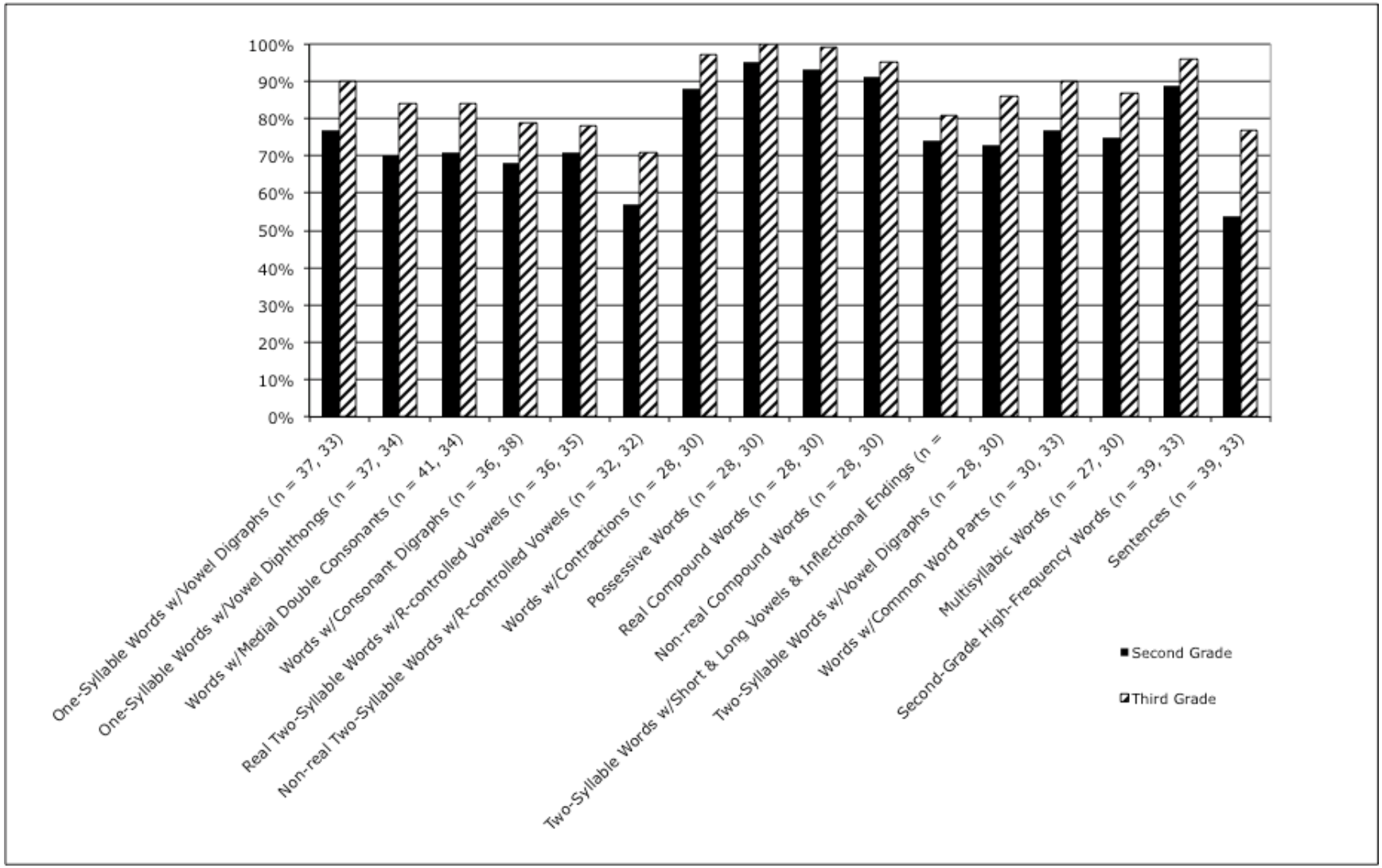

Figure 9. Average Item Percent Correct by Section: Word Reading and Decoding Form 4 Fall Administration

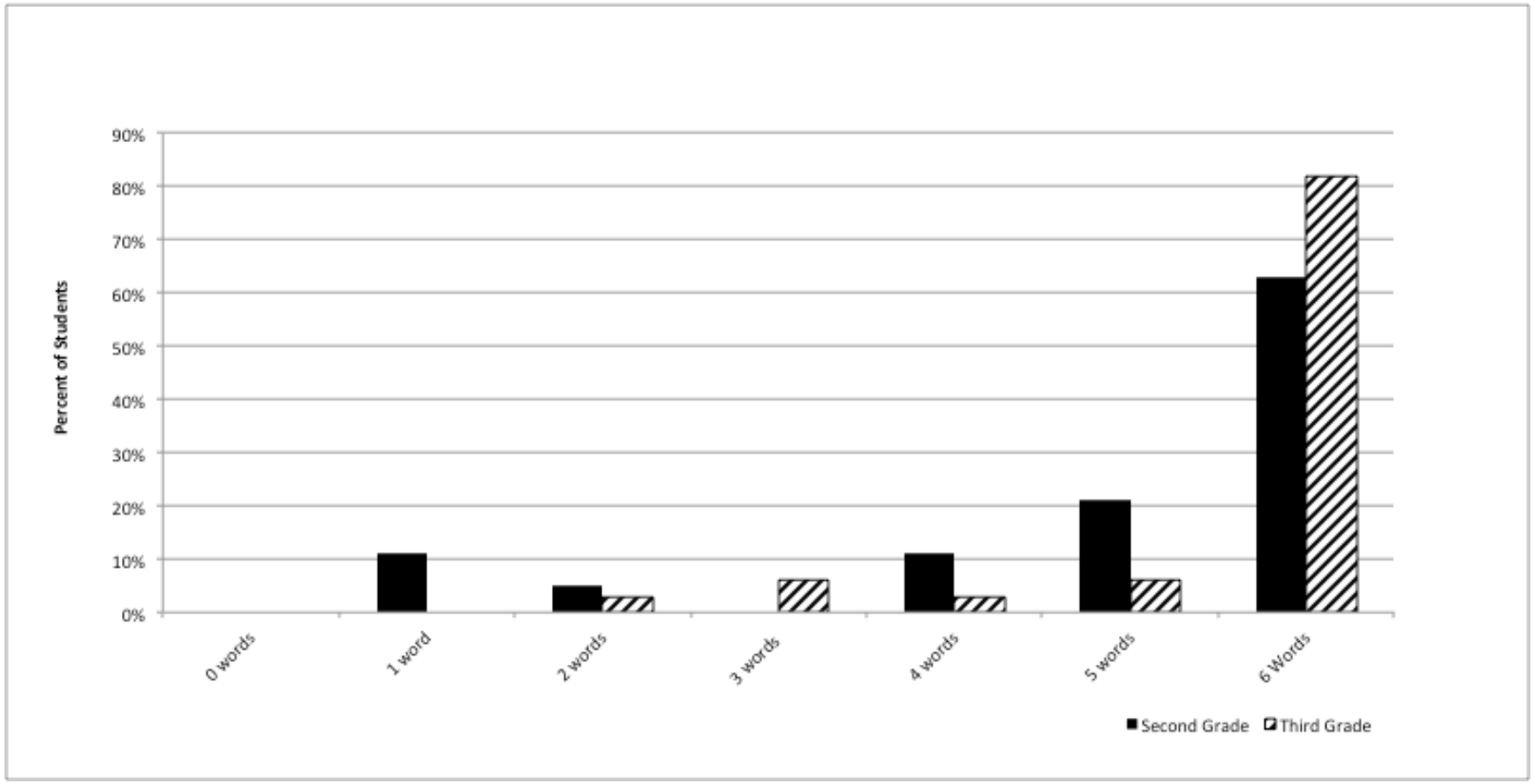

Figure 10. Percent of Students Earning Different Words Correct Scores on Word Reading and Decoding Form 4 Sentence #5

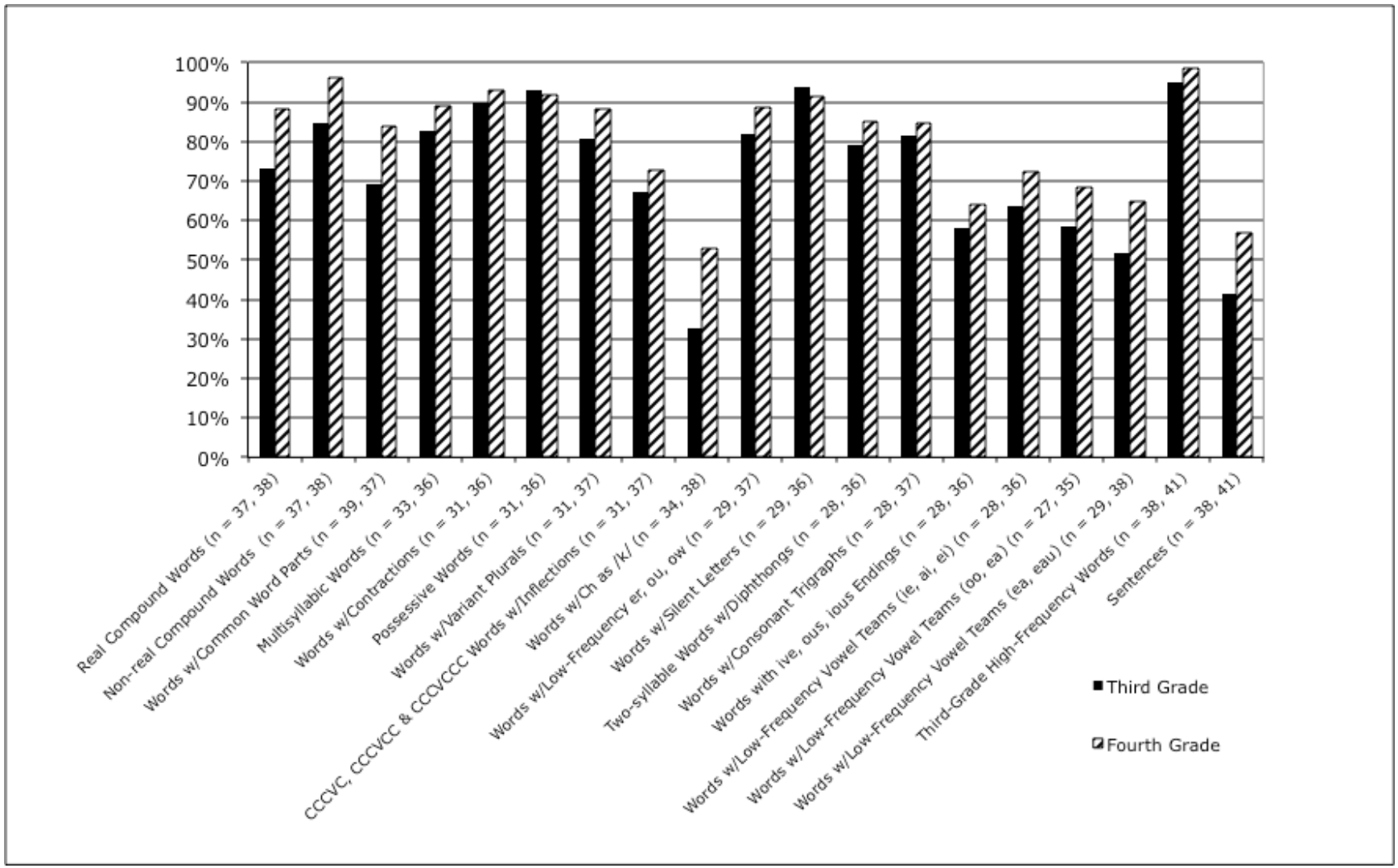

Figure 11. Average Item Percent Correct by Section: Word Reading and Decoding Form 5 Fall Administration

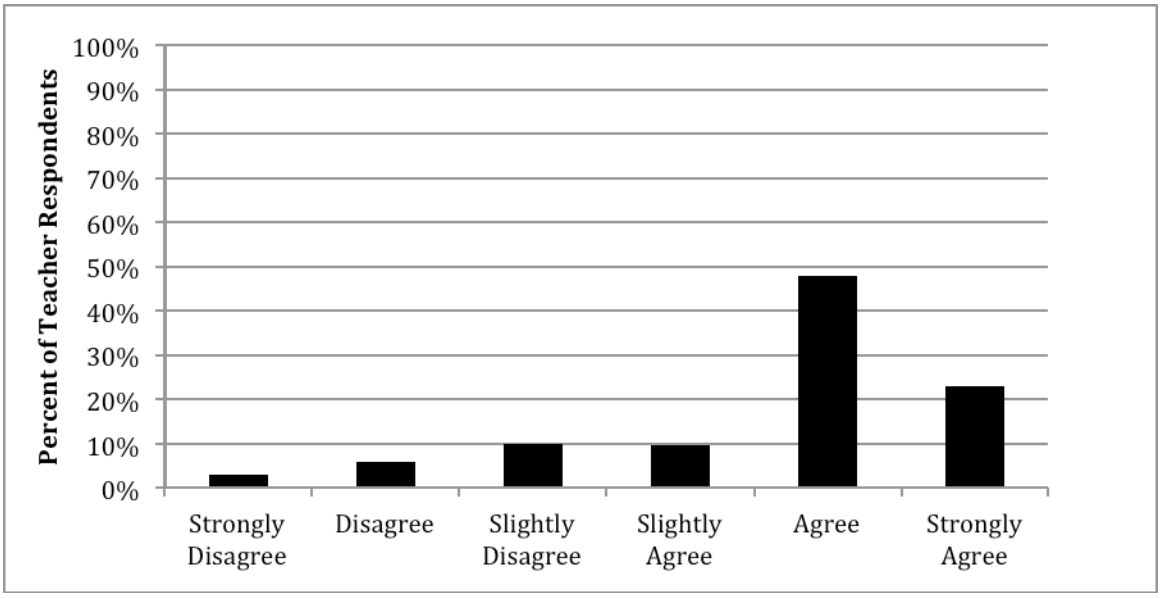

Figure 12. Teacher Ratings of Item 1: The Measures Adequately Covered the Reading Skills in the Grade Level I Teach (n = 31)

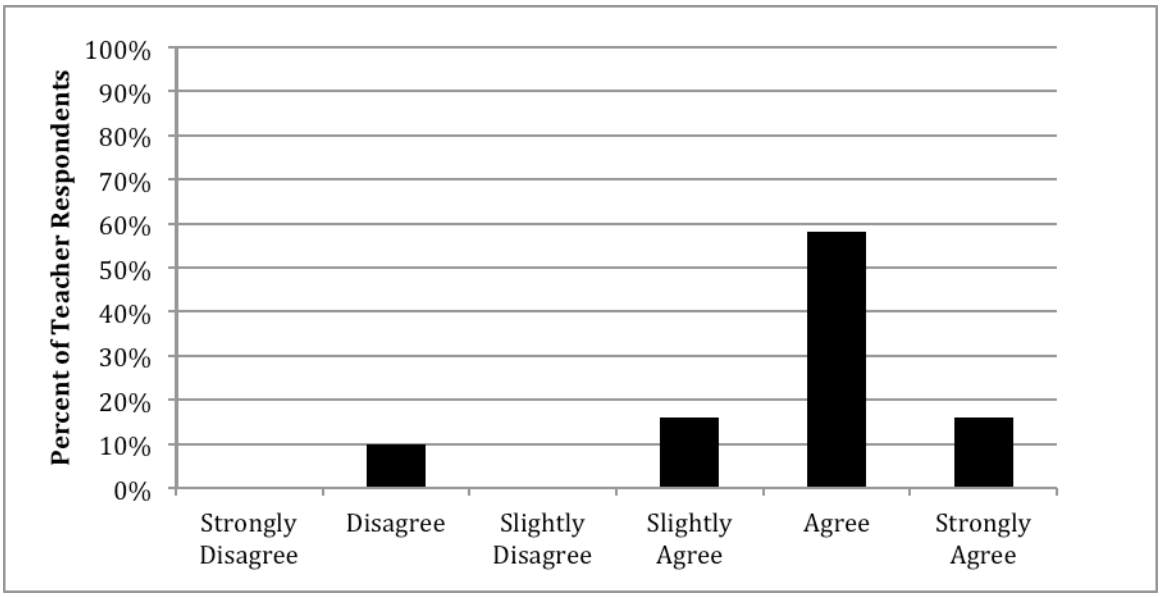

Figure 13. Teacher Ratings of Item 2: Most Teachers Would Find the Measures Appropriate for Assessing Reading Difficulties (n = 31)

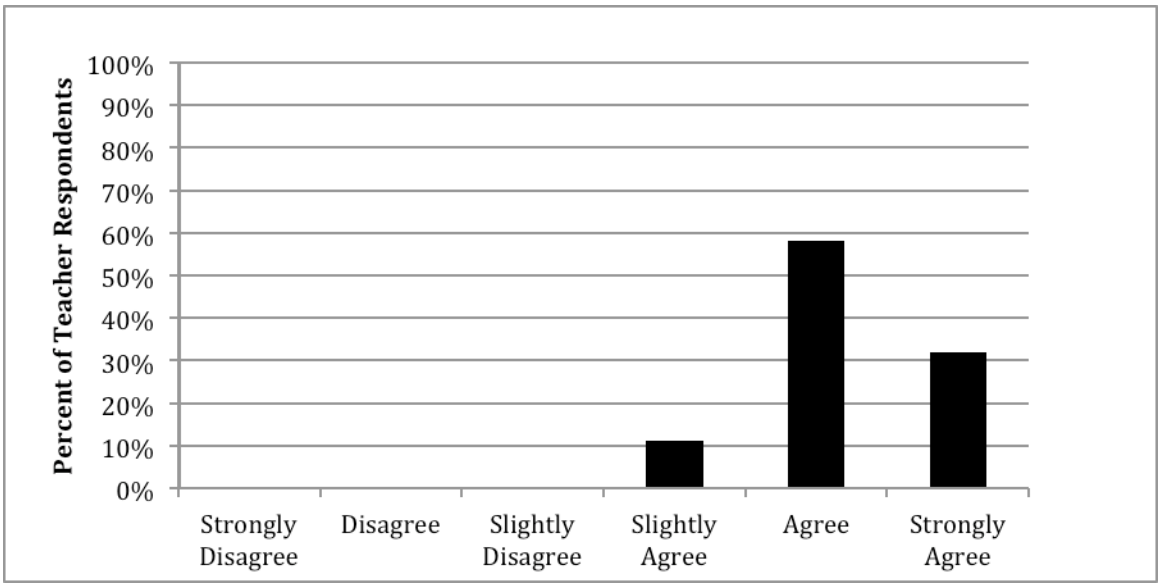

Figure 14. Teacher Ratings of Item 3: I Believe the Measures Would Be Helpful in Planning Reading Instruction for Phonemic Awareness (n = 19)

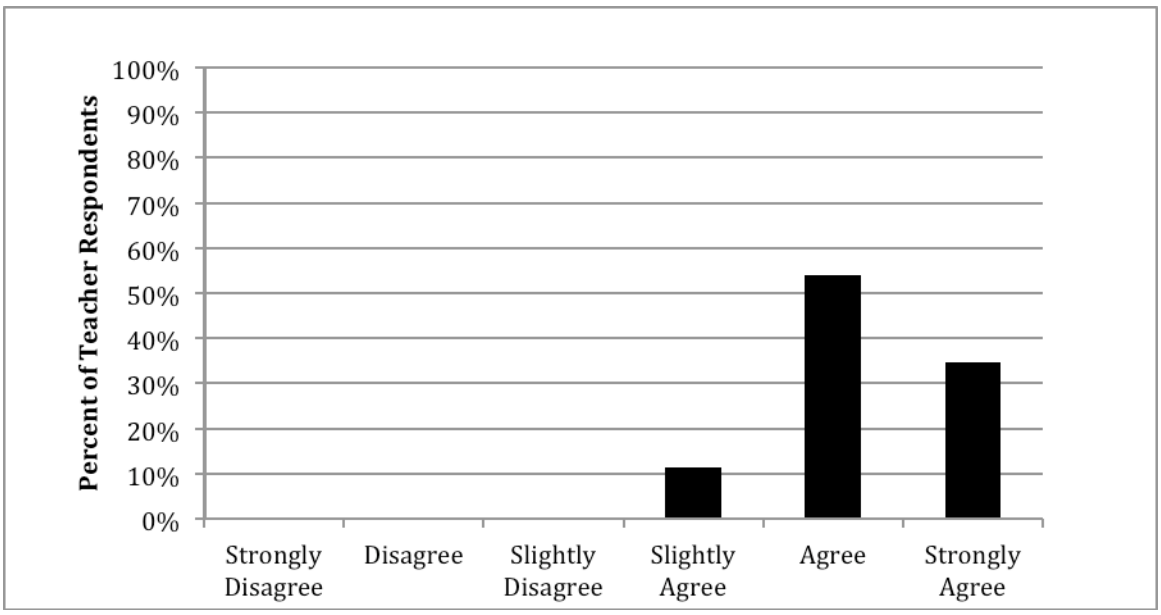

Figure 15. Teacher Ratings of Item 4: I Believe the Measures Would Be Helpful in Planning Reading Instruction for Phonics (Word Reading and Decoding) (n = 26)

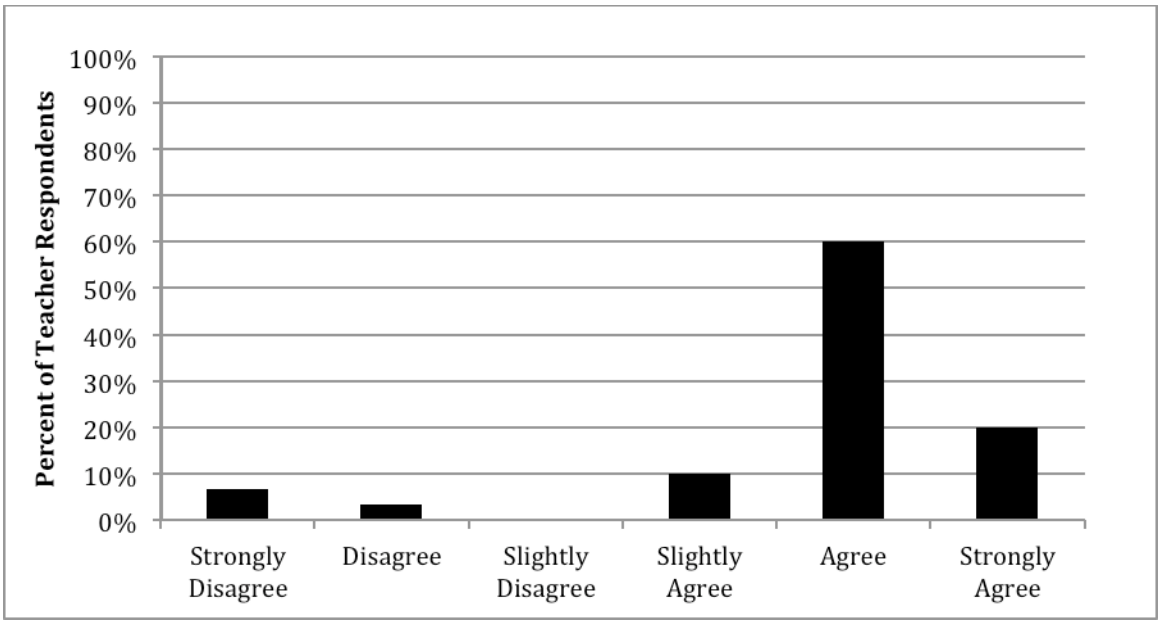

Figure 16. Teacher Ratings of Item 5: I Would Suggest the Use of the Measures to Other Teachers (n = 30)

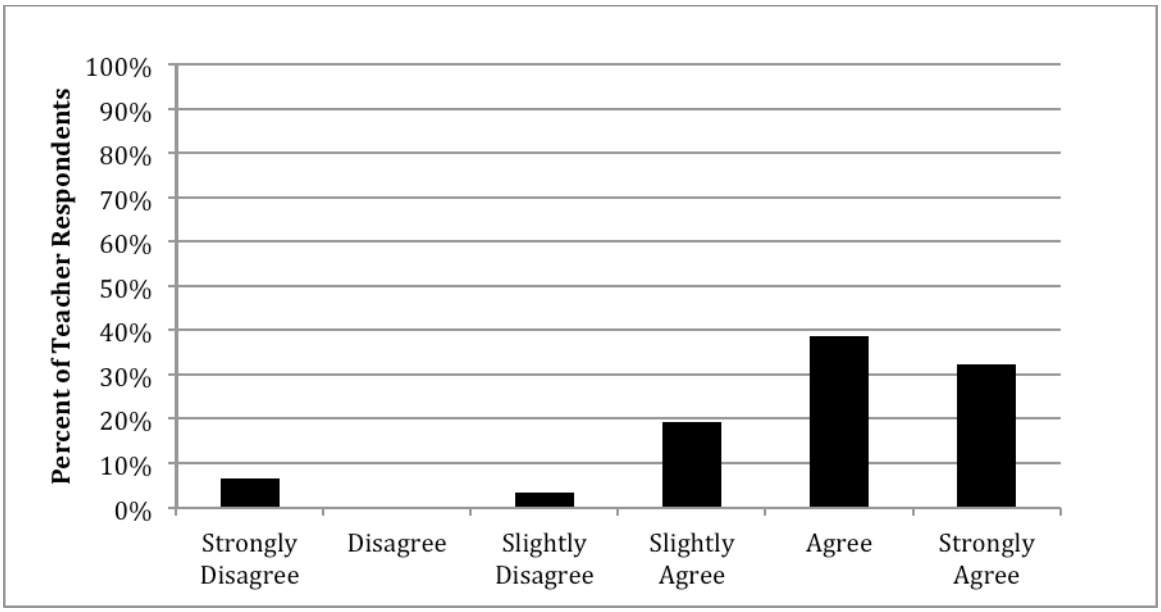

Figure 17. Teacher Ratings of Item 6: I Would Be Willing to Use the Measures in My Classroom (n = 31)

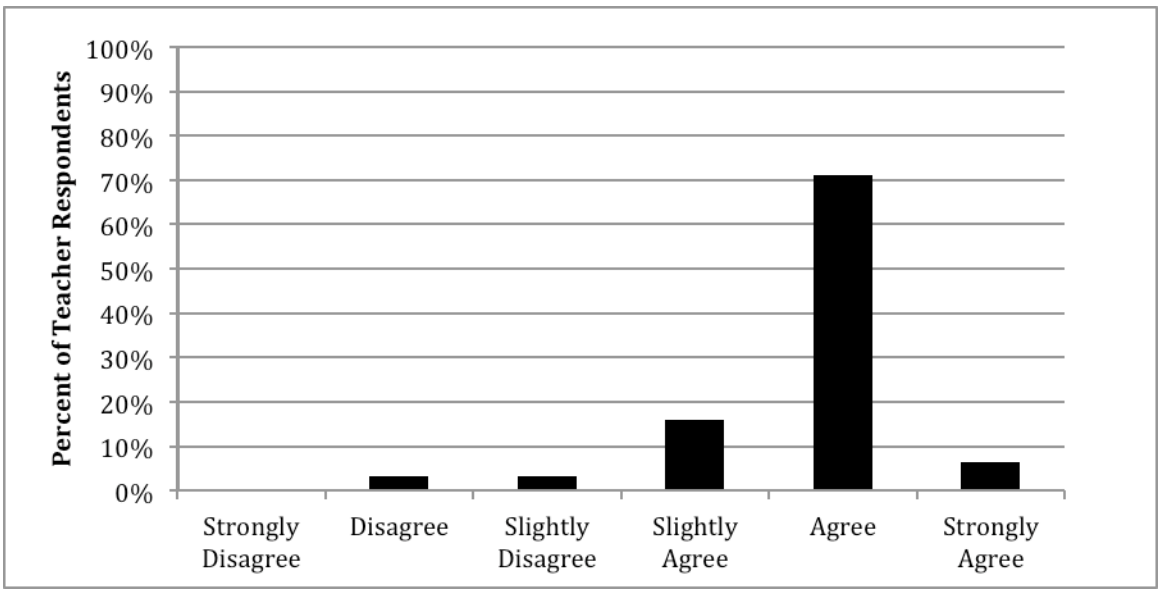

Figure 18. Teacher Ratings of Item 7: I Like the Procedures Used for the Measures (n = 31)

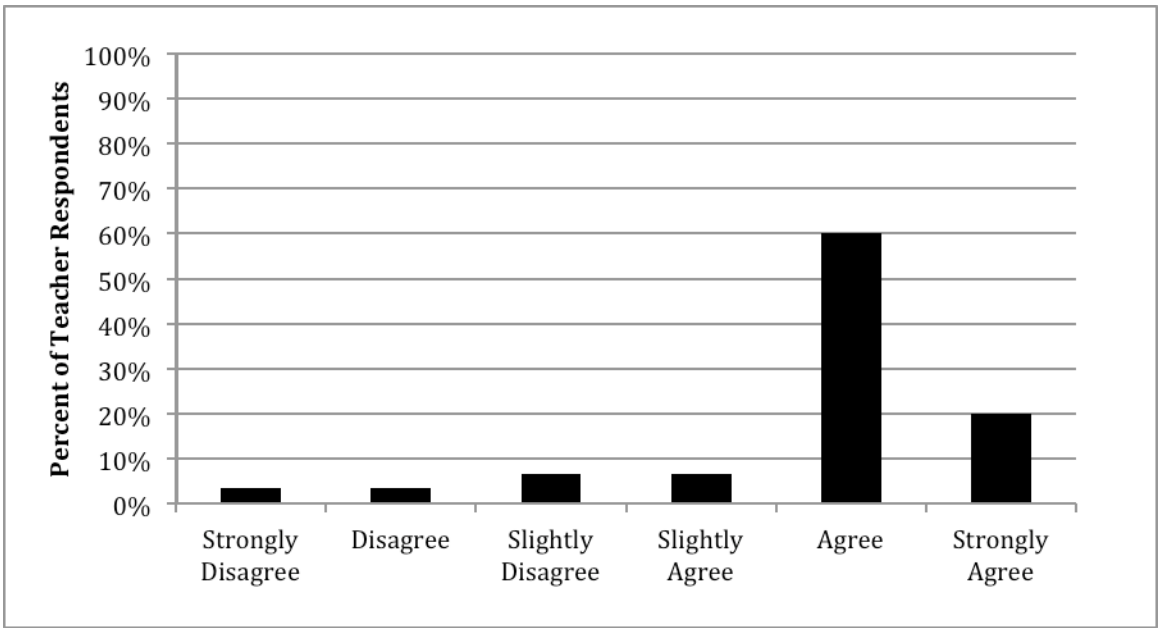

Figure 19. Teacher Ratings of Item 8: The Measures Were A Good Way to Assess Students' Reading Strengths and Weaknesses (n = 30)

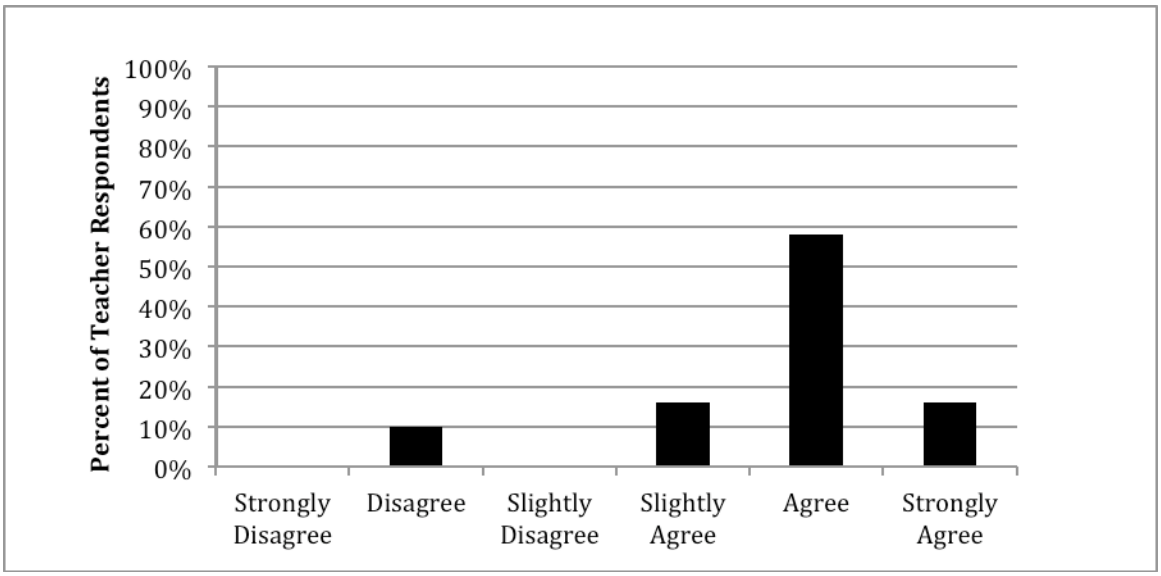

Figure 20. Teacher Ratings of Item 9: Overall, the Measures Would Be Beneficial for Planning Reading Instruction (n = 31)

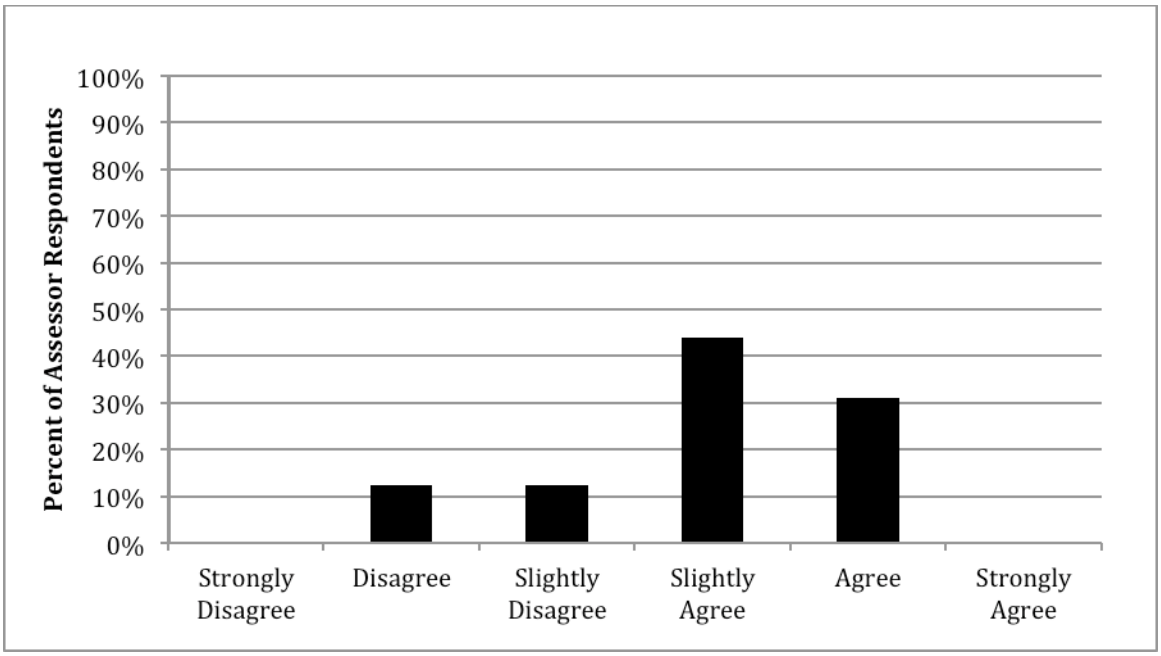

Figure 21. Assessor Ratings of Item 1: The Administration and Scoring Rules Were Easy to Follow (n = 16)

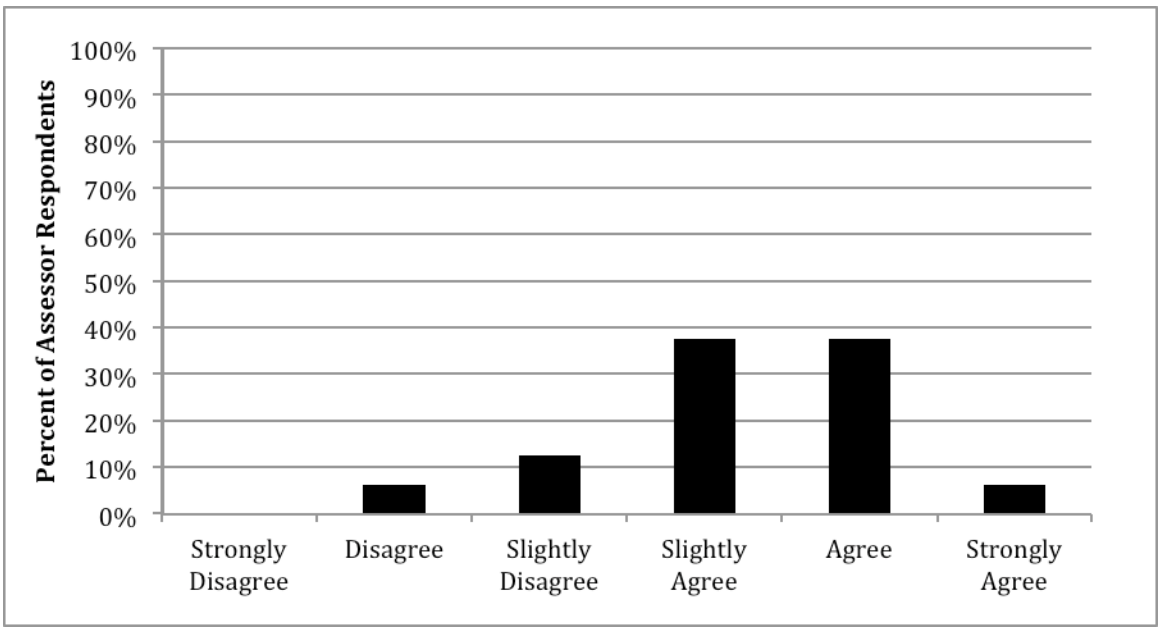

Figure 22. Assessor Ratings of Item 2: The Materials Were Organized Appropriately for Efficient Administration of the Measure(s) (n = 16)

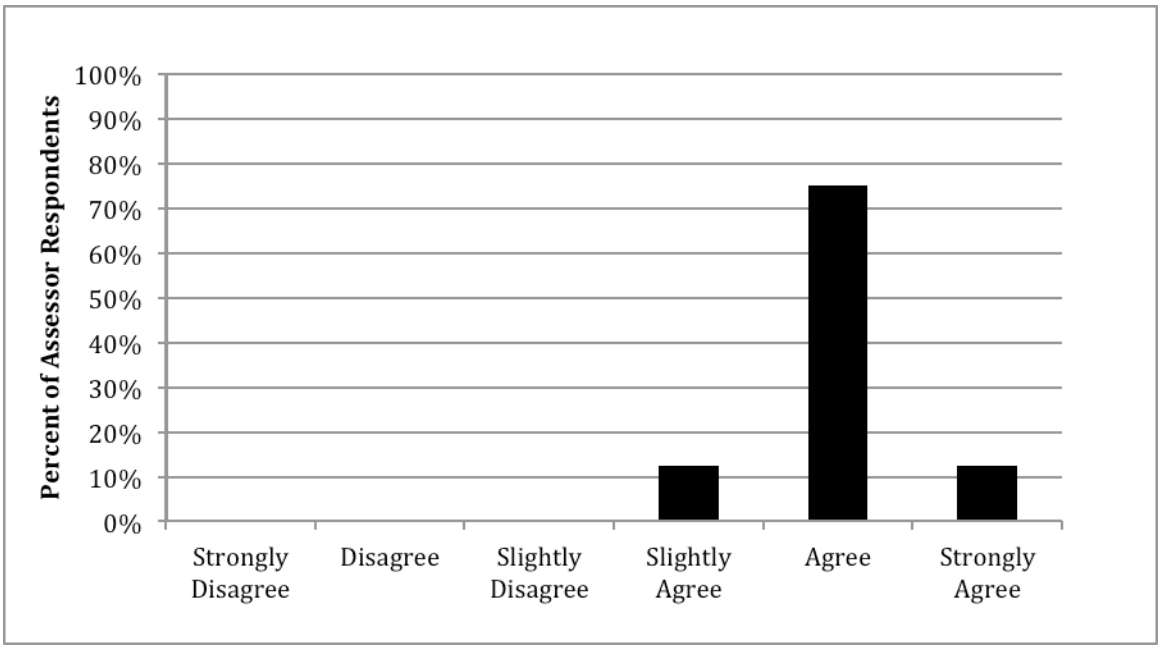

Figure 23. Assessor Ratings of Item 3: I Believe that the Number, Type, and Sequence of Practice Items Were Sufficient to Ensure that the Students Understood the Task (n = 16)

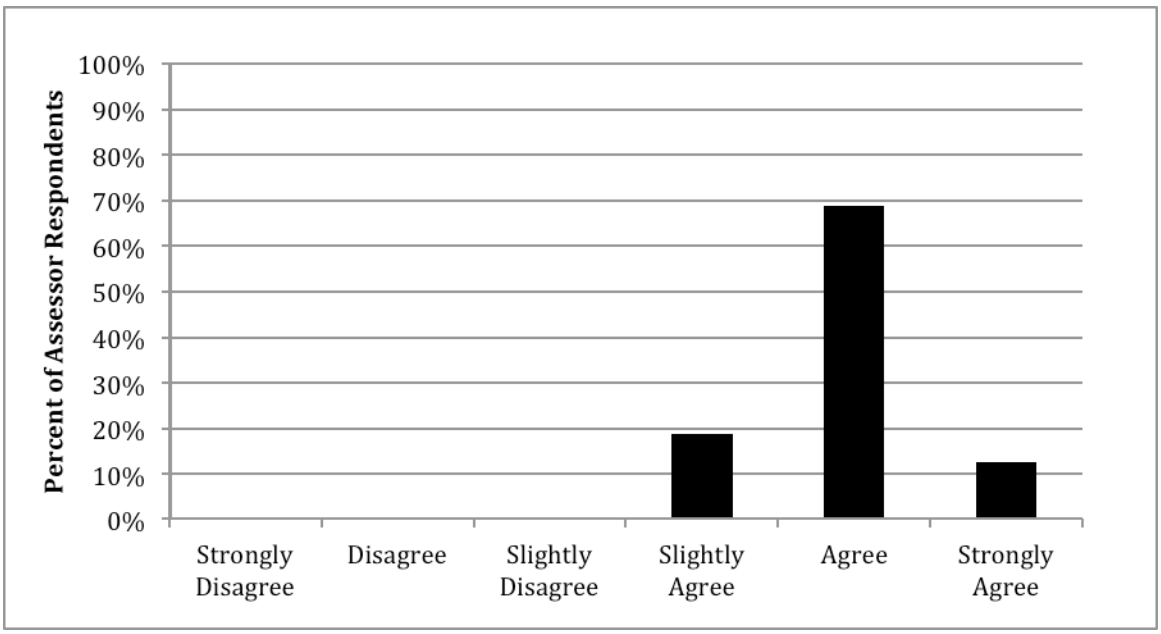

Figure 24. Assessor Ratings of Item 4: I Believe that the Tasks Were Appropriate to the Age/Grade Level of the Students I Tested (n = 16)

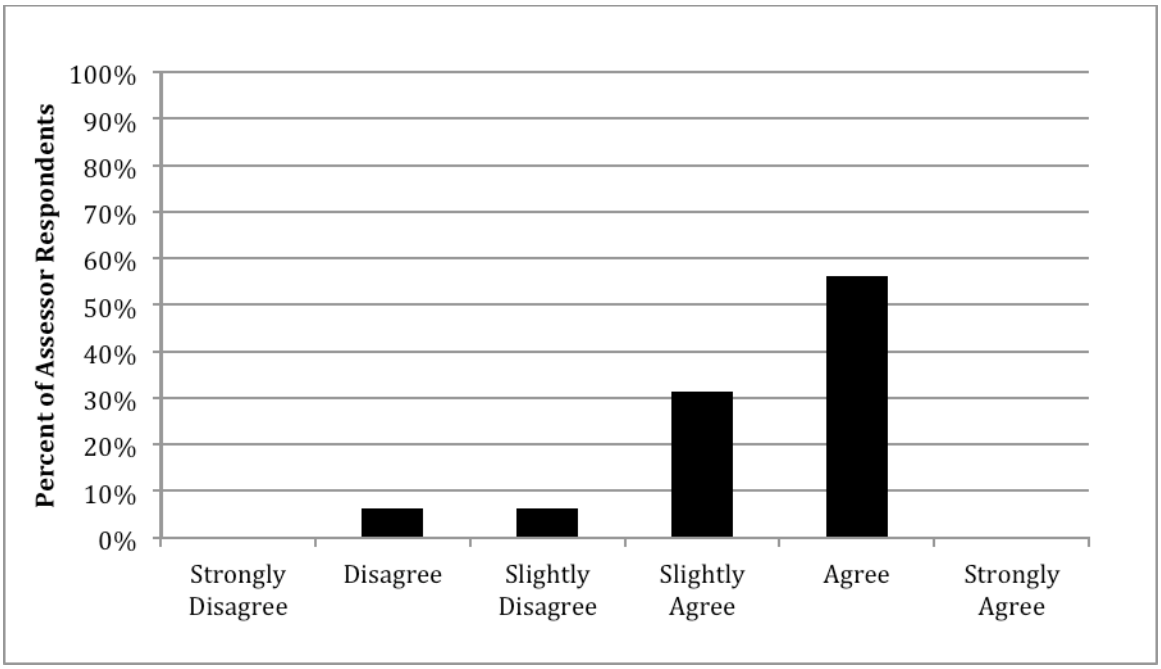

Figure 25. Assessor Ratings of Item 5: All Items Included Within the Measure Were Appropriate (n = 16)

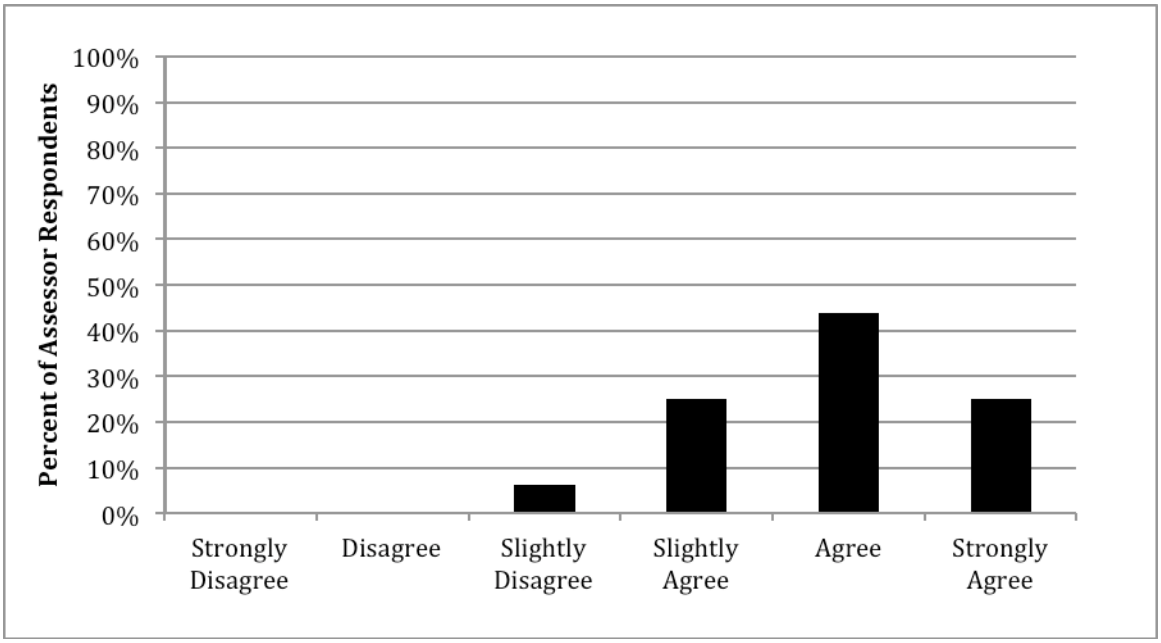

Figure 26. Assessor Ratings of Item 6: I Believe that the Scores Obtained From the Measures Accurately Reflect Students' Skill Level (n = 16)

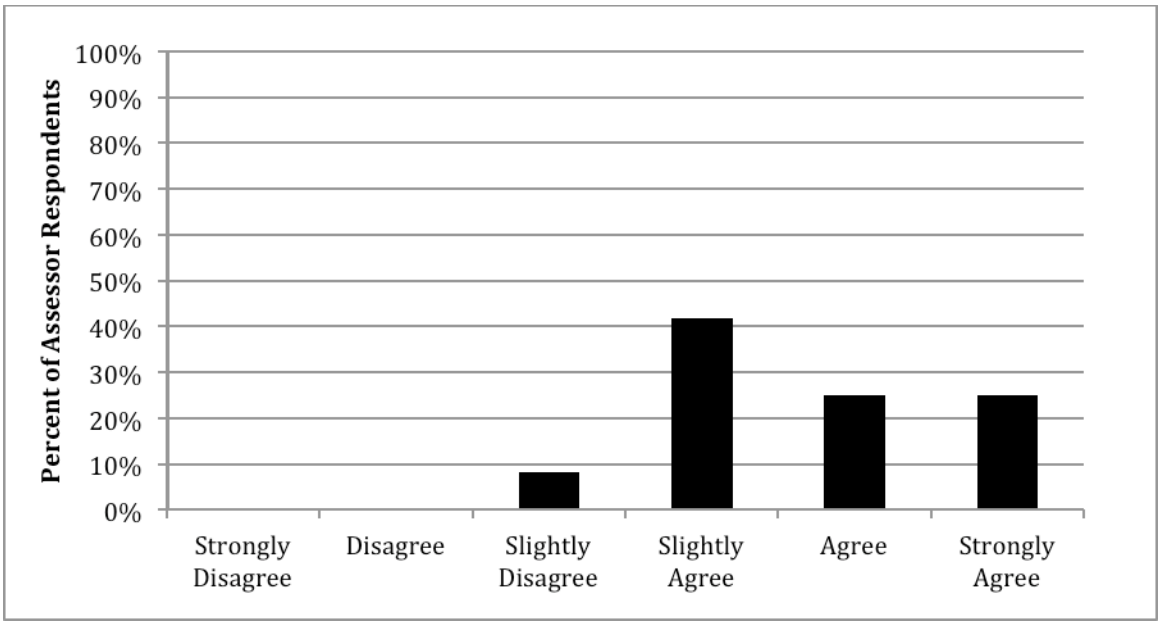

Figure 27. Assessor Ratings of Item 7: I Would Suggest the Use of the Measures to Others (n = 12)

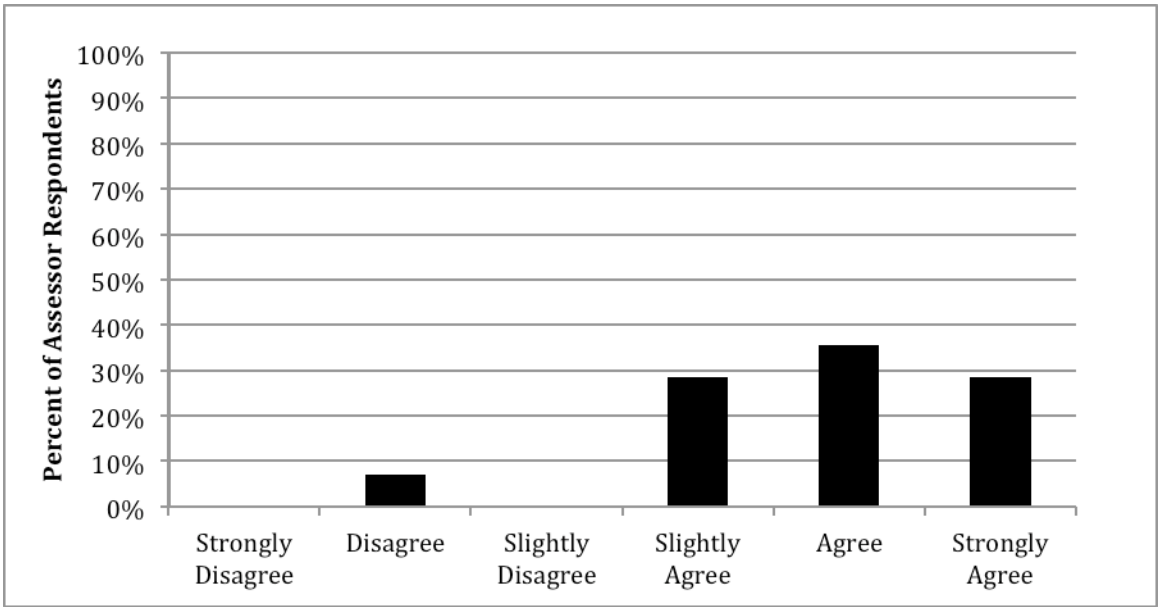

Figure 28. Assessor Ratings of Item 8: The Measures Were A Good Way to Assess Students' Reading Strengths & Weaknesses (n = 14)

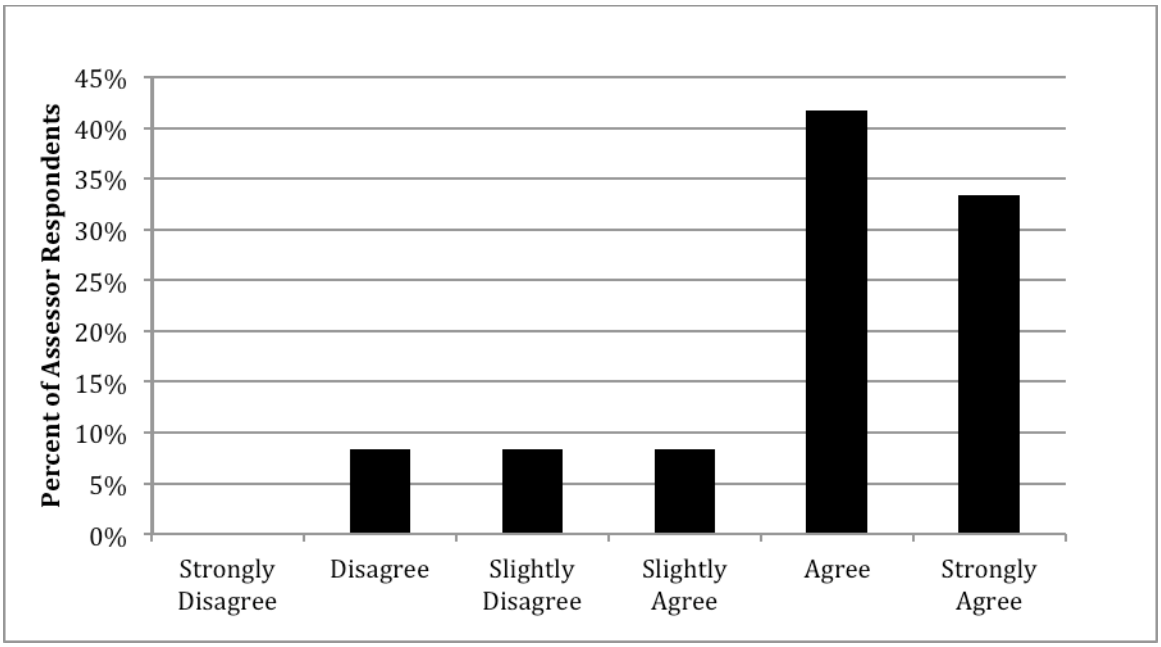

Figure 29. Assessor Ratings of Item 9: Overall, the Measures Would Be Beneficial for Planning Reading Instruction (n = 12)

Appendix A

List of DIBELS Deep Decision Rules (Phase 1 Research Study)

- On any form, if a student does very well, if the next form up in the sequence is marked with an "X" on the attached chart, it may be given. For example, if a first grade student completes the WRD 2 Form successfully during the fall data collection, they may be administered WRD 3.

- *If an WRD form is discontinued before getting to high-frequency words and sentences, proceed to administer high-frequency words and sentences anyway. (Note: this may have been implemented inconsistently in the fall)*

- On any form, if a student meets the discontinue criterion (e.g., less than 3 correct in each of 3 **consecutive** sections), proceed to next form down in the sequence. For example, if testing on WRD 5 and the student gets less than three correct in each section B, C, and D, then test high-frequency words and sentences, and then proceed to administer WRD 4. If testing WRD 4 and the student meets this discontinue criterion, then begin testing WRD 3, and so on.

- **Exceptions** to the discontinue criteria (as noted in the test books), include the following:
    1. On PA Form 1, if the student does not respond correctly to at least three items in each of the first three sections (i.e., Sections A1-B1), then administer Section D of PA Form 1 instead of discontinuing.
    2. On PA Form 1, only administer section C1a if the student does not respond correctly to at least three items on C1.
    3. On PA Form 2, if the student gets fewer than 3 correct on section B1, then drop back to PA 1.
    4. On PA Form 2, if the student gets fewer than 3 items correct on section B2, then drop back to PA 1.
    5. On WRD Form 1 section A, if a student does not appear to understand this task or earns a score of less than five letter sounds correct in the first row after prompting, then discontinue testing WRD Form 1 and begin testing with DIBELS Deep PA Form 2.

- Section Discontinue Rule (applies on every form except QS)—discontinue testing a section if the student answers three consecutive items incorrectly.

- Quick Screen Discontinue Rule—discontinue testing if the child has 5 consecutive incorrect items.