Monitoring Early Reading Development in First Grade: Word Identification Fluency Versus Nonsense Word Fluency

LYNN S. FUCHS
DOUGLAS FUCHS
DONALD L. COMPTON
Peabody College of Vanderbilt University

ABSTRACT: This study contrasted the validity of 2 early reading curriculum-based measurement (CBM) measures: word identification fluency and nonsense word fluency. At-risk children (n = 151) were assessed (a) on criterion reading measures in the fall and spring of first grade and (b) on the 2 CBM measures each week for 7 weeks and twice weekly for an additional 13 weeks. Concurrent and predictive validity for CBM performance level and predictive validity for CBM slopes demonstrated the superiority of word identification fluency over nonsense word fluency. Findings are discussed in terms of the measures’ utility for identifying children in need of intensive instruction and for monitoring children’s progress through first grade.
BACKGROUND INFORMATION ON CBM

CBM differs from most forms of classroom assessment in several ways (Fuchs & Deno, 1991) including these two features: First, CBM is standardized so that the behaviors to be measured and the procedures for measuring those behaviors are prescribed, with documented reliability and validity. Second, CBM’s focus is long-term so that the testing methods and the testing content remain constant, with equivalent weekly tests spanning much, if not all, of the school year. The reason for long-term consistency is so that progress can be monitored systematically over time.

In using the CBM passage reading fluency task, the teacher relies on established methods to identify passages of equivalent difficulty; each equivalent passage represents the material students should be comfortable reading at year’s end. The teacher administers CBM by having the student read aloud a different passage for each assessment, each time for 1 min, and the weekly score is the number of words read correctly. Each assessment produces an indicator of reading competence because it requires a multifaceted performance. This performance entails, for example, a reader’s skill at automatically translating letters into coherent sound representations, unitizing those sound components into recognizable wholes and automatically accessing lexical representations, processing meaningful connections within and between sentences, relating text meaning to prior information, and making inferences to supply missing information. As competent readers translate text into spoken language, they coordinate these skills in an obligatory, seemingly effortless manner (Fuchs, Fuchs, Hosp, & Jenkins, 2001).

Because the CBM passage reading fluency task reflects this complex performance, it can be used to characterize reading expertise and to track its development in the primary grades (e.g., Biemiller, 1977-1978; Fuchs & Deno, 1991). To characterize reading expertise, CBM is interpreted in a norm-referenced or criterion-referenced framework. Within a normative framework, practitioners designate reading difficulty by comparing CBM performance levels between individuals. For example, with local CBM norms, students performing below the third percentile are identified as reading disabled and entitled to intensive reading instruction. Within a criterion-referenced perspective, CBM benchmarks specify the minimum performance levels associated with future reading success. For example, schools might establish that students who fail to attain a CBM score of 75 by the end of second grade have a poor probability of achieving a “proficient” score on their state’s high-stakes fourth-grade reading assessment; therefore, students scoring below this benchmark at the end of second grade are candidates for intensive reading instruction.

Across these normative- and criterion-referenced perspectives, CBM is administered in a single timeframe to identify students who require special attention. By contrast, the major purpose for CBM is to monitor the development of academic competence. Progress monitoring requires an intra-individual framework, where CBM is collected frequently (weekly or monthly); student scores are graphed; a slope is derived from the graphed scores to quantify reading improvement; and the teacher applies decision rules to the slope to formulate instructional decisions.

These strategies for characterizing reading competence in one timeframe and for describing growth over time using CBM have been shown to be more sensitive to inter- and intra-individual differences than those offered by other commercial tests and by other classroom reading assessments (e.g., Marston, Fuchs, & Deno, 1986). In addition, CBM is sensitive to growth made under a variety of treatments (Fuchs, Fuchs, & Hamlett, 1989b; Hintze & Shapiro, 1997; Hintze, Shapiro, & Lutz, 1994; Marston et al., 1986). In a related way, special educators’ instructional plans, developed in response to CBM, incorporate a wide range of reading methods including, for example, decoding instruction, repeated readings, vocabulary instruction, story grammar exercises, and semantic mapping activities (Fuchs, Fuchs, Hamlett, & Ferguson, 1992). So, CBM is not tied to any particular reading instructional method.
Perhaps most important, however, studies indicate that CBM progress monitoring enhances special educators’ capacity to plan programs for and effect achievement among students with serious reading problems. The methods by which CBM informs reading instruction rely on the graphed performance indicator. Decisions are based on a graph displaying time on the horizontal axis and performance (number of words read correctly from text in 1 min) on the vertical axis. If a student’s growth trajectory is judged to be adequate, the teacher increases the student’s goal for year-end performance; if not, the teacher revises the instructional program. Research shows that these decision rules produce more varied instructional programs, which are more responsive to individual needs (Fuchs et al., 1989b), with more ambitious student goals (Fuchs, Fuchs, & Hamlett, 1989a) and stronger end-of-year scores on commercial reading tests (e.g., Fuchs, Deno, & Mirkin, 1984; Wesson, 1991).

In reading, the vast majority of CBM research centers on this passage reading fluency task. When considering the developmental span within the elementary school years, however, the CBM passage reading fluency task is incomplete. The focus of the present study is the early phase of first grade, when the CBM passage reading fluency task produces a floor effect for many students. These students may make progress toward becoming a reader even though performance on the CBM passage reading fluency task, which hovers around zero, indicates a lack of growth. For this reason, alternative tasks, which are sensitive to beginning first-grade reading development, are required.

**Research on CBM Measures for Beginning First-Grade Reading Development**

Prior work provides the basis for supposing that two measures may be potentially useful for indexing and monitoring beginning first-grade reading development: word identification fluency and nonsense word fluency. **Word identification fluency** is one of the CBM measures investigated by Deno and colleagues (Deno, Mirkin, & Chiang, 1982) at the University of Minnesota Institute for Research on Learning Disabilities. With word identification fluency, students have 1 min to read isolated words, presented in lists; the words are selected randomly from high frequency word lists. The score, the number of words read correctly, represents automatic word recognition skill, a hallmark of competent reading behavior. Deno et al. examined the concurrent validity of word identification fluency with 66 children in Grades 1 to 6. The criterion measures were the reading comprehension subtest of the Peabody Individual Achievement Test and the phonetic analysis and the inferential and literal reading comprehension subtests of the Stanford Diagnostic Reading Test. When students read third-grade word lists, correlations with these four respective measures were .76, .68, .71, and .75; when they read sixth-grade word lists, respective correlations were .78, .71, .68, and .74.

**Nonsense word fluency** is the first-grade Dynamic Indicators of Basic Literacy Skills, or DIBELS, measure developed more recently by Roland Good, Ruth Kaminski, and colleagues at the University of Oregon (Good, Simmons, & Kame’enui, 2001). With nonsense word fluency, students have 1 min to read consonant-vowel-consonant pseudowords. The score is the number of sounds produced correctly, with credit earned either by saying individual sounds in the pseudowords or by phonologically recoding the pseudowords (with three sounds awarded for each correctly read pseudoword). The score indexes “letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds” (pp. 272-273). As Good et al. showed with samples of 70 to 242 children, concurrent validity with the Woodcock-Johnson readiness cluster score (i.e., visual auditory learning and letter identification) ranged between .35 in May to .59 in February (median coefficient = .52). The predictive validity coefficients from October of first grade to May of first grade were .71 with respect to CBM passage reading fluency and .52 with respect to the Woodcock-Johnson reading cluster score.
PURPOSE AND IMPORTANCE OF
THE PRESENT STUDY

Traditional concurrent and predictive validity, as demonstrated by the Deno et al. (1982) and Good et al. (2001) research groups, are important. By documenting correspondence between the CBM task and important criterion variables, criterion validity verifies that a test measures relevant behaviors and helps establish that the measure can be used to identify students with serious reading problems. Yet, Deno et al.’s findings were restricted to concurrent validity and did not focus on first graders reading first-grade word lists, and the Good et al. concurrent validity coefficients’ criterion measures did not involve reading. Moreover, although concurrent and predictive validity are necessary features of an adequate progress-monitoring task, they do not constitute sufficient evidence that a measure works well for the purpose of monitoring progress over time. Some measures may function well as correlates or predictors of important criterion outcomes (when those measures are collected at one point in time), but fail to represent reading improvement when administered frequently for progress monitoring.

At the same time, technical features of these early reading measures (word identification fluency and nonsense word fluency) have never been contrasted with the same sample using the same procedures. Therefore, direct comparisons of these two alternative CBM measures are not possible. In contrast, the CBM literature, by convention, offers direct comparisons of the tenability of alternative progress-monitoring tasks (e.g., Deno et al., 1982; Fuchs & Fuchs, 1992) so that progress-monitoring measures may be compared and optimal measures identified. This is important because no absolute criterion exists for judging adequate criterion validity. Instead, “our basis for evaluating any one predictor is in relation to other possible predictors” (Thorndike & Hagen, 1969, p. 169). So, direct comparison between word identification fluency and nonsense word fluency is required.

We designed the present study to fill these gaps in the CBM literature on early progress monitoring in the area of reading. Toward that end, we contrasted the concurrent and predictive validity for the two alternative CBM early reading measures: word identification fluency and nonsense word fluency. For predictive validity, we investigated both CBM performance level and CBM slope of improvement. Whereas the predictive validity of performance level (in one timeframe) is relevant for judging the validity of a measure for screening, the predictive validity of slope (improvement over time) is necessary for evaluating the validity of a measure for progress monitoring.

We assessed a relatively large cohort (n = 151) of at-risk children (the target group for progress monitoring) in the fall of first grade on the two early reading CBM measures and then again in the spring; at each time point, we administered criterion measures that directly involved reading (the Woodcock Reading Mastery Tests Word Identification and Word Attack subtests at both time points; the Comprehensive Reading Assessment Battery, which assesses passage reading fluency and comprehension, added in the spring). In addition, we assessed each of the 151 children for 20 weeks, once weekly for the first 7 weeks and twice weekly for the final 13 weeks, on the two alternative early reading CBM measures.

Findings are potentially important in helping school personnel identify useful measures for monitoring the progress of children as they enter and progress through first grade. As schools struggle with No Child Left Behind’s requirement to monitor the reading progress of children in kindergarten through third grade, this topic takes on a sense of urgency. As special educators strive to infuse greater accountability to the individualized education program (IEP) progress-monitoring mandate, findings are equally useful for special educators. Across general and special education, CBM users need to have confidence that the measures they use to identify students for expensive, intensive services, in fact, yield the students most at risk for poor reading outcomes.

Across general and special education, CBM users need to have confidence that the measures they use to identify students for expensive, intensive services, in fact, yield the students most at risk for poor reading outcomes.
students most at risk for poor reading outcomes. Practitioners also need to be assured that when students’ CBM scores are increasing over time, those CBM slopes are associated with improved scores on high-stakes tests and other important outcomes that bespeak adequate end-of-first grade reading.

**METHOD**

**PARTICIPANTS**

The sample for the present set of analyses was derived from an intervention study examining the effects of Peer-Assisted Learning Strategies (PALS) in first grade (McMaster, Fuchs, Fuchs, & Compton, in press). For that study, eight schools were identified in a large, Southeastern, metropolitan public school system; in each school, at least three first-grade teachers volunteered to participate. Half the schools were high poverty (i.e., Title I), and half were designated as non-Title I. In these schools, we identified 33 teachers for participation and, blocking within school, randomly assigned teachers to control (one third of the teachers) or PALS (two thirds of the teachers). Children in the control classrooms were not included in the present database because we did not monitor their performance with either CBM measure; therefore, control students had no data to contribute to the analyses described in this article.

In PALS classrooms, we screened all children on whom we had parental consent using rapid letter naming (number of letters named in 1 min in response to 26 randomly displayed, lowercase letters). Using these scores, we designated the lowest 7 children per class (i.e., Title I), and half were designated as non-Title I. In these schools, we identified 33 teachers for participation and, blocking within school, randomly assigned teachers to control (one third of the teachers) or PALS (two thirds of the teachers). Children in the control classrooms were not included in the present database because we did not monitor their performance with either CBM measure; therefore, control students had no data to contribute to the analyses described in this article.

On average, the age of these children in October of first grade was 6.72 years ($SD = 0.46$). Of these 151 children, 74 were in Title I schools and 77 were in non-Title I schools; 89 received subsidized lunch; 79 were male; 26 were English language learners; 11 had IEPs (6 had speech/language disability, and 5 had a learning disability); 17 had been retained (7 in kindergarten and 10 in first grade); and 58 were African American, 53 were European American, 36 were Hispanic, and 4 were Asian. During the fourth 6-week marking period, 63 children had perfect attendance, 27 had 1 absence, 28 had 2 absences, 11 had 3 absences, 8 had 4 absences, 6 had 5 absences, 3 had 6 absences, 1 each had 7, 8, and 11 absences, and 2 had 13 absences.

**CRITERION MEASURES**

**Word Attack Subtest of the Woodcock Reading Mastery Test-Revised, Form G** (Woodcock, 1987). This measure evaluates students’ ability to pronounce pseudowords. It contains 45 nonsense words, ordered from most easy to most difficult. The test is discontinued after six consecutive errors. Students earn 1 point for each correctly pronounced pseudoword. Scores range from 0 to 45. Split-half and test–retest reliabilities are .95 and .90, respectively, for first grade. Concurrent validity with respect to the Woodcock Johnson at first grade was .57 with Letter-Word Identification, .64 with Word Attack, .43 with Passage Comprehension, and .69 with Total Reading.

**Word Identification Subtest of the Woodcock Reading Mastery Test-Revised, Form G** (Woodcock, 1987). The Word Identification subtest requires children to read single words. It consists of 100 words ordered in difficulty. Testing is discontinued after six consecutive errors. Students earn 1 point for each correctly pronounced word. Scores range from 0 to 100. Split-half and test–retest reliabilities are .99 and .94, respectively, for first grade. Concurrent validity with respect to the Woodcock Johnson at first grade was .57 with Letter-Word Identification, .48 with Word Attack, .75 with Passage Comprehension, and .82 with Total Reading.

**Comprehensive Reading Assessment Battery (CRAB)**. The CRAB (Fuchs, Fuchs, & Hamlett, 1989c) employs 400-word traditional folktales used in previous studies of reading comprehension (e.g., Brown & Smiley, 1977; Jenkins, Heliotis, Haynes, & Beck, 1986). The folktales had been rewritten by Jenkins et al. to approximate a
second- to-third-grade readability level (Fry, 1968), while preserving the gist of the stories. We rewrote them at a readability grade level of 1.5. Students first read aloud one folktale for 3 min and then answer 10 comprehension questions. This is repeated on a second folktale. The questions, developed by Jenkins et al., require short answers, reflecting recall of information contained in idea units of high thematic importance. To generate the fluency score, examiners mark insertions, omissions, mispronunciations, hesitations of more than 4 s, and substitutions as errors (self-corrections are not considered errors). The score is the average number of correct words-read-across the two passages. In this study, we divided the score by 3 min to produce a words-read-correctly per-minute metric. Test–retest reliability ranges from .93 to .96; concurrent validity with the reading comprehension subtest of the Stanford Achievement Test (SAT) was .91 (Fuchs, Fuchs, & Maxwell, 1988). To generate the comprehension score, the tester records student answers to the comprehension questions. When the student makes five consecutively incorrect responses, questioning is terminated. The score is the average number of questions answered correctly across the two passages. For the number of questions answered correctly score, test–retest reliability was .92; the correlation with the SAT was .82 (Fuchs et al., 1988). On the present sample, correlations for the CRAB fluency score were .81 with Woodcock Word Identification and .58 with Woodcock Word Attack; for the CRAB comprehension score, .71 with Woodcock Word Identification and .59 with Woodcock Word Attack.

**Progress-Monitoring Measures**

**Word Identification Fluency.** With word identification fluency, the child is presented with a single page of 50 high-frequency words. Alternate forms were generated by randomly sampling words, with replacement, from 100 high frequency words from the Dolch preprimer, primer, and first-grade-level lists. The student has 1 min to read words. If a student hesitated on an item for 4 seconds, the examiner prompted him/her to proceed to the next word. The alternate test-form/stability coefficient from 2 consecutive weeks was .88.

**Nonsense Word Fluency.** With nonsense word fluency, the child is presented with a single page of 50 consonant-vowel-consonant or vowel-consonant pseudowords. Alternate forms were printed from the DIBELS Web site (http://dibels.uoregon.edu). The student has 1 min to say the sounds constituting the pseudowords or to read the pseudowords. If a student lingered on an item for 4 s, the examiner prompted him/her to proceed to the next item. The score is the number of correctly spoken sounds (with three sounds awarded for a correctly read pseudoword). The median alternate test-form/stability coefficient at 2-month intervals was .83 (Good et al., 2001). As calculated on the current sample, alternate test-form/stability from 2 consecutive weeks was .87.

**Data Collection**

Data collectors were full-time master’s students, full-time doctoral students, or full-time employees with master’s degrees, all of whom had been trained to 100% accuracy in data-collection procedures prior to any data collection. They administered tests in quiet locations in the schools, working with one child at a time. Interscorer agreement, calculated on 20% of protocols by two independent scorers, ranged between 98% and 100%. The fall battery of criterion measures (Woodcock Word Identification and Word Attack) was administered in October. The spring battery of criterion measures (Woodcock Word Identification and Word Attack plus Comprehensive Reading Assessment Battery) was administered in May. Progress-monitoring measures were administered for 20 weeks, once weekly for the first 7 weeks and twice weekly for the final 13 weeks, with word identification fluency and nonsense word fluency administered in the same session and with make-ups completed within 1 week of the targeted data-collection date.

**Data Analysis and Results**

To obtain a level of performance on each progress-monitoring measure, two measurements were averaged for the fall (first two measurements) and for the spring (last two measurements). To obtain
a slope of improvement on each progress-monitoring measure, an ordinary least-squares regression was calculated between calendar days and scores. That slope was converted to a weekly slope (by multiplying the derived slope by 7 days) to represent the weekly increase in score. Slopes were derived for fall (October through December), spring (January through May), and the year (October through May).

Means and standard deviations for these scores and for the criterion variables are shown in Table 1. Mean raw scores on the Woodcock measures correspond to near-average standard scores. This is surprising given that this at-risk sample represented the lowest third of the classrooms in an urban setting. An inflated normative profile at first grade on the Woodcock Reading Mastery Tests has, however, been documented elsewhere (e.g., Fuchs, et al., 2004; Vellutino, Scanlon, Small, & Fanuele, 2003), and evidence that this sample was, in fact, at risk is provided in other ways. That is, relative to a normative profile of 179 children representing the full range of first-grade performance (e.g., Fuchs et al., 2004), the fall CBM word identification fluency score in the present study is .85 standard deviations below the mean and the spring CBM word identification fluency score in the present study is 1.24 standard deviations below the mean. Moreover, the end-of-year CRAB fluency score is considerably below a benchmark performance of 40, even though these children participated in the research-validated PALS reading treatment.

We next ran a series of correlations between word identification fluency and the criterion variables and between nonsense word fluency and the criterion variables. Each pair of correlations (word identification fluency vs. nonsense word fluency) was compared using Walker and Lev’s (1953) formula, which tested the difference between correlations calculated on dependent samples.

To index concurrent validity, we ran correlations between the fall progress-monitoring level and the fall criterion measures (Woodcock Word Identification and Word Attack) and between the spring progress-monitoring level and the spring criterion measures (Woodcock Word Identification and Word Attack; CRAB Fluency and Comprehension). These correlations are reported in the top six rows of Table 2. As these t values in Table

---

### Table 1

**Means and Standard Deviations (n = 151)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fall</th>
<th>Spring</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X (SD)</td>
<td>X (SD)</td>
<td>X (SD)</td>
</tr>
<tr>
<td><strong>Criterion Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRMT-R WID</td>
<td>9.01 (7.98)</td>
<td>30.92 (11.43)</td>
<td>NA</td>
</tr>
<tr>
<td>WRMT-R WAT</td>
<td>3.27 (4.37)</td>
<td>11.64 (6.87)</td>
<td>NA</td>
</tr>
<tr>
<td>CRAB Fluency</td>
<td>NA</td>
<td>29.65 (22.17)</td>
<td>NA</td>
</tr>
<tr>
<td>CRAB Comp</td>
<td>NA</td>
<td>1.33 (1.58)</td>
<td>NA</td>
</tr>
<tr>
<td><strong>CBM Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWF</td>
<td>31.29 (14.47)</td>
<td>52.17 (24.07)</td>
<td>NA</td>
</tr>
<tr>
<td>WIF</td>
<td>10.11 (9.26)</td>
<td>29.72 (19.96)</td>
<td>NA</td>
</tr>
<tr>
<td><strong>CBM Slope</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWF</td>
<td>1.92 (2.04)</td>
<td>1.49 (1.23)</td>
<td>1.24 (0.82)</td>
</tr>
<tr>
<td>WIF</td>
<td>0.90 (0.90)</td>
<td>1.31 (0.95)</td>
<td>1.02 (0.68)</td>
</tr>
</tbody>
</table>

Note: WRMT-R is Woodcock Reading Mastery Test-Revised; WID is the Word Identification subtest; WAT is the Word Attack subtest; CRAB is the Comprehensive Reading Assessment Battery; Fluency is number of words read correctly aloud per minute; Comp is comprehension questions answered correctly; WIF is word-identification fluency; NWF is nonsense word fluency.
## Table 2
Concurrent and Predictive Validity for Word Identification Fluency Versus Nonsense Word Fluency in First Grade (n = 151)

<table>
<thead>
<tr>
<th>Validity</th>
<th>CBM Measure</th>
<th>WIF</th>
<th>NWF</th>
<th>t(148)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concurrent Validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall CBM Level</td>
<td>WRMT-R WID</td>
<td>.77</td>
<td>.58</td>
<td>4.93***</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.59</td>
<td>.50</td>
<td>1.12</td>
</tr>
<tr>
<td>Spring CBM Level</td>
<td>WRMT-R WID</td>
<td>.82</td>
<td>.64</td>
<td>3.82***</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.52</td>
<td>.51</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>CRAB Fluency</td>
<td>.93</td>
<td>.80</td>
<td>2.72**</td>
</tr>
<tr>
<td></td>
<td>CRAB Comprehension</td>
<td>.73</td>
<td>.54</td>
<td>3.23**</td>
</tr>
<tr>
<td><strong>Predictive Validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall CBM Level</td>
<td>WRMT-R WID</td>
<td>.63</td>
<td>.57</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.45</td>
<td>.46</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>CRAB Fluency</td>
<td>.80</td>
<td>.64</td>
<td>4.27***</td>
</tr>
<tr>
<td></td>
<td>CRAB Comprehension</td>
<td>.66</td>
<td>.50</td>
<td>5.79***</td>
</tr>
<tr>
<td>Fall CBM Slope</td>
<td>WRMT-R WID</td>
<td>.43</td>
<td>.05</td>
<td>3.96***</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.27</td>
<td>.03</td>
<td>2.93**</td>
</tr>
<tr>
<td></td>
<td>CRAB Fluency</td>
<td>.54</td>
<td>.16</td>
<td>4.27***</td>
</tr>
<tr>
<td></td>
<td>CRAB Comprehension</td>
<td>.49</td>
<td>.04</td>
<td>5.71***</td>
</tr>
<tr>
<td>Spring CBM Slope</td>
<td>WRMT-R WID</td>
<td>.61</td>
<td>.35</td>
<td>3.52***</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.32</td>
<td>.27</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>CRAB Fluency</td>
<td>.63</td>
<td>.49</td>
<td>2.08*</td>
</tr>
<tr>
<td></td>
<td>CRAB Comprehension</td>
<td>.45</td>
<td>.27</td>
<td>2.13*</td>
</tr>
<tr>
<td>Year CBM Slope</td>
<td>WRMT-R WID</td>
<td>.79</td>
<td>.38</td>
<td>8.18***</td>
</tr>
<tr>
<td></td>
<td>WRMT-R WAT</td>
<td>.50</td>
<td>.28</td>
<td>3.83***</td>
</tr>
<tr>
<td></td>
<td>CRAB Fluency</td>
<td>.85</td>
<td>.58</td>
<td>6.84***</td>
</tr>
<tr>
<td></td>
<td>CRAB Comprehension</td>
<td>.66</td>
<td>.27</td>
<td>6.80***</td>
</tr>
</tbody>
</table>

Note: WIF is word-identification fluency; NWF is nonsense word fluency; WRMT-R is Woodcock Reading Mastery Test-Revised; WID is the Word Identification subtest; WAT is the Word Attack subtest; CRAB is the Comprehensive Reading Assessment Battery.

'a To test the difference between correlations calculated on the same sample, we relied on Walker and Lev's (1953) formula.

*p < .05. **p < .01. ***p < .001.
2 reveal, correlations for the CBM word identification fluency measure were reliably higher than for nonsense word fluency for one of the two fall concurrent validity criterion variables and for three of the four spring concurrent validity criterion variables. Basically, comparisons favored word identification fluency except when the criterion variable was highly aligned with nonsense word fluency (i.e., Woodcock Word Attack). Even there, where we would expect the comparison to favor nonsense word fluency, concurrent validity for the two CBM measures was, in fact, comparable.

To index predictive validity, the following indices were correlated with the spring criterion measures (Woodcock Word Identification and Word Attack; CRAB Fluency and Comprehension): fall progress-monitoring level (see rows 7–10 of Table 2), fall progress-monitoring slope (see rows 11–14 of Table 2), spring progress-monitoring slope (see rows 15–18 of Table 2), and the year’s progress-monitoring slope (see rows 19–22 of Table 2). As with concurrent validity, the vast majority of predictive validity coefficients reliably favored word identification fluency over nonsense word fluency.

We then looked at how the predictive validity of the two CBM measures compared to the predictive validity of the two Woodcock measures. These correlations are shown in Table 3. When spring Woodcock Word Attack was the criterion variable, all four predictors performed comparably. Nonsense word fluency performed comparably to the Woodcock measure when predicting Woodcock Word Identification and when predicting CRAB comprehension, but outperformed both Woodcock measures when predicting CRAB fluency. By contrast, word identification fluency reliably outperformed both Woodcock measures in predicting Woodcock Word Identification, in predicting CRAB fluency, and in predicting CRAB comprehension.

Finally, we performed dominance analysis (Budescu, 1993; Schatschneider, Francis, Fletcher, & Foorman, 2004), which is an extension of multiple regression. Dominance analysis involves the pairwise comparison of all predictors (i.e., fall nonsense word fluency level, full-year nonsense word fluency slope, fall word identification fluency level, full-year word identification fluency slope) as they relate to a spring criterion (i.e., Woodcock Word Identification, Woodcock Word Attack, CRAB fluency, CRAB comprehension). With dominance analysis, a variable is considered dominant over another if the predictive ability of

| **Table 3** Predictive Validity for Fall Woodcock Versus Fall CBM Measures in First Grade (n = 151) |
|----------------------------------|------------------|----------------|------------------|
| **Fall Predictor**               | **Spring Criterion** | **t(148)** |
|                                 | WRMT-R            |             |                 |
|                                 | WID               | WAT          | NWF              |
| WRMT-R WID                      | .60               | .44          | .63              | .57               | 0.71 0.53 | **3.30** | **1.98** |
| WRMT-R WAT                      | .46               | .49          | .45              | .46               | 0.20 0.00 | 0.63    | 0.43    |
| CRAB-F                          | .63               | .47          | .80              | .64               | **5.01** | 0.19    | **7.46** | **2.76** |
| CRAB-C                          | .59               | .55          | .66              | .66               | 1.70    | 0.18    | **2.05** | **0.76** |

Note: WRMT-R is Woodcock Reading Mastery Test-Revised; WID is the Word Identification subtest; WAT is the Word Attack subtest; WIF is word-identification fluency; NWF is nonsense word fluency; CRAB is the Comprehensive Reading Assessment Battery; F is the CRAB passage reading fluency score; C is the CRAB comprehension score.

* Using Walker and Lev’s (1953) formula, these t tests compare correlations between the top fall predictor versus the bottom fall predictor in the relevant column with respect to spring criterion in the relevant row (e.g., the first t test compares the correlation between fall WRMT-R WID and spring WRMT-R WID to the correlation between fall word identification fluency and spring WRMT-R WID). For 148 degrees of freedom, the t value associated with p < .05 is 1.976; the t value associated with p < .01 is 2.61; and the t value associated with p < .001 is 3.36. Any t value associated with p < .05 is bolded.
that variable exceeds the other—both alone and in the presence of all other predictors in the model. Dominance analysis uses asymptotic confidence limits to test differences in the unique effects among pairwise comparisons. These pairwise comparisons are not a test of the amount of unique variance each predictor contributes, but instead are a direct comparison of the differing amounts of unique variance attributed to the two predictors as they relate to the spring criterion.

In Table 4, we show the asymptotic confidence intervals for each of the six pairs of predictors. Each row shows the pair of variables compared. Each column shows results of the dominance analysis for one spring outcome variable. Under each spring outcome variable, three statistics for the dominance analysis are displayed: \( R^2_D \) is the difference between the squared multiple correlations; asymptotic \( SE \) is the standard error of the differences; \( p \) indicates whether the lower and upper bounds of the 95% asymptotic confidence interval included zero. NWF is nonsense word fluency; WIF is word-identification fluency; L is fall level; S is full-year slope. Significant differences are bolded.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>WRMT-WID</th>
<th>WRMT-WAT</th>
<th>CRAB-F</th>
<th>CRAB-C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2_D )</td>
<td>Asy. SE</td>
<td>( p )</td>
<td>( R^2_D )</td>
</tr>
<tr>
<td>NWF-L v. WIF-L</td>
<td>-.03</td>
<td>.02</td>
<td>ns</td>
<td>-.12</td>
</tr>
<tr>
<td>NWF-L v. NWF-S</td>
<td>-.00</td>
<td>.04</td>
<td>ns</td>
<td>-.01</td>
</tr>
<tr>
<td>NWF-S v. WIF-S</td>
<td>-.21</td>
<td>.05</td>
<td>&lt; .05</td>
<td>-.13</td>
</tr>
<tr>
<td>WIF-L v. NWF-S</td>
<td>.03</td>
<td>.02</td>
<td>ns</td>
<td>.11</td>
</tr>
<tr>
<td>WIF-L v. WIF-S</td>
<td>-.18</td>
<td>.05</td>
<td>&lt; .05</td>
<td>-.03</td>
</tr>
</tbody>
</table>

Note: WRMT-R is Woodcock Reading Mastery Test-Revised; WID is the Word Identification subtest; WAT is the Word Attack subtest; CRAB-F is the passage reading fluency score of the Comprehensive Reading Assessment Battery; CRAB-C is the comprehension score of the Comprehensive Reading Assessment Battery. \( R^2_D \) is the difference between the squared multiple correlations; Asy. SE is the standard error of the differences; \( p \) indicates whether the lower and upper bounds of the 95% asymptotic confidence interval included zero. NWF is nonsense word fluency; WIF is word-identification fluency; L is fall level; S is full-year slope. Significant differences are bolded.
variances are significantly different from zero and, in those cases, highlighted the predictor variable that accounts for more unique variance.

With these analyses, we were interested in whether word identification fluency dominated nonsense word fluency. We were also interested in examining whether slope provided additional predictive value over performance level indices. Word identification fluency dominated nonsense word fluency in 10 of 16 comparisons (see rows 1, 3, 4, and 6 of Table 4); word identification fluency slope provided additional predictive value over performance level indices in 4 of 8 comparisons (see rows 3 and 5 of Table 4); nonsense word fluency slope provided additional predictive value over performance level indices in 2 of 8 comparisons (see rows 2 and 4 of Table 4); and word identification fluency slope dominated nonsense word fluency slope in 3 of 4 comparisons (see last row of Table 4).

DISCUSSION

To identify optimal progress-monitoring tasks, direct comparison of various measures is necessary because no absolute criterion exists for judging adequate criterion validity. Instead, the basis for evaluating a predictor is in relation to other possible predictors (Thorndike & Hagen, 1969). Toward that end, this study compared two potentially useful CBM measures for monitoring early reading development in first grade: word identification fluency and nonsense word fluency. For these measures, we examined (a) concurrent validity for CBM level at fall and spring; (b) predictive validity for CBM level from fall to spring; and (c) predictive validity for fall CBM slope to spring final status, for spring CBM slope to spring final status, and for full-year CBM slope to spring final status. Almost all comparisons favored word identification fluency.

To explore concurrent validity near the beginning of first grade, we ran correlations between the two CBM measures and Woodcock Word Identification (which requires students to read words in untimed fashion) and Woodcock Word Attack (which requires students to decode pseudowords in untimed fashion). Given the nature of the two CBM measures, one might expect word identification fluency to correlate more strongly with Woodcock Word Identification and expect nonsense word fluency to correlate more strongly with Woodcock Word Attack. This was only partly true. The correlation with Woodcock Word Identification was, in fact, statistically significantly higher for word identification fluency than for nonsense word fluency (.77 vs. .58). Contrary to expectations, however, the correlations for the two CBM measures with Woodcock Word Attack were comparable (.59 for word identification fluency vs. .50 for nonsense word fluency). So, even at the beginning of the year, when one might assume a lower floor (and therefore greater range) for nonsense word fluency (which awards credit for saying sounds without requiring decoding), greater validity was demonstrated for the CBM word identification fluency measure.

At spring, where we included a better variety of criterion measures, tapping text reading fluency and comprehension with the CRAB, findings again supported word identification fluency over nonsense word fluency, this time even more strongly. Across Woodcock Word Identification, CRAB fluency, and CRAB comprehension, correlation coefficients ranged between .73 to .93 for word identification fluency; between .51 and .80 for nonsense word fluency. The one criterion measure for which correlations between the two CBM measures were comparable (.52 vs. .51) was the Woodcock Word Attack, where the task (reading pseudowords) is more similar to nonsense word fluency than to word identification fluency. Thus, at the end of first grade, the word identification fluency task remains a stronger concurrent correlate of important reading behaviors.

In a similar way, for predictive validity, results favored CBM word identification fluency over nonsense word fluency. Fall CBM word identification fluency scores demonstrated superior predictive validity with respect to the CRAB fluency and comprehension criterion measures. Consequently, for identifying children in October of first grade who are at risk for poor end-of-year reading outcomes, the CBM word identification fluency measure outperforms the nonsense word fluency task.
To supplement these analyses, we also asked whether we could predict spring outcomes just as well—simply by using fall Woodcock scores. The answer was no. In predicting spring Woodcock Word Identification, the fall CBM measures did comparably well to fall Woodcock Word Identification. Moreover, in predicting spring Woodcock Word Attack, both CBM measures outperformed the fall Woodcock Word Attack measure, and the significant differences in predictions of spring CRAB performance consistently favored the fall CBM measures over the fall Woodcock measures. So, in all cases, there was no advantage to using fall Woodcock scores to predict spring reading performance.

Of course, the major purpose for CBM is progress monitoring, where the criterion validity of the CBM slope is more important than CBM level. With CBM progress monitoring, slope is a critical index because CBM slope is used to formulate decisions about whether reading progress is adequate. If the CBM slope indicates adequate progress, the instructional program remains intact; if not, the program is revised. So, the relevant technical question is: Does improvement on the CBM measure, as indexed with slope, reflect meaningful reading development, which predicts reading accomplishment at the end of the year?

Across the fall semester, CBM slopes correlated statistically significantly higher for word identification fluency than for nonsense word fluency with all four spring criterion measures. In fact, coefficients for the nonsense word fluency measure slopes were disappointingly low, ranging from -.04 to .16. Because nonsense word fluency is recommended for progress monitoring in the fall of first grade within the DIBELS system (Good et al., 2001), these findings raise serious concern. An increasing pattern of scores through the first semester of first grade on DIBELS nonsense word fluency appears to bear little relationship to students’ end-of-year reading status. By contrast, for word identification fluency, fall slopes of improvement correlated more strongly, with coefficients of .43 with end-of-year Woodcock Word Identification, .54 with CRAB fluency, and .49 with CRAB comprehension. These correlations are modest, but fall in the range of many predictive measures (Jensen, 1981). And the correlations are noteworthy given the difficulty of achieving strong correlations when measures of change, such as slope, are used as predictors. At this time, therefore, fall word identification fluency slope appears to represent an acceptable index for predicting end-of-year reading outcome. Clearly, it represents a better alternative than fall nonsense word fluency slope.

As might be expected, given the more proximate timeframe, spring slopes of improvement correlated with final reading status measures more strongly than did fall slopes. However, here again, coefficients for spring word identification fluency slopes were reliably stronger than for nonsense word fluency slopes. Perhaps most important, however, very large (and reliable) differences in the magnitude of correlations were observed for the full-year slopes: Coefficients for full-year nonsense word fluency slopes ranged between .27 (with CRAB comprehension) to .58 (with CRAB fluency). By contrast, coefficients for full-year word identification fluency slope ranged from .50 (with Woodcock Word Attack) to .85 (with CRAB fluency). In addition, the superiority of word identification fluency slope (over nonsense word fluency slope) was demonstrated even for Woodcock Word Attack, which seems more transparently related to nonsense word fluency than to word identification fluency.

Dominance analysis provided a supplementary and elegant method for exploring the predictive value of the two early reading progress-monitoring measures. In each case, word identification fluency, level, or slope, dominated nonsense word fluency. Word identification fluency level dominated nonsense word fluency level in predicting CRAB fluency and CRAB comprehension; word identification fluency slope dominated nonsense word fluency level in predicting all spring outcome variables except Woodcock Word Attack; word identification fluency level dominated nonsense word fluency slope in predicting CRAB comprehension; word identification fluency slope dominated nonsense word fluency level in predicting all spring outcome variables except Woodcock Word Attack; and word identification fluency slope dominated nonsense word fluency slope in predicting Woodcock Word Identification; and word identification fluency slope dominated nonsense word fluency slope in predicting all spring outcome variables except Woodcock Word Attack. These summative analyses not only corroborate the superiority of word identification fluency, but also suggest that collecting
word identification fluency slope provides additional predictive value beyond simply collecting fall word identification fluency level data. This indicates the benefit of progress monitoring beyond initial screening for identifying students likely to experience reading difficulty.

In sum, results suggest that word identification fluency functions better than nonsense word fluency as a CBM tool for assessing early reading development in first grade. Because predictive validity with respect to end-of-year text-reading fluency and comprehension is stronger for word identification fluency than for nonsense word fluency, word identification fluency provides a stronger basis for formulating screening decisions in October of first grade. Moreover, the superiority of the word identification fluency over nonsense word fluency is most clearly demonstrated for progress monitoring decisions, where 11 of 12 correlations for CBM slope with respect to end-of-year outcomes were stronger for word identification fluency than for nonsense word fluency. Dominance analysis demonstrated how slope provides additional predictive value beyond one-time screening decisions in the fall. Dominance analysis also corroborates that word identification fluency, fall level, and full-year slope dominates nonsense word fluency, when both metrics for both CBM measures are entered simultaneously to predict spring outcomes. For these reasons, practitioners can have confidence that increases in word identification fluency over time reflect improved performance on important end-of-year reading outcomes. As our results suggest, the same is not true for DIBELS nonsense word fluency, and findings are particularly compelling because data were collected on the same group of children using the same methods.

Why is predictive validity for word identification fluency performance level and slope better than for nonsense word fluency? Although findings do not yield a direct answer to this question, we offer two possible explanations, which correspond to two difficulties with the nonsense word fluency task. First, on the nonsense word fluency task, two students with very different performance patterns may receive equal credit. That is, a student who says three separate sounds in response to a consonant-vowel-consonant pseudoword earns three points—the same score as a student who blends those three sounds into the pseudoword. Clearly, the student who blends the sounds has stronger reading capacity than the child who can only represent the separate sounds. Moreover, in our sample, we observed low-performing students who, when monitored with nonsense word fluency, were increasingly capable of saying many sounds very quickly, without achieving the alphabetic insight required for blending.

A second problem is that competent phonological decoding is, especially as the year progresses, better represented by the capacity to decode a variety of phonetic patterns. So, students who perform well on nonsense word fluency’s consonant-vowel-consonant pseudowords may or may not be skilled at reading consonant-vowel-consonant -e words, r-controlled words, dual vowel words, multisyllabic words, etc. The restriction of the nonsense word fluency task to a single, easy phonetic pattern may reduce the correlation between nonsense word fluency and important criterion measures.

Present findings are reminiscent of earlier work demonstrating the superiority of word identification fluency over nonsense word fluency with a different sample of first graders with more severe reading difficulties (Fuchs, 2003). For that sample of 36 at-risk students who received one-to-one tutoring across the second semester of first grade, nonsense word fluency slopes failed to reliably discriminate student performance on key indicators of reading competence at the end of first grade. In that study, a median split was performed on the slopes of these 36 children, creating a group of children with the top 18 slopes and another group with the bottom 18 slopes. The average effect size comparing these two groups of children on end-of-year indicators of reading competence and on fall-to-spring reading growth was .4 standard deviations, and the difference in the performance of these two groups was statisti-
cally significantly different on only one of the eight criterion measures. By contrast, when top versus bottom groups were formed on the basis of word identification fluency slopes, the average effect size comparing the groups exceeded 1 standard deviation, and the performance of students with the top-half slopes versus those with bottom-half slopes were statistically significantly different on all eight year-end indicators of reading competence and fall-to-spring reading growth. Current findings corroborate those earlier findings, showing how improvement across time on word identification fluency functions better than nonsense word fluency for forecasting end-of-first-grade reading status (as well as reading improvement).

It is important to note that this study employed a restricted sample of at-risk pupils and that we might expect correlations generally to be higher if we were to conduct the study with a greater range of performance. Of course, this renders the large correlations for word identification fluency even more impressive, but as with any study, results should be corroborated with additional samples.

Nevertheless, findings are particularly timely given the press to implement the progress-monitoring component of No Child Left Behind and as special educators ratchet up the IEP progress-monitoring requirement of the Individuals with Disabilities Education Act. Moreover, given widespread adoption of DIBELS, practitioners may wish to reconsider nonsense word fluency in favor of the CBM word identification fluency measure. Clearly, as schools select measures for monitoring the reading progress of children with and without disabilities in the early stages of reading development, results provide a strong basis for selecting word identification fluency over nonsense word fluency. Findings also indicate that monitoring student progress frequently with word identification fluency can contribute importantly, beyond the simple collection of fall screening data, to the identification of students likely to experience difficulty in learning to read in the first grade.

REFERENCES


---

**ABOUT THE AUTHORS**

LYNN S. FUCHS (CEC #185), Nicholas Hobbs Professor; DOUGLAS FUCHS (CEC #185), Nicholas Hobbs Professor; and DONALD L. COMPTON (Tennessee Federation), Assistant Professor, Special Education, Peabody College of Vanderbilt University, Nashville, Tennessee.

Inquiries should be addressed to Lynn S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.

The research described in this paper was supported in part by Grant #H324C000022 from the U.S. Department of Education, Office of Special Education Programs, and Grant HD 15052 from the National Institute of Child Health and Human Development to Vanderbilt University. Statements do not reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

Manuscript received October 2003; accepted January 2004.
If you're an Educator, write down this address

www.cec.sped.org/pd/webcourses

CEC Online Courses Offer:

Convenience  All you need is a computer with Internet access to attend these classes from your home, school or office

Flexibility  Each course is self-paced and offered continuously, so you decide when to begin and how quickly to finish

Value  Affordable courses with volume discounts for groups of 15 or more

Support  24/7 customer support and technical assistance

Professional Credit  Continuing Education Units (CEUs) are awarded upon course completion

Courses Now Available!

CEC 101: Adopting Language Arts, Social Studies, and Science Materials for the Inclusive Classroom
CEC 102: Adopting Reading and Math Materials for the Inclusive Classroom

Assessing One and All Series
CEC 111: Educational Assessment Principles and Practices
CEC 112: Large-Scale Assessment
CEC 113: Accommodations and Alternate Assessment

Coming Soon!

Introducing a new series of courses in Fall 2004, with up-to-date, practical information on Classroom Management, Classroom Assessment, Direct Instruction Reading, Inclusive Classroom Strategies, and much more!

Council for Exceptional Children
The voice and vision of special education