Article

HAMMILL INSTITUTE ON DISABILITIES

Journal of Learning Disabilities
2014, Vol. 47(3) 254–270

DOI: 10.1177/0022219412454455
journaloflearningdisabilities.sagepub.com
SAGE

Monitoring Early First-Grade Reading Progress: A Comparison of Two Measures

Nathan H. Clemens, PhD[1], Edward S. Shapiro, PhD[2], Jiun-Yu Wu, PhD[3], Aaron B. Taylor, PhD[1], and Grace L. Caskie, PhD[4]

Abstract

This study compared the validity of progress monitoring slope of nonsense word fluency (NWF) and word identification fluency (WIF) with early first-grade readers. Students (*N* = 80) considered to be at risk for reading difficulty were monitored with NWF and WIF on a 1-2 week basis across 11 weeks. Reading skills at the end of first grade were assessed using measures of passage reading fluency, real and pseudoword reading efficiency, and basic comprehension. Latent growth models indicated that although slope on both measures significantly predicted year-end reading skills, models including WIF accounted for more variance in spring reading skills than NWF, and WIF slope was more strongly associated with reading outcomes than NWF slope. Analyses of student growth plots suggested that WIF slope was more positively associated with later reading skills and discriminated more clearly between students according to successful or unsuccessful year-end reading outcomes. Although both measures may be used to monitor reading growth of at-risk students in early first grade, WIF may provide a clearer index of reading growth. Implications for data-based decision-making are discussed.

Keywords

Curriculum-based measurement, reading/early literacy, at-risk learners

In the late 1970s, Stanley Deno and colleagues at the University of Minnesota developed curriculum-based measurement (CBM), a framework designed to allow special education teachers to repeatedly monitor students' progress toward individual educational goals using brief samples of performance (Deno, 2003; Deno & Mirkin, 1977). Central to CBM was the concept of general outcomes measurement (GOM), in which an equivalent set of measures, independent of the curriculum of instruction, were used to reflect achievement in an overall academic domain. Thus, GOM permits the frequent evaluation of students' academic "vital signs," change in students' achievement over time, and effectiveness of instruction (Deno).

Research has indicated that teachers who use CBM are more realistic about their students' progress, demonstrate greater structure in their teaching, and facilitate greater student awareness of progress (Fuchs, Deno, & Mirkin, 1984; Shinn & Hubbard, 1992). When teachers use CBM frequently with a set of systematic decision rules, students have been shown to display superior academic gains over students monitored less frequently with more traditional forms of assessment (e.g., unit-end tests) or without a set of decision rules, (Fuchs & Fuchs, 1986; Fuchs, Fuchs, & Hamlett, 1989; Stecker, Fuchs, & Fuchs, 2005), primarily because more frequent and direct feedback on students' progress enables more timely changes to instruction.

With CBM, decisions to continue or change instructional programs are often based on student progress (or lack thereof). Rate of improvement (i.e., slope) is a recommended variable on which to base instructional decisions (Jenkins & Terjeson, 2011), with the assumption that an increasing pattern of scores is indicative of skill acquisition, and when slope is consistent with expectations, reflective of effective instruction. Conversely, lower-than-expected rates of growth are a signal to adjust instruction. For a measure to be considered useful for monitoring progress, growth indicated by the measure should be associated with important outcomes. Otherwise, slope may provide an ambiguous or potentially misleading index of student progress, and may not yield information that is accurate for instructional decision-making.

[1]Texas A&M University, College Station, TX, USA
[2]Lehigh University Center for Promoting Research to Practice, Bethlehem, PA, USA
[3]National Chiao Tung University, Hsinchu, Taiwan
[4]Lehigh University, Bethlehem, PA, USA

Corresponding Author:
Nathan H. Clemens, Ph.D., Texas A&M University, 721 Harrington Tower, 4225 TAMU, College Station, TX 77843-4225, USA.
Email: nclemens@tamu.edu

CBM in Reading and Early Literacy

Early CBM research found that the rate at which students read words, either in lists or connected text, corresponded closely to reading comprehension (Deno, Mirkin, & Chaing, 1982). Subsequently, the rate at which students read from a passage of text (CBM-R) became the most widely studied form of CBM, as research established CBM-R as a valid measure of overall reading achievement (e.g., Fuchs, Fuchs, & Maxwell, 1988; Reschley, Busch, Betts, Deno, & Long, 2009; Shinn, Good, Knutson, Tilly, & Collins, 1992). Using CBM-R assumes that students have at least some degree of skill in reading connected text, which in many cases is not often expected until the latter half of first grade. This presents a unique challenge for educators in the early part of first grade: Although it is considered a critical transitional period in a child's reading development as early literacy skills in letter-sound knowledge and phonological awareness are utilized to facilitate skills in reading words (Snow, Burns, & Griffin, 1998), compared to CBM-R, far less research exists on the most appropriate measures for monitoring progress during this time.

For young children, specific skills have been shown to be highly predictive of later reading, including skills in naming letters or letter sounds, identifying or manipulating phonemes, or reading words or pseudowords (see National Early Literacy Panel, 2008, for a review). Logic suggests that monitoring these skills would provide information on growth toward later reading outcomes, thinking that has led to downward extensions of CBM into the areas of early and preliteracy with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), AIMSweb, and EasyCBM as popular sources of measures. Although a good deal of research on these and other early literacy measures exists, studies have primarily focused on their validity and screening accuracy from measurement at a single point in time (for reviews see Goffreda & DiPerna, 2010; Jenkins, Hudson, & Johnson, 2007).

One measure that has been suggested for monitoring progress in early grades levels is nonsense word fluency (NWF). Versions of NWF are available through DIBELS and AIMSweb, and the measure is widely used across the United States. Beginning in the middle of kindergarten, NWF was designed to measure students' acquisition of letter-sound relationships and basic decoding skills. NWF consists of a list of vowel-consonant (VC) and consonant-vowel-consonant (CVC) pseudowords in which students are asked to read as whole units or by naming individual letter sounds. In kindergarten and first grade, NWF has demonstrated moderate to strong predictive validity with reading skills in later first and second grade (Catts, Petscher, Schatschneider, Sittner Bridges, & Mendoza, 2009; Clemens, Shapiro, & Thoemmes, 2011; Cummings, Dewey, Latimer, & Good, 2011; Fien et al., 2010; Fuchs, Fuchs, & Compton, 2004; Goffreda & DiPerna, 2010; Goffreda, DiPerna, & Pedersen, 2009; Good, Baker, & Peyton, 2009; Harn, Stoolmiller, & Chard, 2008; Johnson, Jenkins, Petcher, & Catts, 2009; Kim, Petscher, Schatschneider, & Foorman, 2010; Ritchey, 2008).

Much of the validity evidence for NWF is based on scores from a single point in time; relatively less is known about the degree to which slope of improvement is associated with later reading outcomes. Fuchs et al. (2004) found that when monitoring the progress of at-risk first graders, NWF slope was weakly related to reading outcomes at the end of first grade. Other studies have investigated the validity of NWF slope measured across two or three screening assessment points across first grade. Vanderwood, Linklater, and Healy (2008) noted that NWF growth in first grade was positively and significantly associated with third-grade reading skills for English language learners; however, growth was no longer significant after accounting for year-end NWF scores. Further, Kim et al. (2010) found that first-grade NWF growth across benchmark assessments was negatively associated with reading fluency and comprehension at the end of first, second, and third grade, suggesting that higher NWF slopes were associated with lower reading scores in the future.

Several additional studies have investigated whether the relationship between NWF slope and later reading skills is moderated by students' initial NWF scores. Four studies (Cummings et al., 2011; Fien et al., 2010; Good et al., 2009; Harn et al., 2008) investigated NWF slope with first-grade students, each arriving at similar conclusions: NWF slope was weakly correlated with later reading skills for students whose initial NWF scores were within the upper range of the distribution, and more strongly correlated for students with lower initial NWF scores. According to Harn et al., weak associations between slope and outcome for students with higher initial skills had an attenuating effect on the overall correlation between NWF slope and later reading when considering the full population.

Word Identification Fluency

Fluency reading words in lists has also been considered as a progress monitoring measure for early grade levels. Initial CBM investigations found that word list reading demonstrated criterion-related validity similar to that of fluency reading connected text (e.g., Deno et al., 1982). Automaticity in reading words is a critical reading skill (Gough, 1996), as accurate and effortless word recognition skills facilitate reading comprehension by freeing cognitive resources from costly decoding processes which can then be dedicated to comprehending what was read (LaBerge & Samuels, 1974; Nathan & Stanovich, 1991; Perfetti, 1984). Problems with individual word recognition represent the most common form of reading disability (Fletcher, Lyon, Fuchs, & Barnes, 2007). Speed of word recognition powerfully discriminates between skilled and disabled readers (Compton & Carlisle, 1992; Ehri & Saltmarsh, 1995; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003), and context-free word recognition (i.e., words in isolation or in lists) is considered a more direct assessment of word recognition skills than reading connected text (Torgesen, Wagner, & Rashotte, 1999).

Word identification fluency (WIF) is a measure of word list reading that has been used in first grade. Studies have found moderate to strong predictive relationships between WIF measured in first grade and reading skills in later first and second grade (Clemens et al., 2011; Compton, Fuchs, Fuchs, & Bryant, 2006; Compton et al., 2010; Fuchs et al., 2004; Speece et al., 2011; Zumeta, Compton, & Fuchs, 2011), and research has also extended the validity of word reading fluency measures downward to kindergarten (Lai, Nese, Jamgochian, Alonzo, & Tindal, 2009). Rapid reading of words (i.e., "by sight") is facilitated by a process in which knowledge of letter-sound relationships and basic phonemic awareness are used to link increasingly larger letter strings to pronunciations in memory (Adams, 1990; Ehri, 2002, Share 1995; 2008). Thus, WIF might be a useful measure for monitoring early reading development by providing an index of how well students are consolidating early literacy skills in developing automatic word recognition, an important goal for first grade instruction (Snow et al.,1998).

Limited research has investigated the validity of WIF slope. Among existing studies, Fuchs et al. (2004) compared the validity of WIF and NWF progress monitoring slopes in a sample of first-grade students considered to be at risk for later reading difficulty, finding that WIF slope was moderately correlated with year-end reading skills, and coefficients were significantly higher than those observed for NWF slope. Dominance analyses suggested that WIF slope accounted for more unique variance than NWF slope in year-end reading skills including word identification, reading fluency, and comprehension (but not word attack).

Studies have found that WIF slope in first grade is predictive of later reading skills, and that the addition of WIF slope to first grade reading screening batteries results in improved accuracy in predicting later reading difficulty (Compton et al., 2006, 2010; Speece et al. 2011). Additionally, Zumeta et al. (2011) demonstrated validity of WIF progress monitoring slope for first grade students considered at risk for reading difficulty, as well as typical and high achieving students, in predicting reading skills at the end of first grade.

Summary and Purpose of the Current Study

In summary, both NWF and WIF have demonstrated validity as first-grade predictors of later reading skills, and both have been used for monitoring the reading progress of students in early first grade. However, with the exception of Fuchs et al. (2004), studies have not investigated the validity of NWF and WIF slope on a frequent progress monitoring basis within a common sample of first-grade students considered to be at risk for later reading difficulty. With the focus of NWF on alphabetic understanding, and WIF's emphasis on word recognition, both appear to assess skills important in reading development that are relevant to first grade reading instruction. However, research is needed to better inform educators' decisions regarding measures most valid for monitoring early reading progress. Thus, the purpose of this study was to compare the validity of NWF and WIF progress monitoring slope in predicting reading outcomes at the end of first grade. Because progress monitoring was intended for (and primarily used with) students who are lower achieving, we chose a sample of first-grade students considered to be at risk for later reading failure to investigate the validity of progress monitoring slope. Additionally, the degree to which slope predicted later outcomes was investigated while simultaneously considering initial level, based on recent findings indicating that NWF slope demonstrates greater validity in predicting later outcomes for students with lower initial skills (Cummings et al., 2011; Fien et al., 2010; Good et al., 2009; Harn et al., 2008). Consistent with the results of Fuchs et al. (2004), it was hypothesized that WIF slope would be more strongly associated with later reading skills than NWF slope.

Method

Participants

Participants were first-grade students from three elementary schools (nine classrooms) in Pennsylvania. Data were collected as part of a project investigating a school-wide three-tier response to intervention (RTI) reading model (described below). The schools were located within the same school district in a suburban community.

Analyses were conducted with 80 students who, based on DIBELS universal screening assessments (NWF, Letter Naming Fluency; and Phoneme Segmentation Fluency) conducted at the beginning of the school year with all first grade students (N = 181), were considered to be at risk for reading difficulty and were receiving reading intervention supplemental to core instruction. The average NWF score for the group was 10.34 on the fall administration of NWF (DIBELS cut score for risk status is 24 sounds per minute for fall of first grade), compared to 45.38 for students not considered for the risk sample for progress monitoring. The average WIF score for the group was 4.33, compared to 35.51 for the sample considered not at risk. If considering the strata utilized by Fien et al. (2010), 98% of the students had fall NWF scores that fell within the lowest two strata (0-12 and 12-23) in which Fien et al. found the strongest relations between NWF slope and later outcomes. The sample was 63% male, 60% White, 23% Black, 15% Hispanic/Latino, and 3% Asian/Pacific Islander. Three students spoke

English as a second language and were included in the analyses. No students had been identified with learning disabilities; five students with autism spectrum disorder were excluded from the sample. All three schools were eligible for school-wide Title 1 funding. An average of 32% of students across the three schools was eligible for free or reduced-priced lunch (sample-specific lunch eligibility data were not available).

RTI Model

The RTI model was implemented on a school-wide basis (kindergarten through fifth grade) and is described more completely in other publications (Hilt-Panahon, Shapiro, Clemens, & Gischlar, 2011), but a brief description follows to provide context for the present study. Universal screening was conducted in early September, and the data were subsequently used to place students into one of three tiers: Tier 1 was intended for students who were meeting reading expectations and not in need of additional intervention, Tier 2 for students at some degree of risk of reading difficulties, and Tier 3 for students with more significant achievement deficits.[1] Decisions on tier placement were based on DIBELS (6th ed.; Good & Kaminski, 2007) screening data as well as achievement information from the teacher of the current and prior year (i.e., unit tests). Nearly all students placed into Tier 2 or Tier 3 interventions scored within the "some risk" or "at-risk" levels based on the DIBELS instructional recommendations for fall of first grade (Good & Kaminiski).

Regardless of tier placement, all students received 90 minutes per day of core reading instruction using the *Houghton Mifflin Reading* series (Houghton Mifflin, 2005). In addition to core instruction, 30-40 minutes were allotted each day for supplemental "skills groups" intervention in which students in Tiers 2 or 3 received small-group supplemental reading intervention. Tier 2 intervention utilized the *Language Circle Project Read* (Greene & Enfield, 1997) program administered in groups of 8-10 students; Tier 3 intervention consisted of the *My Sidewalks on Reading Street* program by Scott Foresman (Juel, Paratore, Simmons, & Vaughn, 2008) delivered in groups of 3-6 students. Intervention skills groups were implemented four to five times per week, with one day typically allotted for progress monitoring data collection. All students included in the study received either Tier 2 or Tier 3 supplemental interventions at the start of first grade. Throughout the study, all first grade teachers had access to their students' NWF and WIF progress monitoring data in the form of line graphs. We did not have detailed information on instructional changes or the degree to which teachers based decisions on the NWF or WIF data (or other information, such as classroom achievement or behavior), but tier movement was minimal; 96% of the students remained in either tier 2 or tier 3 intervention during progress monitoring data collection, and by the end of the school year 74% of the sample remained in tier 2 or tier 3 interventions.

Progress monitoring Measures

NWF. This study used NWF progress monitoring probes from the DIBELS Sixth Edition, which was the current edition at the time of the study. The measure consists of a list of VC and CVC pseudowords in which students were instructed to read by naming the sounds of each letter or by reading the whole word. Consistent with the correct letter sequences (CLS) scoring method of the DIBELS Sixth Edition[2], a student's score on the measure was the number of letter sounds produced correctly in one minute, either in isolation (e.g., /t/ /i/ /b/), as part of segments (e.g., /t/ /ib/), or as a whole word (e.g., /tib/). NWF test-retest reliability ranges from .84 to .90 (Good et al., 2009), and the two-week alternate-form reliability in our sample was .80.

WIF. WIF (Fuchs et al., 2004) is a measure of fluency reading lists of high-frequency words. The WIF measure used in this study was obtained from the test authors. Each WIF probe consisted of 100 words, arranged in vertical lists, sampled from the Dolch preprimer, primer, and first-grade lists. Students were scored according to the number of words read correctly in one minute. Two-week alternate-form reliability in our sample was .81.

Spring First-Grade Reading Outcome Measures

DIBELS Oral Reading Fluency (D-ORF). D-ORF (Good & Kaminski, 2007) is a measure of text reading fluency similar to CBM-R. Within DIBELS, D-ORF is used as a screening measure at benchmark assessment points (fall, winter, spring). At each benchmark administration students read three passages of grade-level readability. A D-ORF score consists of the median number of correctly read words in one minute from three probes. The median alternate-form reliability of the D-ORF passages is .94 (Good et al., 2004).

Test of Word Reading Efficiency (TOWRE). The TOWRE (Torgesen et al., 1999) is a standardized, norm-referenced measure that assesses word reading accuracy and fluency on two subtests: Sight Word Efficiency (SWE), which consists of a list of real words; and Phonemic Decoding Efficiency (PDE), which consists of a list of decodable pseudowords. PDE and SWE both contain words that progress in difficulty from VC and CVC words to words of up to three to four syllables. Students are scored according to the number of words read correctly in 45 seconds. SWE and PDE have demonstrated test-retest reliabilities of .96 and .90, respectively, and alternate-form reliabilities of .97 (Torgesen et al.). Raw scores from the SWE and PDE subtests were used in the correlational and structural equation model analyses. For latent class analyses, scores were converted to grade-based standard scores in order to categorize student performance according to percentile achievement.

Maze. A variation of the cloze procedure, CBM Maze consists of a reading passage in which every seventh word is deleted and replaced by three choices, only one of which correctly completes the sentence. Students read the passage silently and circle the word that best completes each sentence, and the measure is scored according to the number of correct word choices in 3 minutes. This study used Maze measures from the AIMSweb system. Administered in second grade, Maze has demonstrated average test-retest reliabilities of .83, .87, and .80 across one, two, and three months, respectively (Shin, Deno, & Espin, 2000).

Procedures

All data were collected by university research assistants and school personnel trained in administration of the measures. Data collectors were trained by the first author using written instructions and materials from the administration and scoring manuals for the respective measures; these were also provided to each data collector during testing. Training included practice administrations with the lead author to ensure administration fidelity.

Data collection. NWF and WIF were administered as progress monitoring measures with students beginning in early October of the school year. As part of the RTI model implemented in the participating schools, students in Tier 3 received progress monitoring once per week, whereas students in Tier 2 received progress monitoring once every two weeks. Analyses utilized data from an 11-week period through mid-December. This limited time frame (as opposed to a full year) was selected based on the idea that in practice, ongoing instructional decisions would be based on data collected across a series of preceding weeks (i.e., 6-10 weeks), not the full year. Additionally, schools might make use of the mid-year point to evaluate the need for instructional changes and slope might be used within these decision-making processes.

All spring reading outcome measures were administered at the end of first grade (late May). With the exception of Maze, administration took place in a quiet room outside the students' classrooms. Maze measures were administered on a whole-class basis by the classroom teacher using standard procedures described by Shinn and Shinn (2002).

Interscorer agreement. Approximately 5% of the progress monitoring administrations of NWF and WIF, and 20% of the D-ORF, SWE, and PDE measures were assessed for interscorer agreement using audio recordings scored by independent observers. Interscorer agreement was calculated on a word-by-word (or sound-by-sound for NWF) basis. Percentage of total agreement was determined by dividing the number of occurrence and nonoccurrence agreements by the number of occurrence and nonoccurrence agreements plus disagreements and multiplied by 100. All Maze measures were checked to ensure accurate tallies of correct responses. Agreement data for D-ORF, PDE, SWE, and Maze were utilized from a screening study (Clemens et al., 2011) with all first-grade students from which the sample was drawn. Average agreement for each measure was as follows: NWF = 98% (range = 80%-100%), WIF = 99% (range = 82%-100%), D-ORF = 98% (range = 90%-100%), and SWE = 98% (range = 88%-100%), PDE = 95% (range = 86%-100%). Maze was group-administered, and administration fidelity was assessed by the first author via direct observation of over half (56%) of the administrations using a standard checklist (Shinn & Shinn, 2002). Teachers implemented an average of 96% of the Maze steps correctly (range = 82%-100%).

Results

The descriptive statistics for the NWF and WIF measures across each data collection point, as well as the spring reading outcome measures are displayed in Table 1. Not all students were administered the progress monitoring measures each week indicated, as some students were monitored weekly and others were monitored once every two weeks (see "missing data procedures" below). The latent growth models (LGMs) included data collected across the 11 weeks between early October and mid-December. Data points from Weeks 7 and 8 were eliminated due to a holiday break that interrupted data collection and instruction.

Data inspection revealed that WIF scores were highly skewed at several time points with skewness statistics over 2.00. LGM analyses assume normally distributed variables (Duncan, Duncan, & Stryker, 2006); thus, a square root transformation was applied to the WIF data points (the transformation was applied to all WIF data points for consistency), which was effective at reducing skew. These scores were used in the WIF LGM analyses. NWF scores demonstrated skewness statistics within normal tolerances and were not transformed.

Latent Growth Models Analyses

Spring reading skills latent variable. To provide an index of reading skills at the end of first grade, a latent variable was specified that was measured by the four criterion variables (D-ORF, SWE, PDE, Maze). A best fitting model was achieved by correlating the error terms of SWE and PDE, which was not unexpected given that they are subtests of the same measure (TOWRE) and have similar content and administrative qualities. The final reading outcome model fit the data well, χ^2 (2) = 2.20, p = .33; comparative fit index (CFI) = 1.0, Tucker Lewis index (TLI) = 1.0, root mean square error of approximation (RMSEA) = .04, standardized root-mean square residual (SRMR) = .01. The resultant

Table 1. Descriptive Statistics for NWF and WIF Data Points, Spring Reading Measures, and Correlations With Spring Reading Outcome Measures.

						*r**			
Variable	*M*	*SD*	Range	Skewness	Kurtosis	ORF	SWE	PDE	Maze
NWF1	20.59	10.10	0 – 43	−.29	−.24	.55	.57	.39	.38
NWF2	23.00	11.62	0 – 47	−.26	−.45	.58	.63	.49	.33
NWF3	23.35	11.34	0 – 56	.16	.87	.53	.52	.34	.27
NWF4	25.34	14.22	0 – 54	.23	−.43	.54	.61	.48	.32
NWF5	30.57	12.75	2 – 55	−.45	−.37	.61	.68	.57	.23 *ns*
NWF6	34.46	17.25	3 – 71	.04	−.41	.63	.70	.61	.30
NWF9	37.22	17.80	4 – 73	−.09	−.58	.70	.75	.66	.42
NWF10	36.63	15.35	2 – 69	−.45	−.39	.61	.65	.43	.26 *ns*
NWF11	39.63	15.80	5 – 76	−.21	−.02	.66	.75	.61	.30
WIF1	3.30	3.12	0 – 15	1.92	4.19	.54	.48	.20 *ns*	.47
WIF2	2.95	3.21	0 – 15	1.81	4.00	.56	.42	.26	.51
WIF3	4.93	4.76	0 – 26	2.66	8.58	.56	.48	.19 *ns*	.47
WIF4	5.82	5.66	0 – 24	1.57	1.96	.66	.57	.48	.47
WIF5	5.08	6.89	0 – 36	2.61	7.50	.55	.38	.04 *ns*	.42
WIF6	7.75	7.39	0 – 33	1.53	2.30	.73	.60	.46	.52
WIF9	9.65	7.52	0 – 36	1.54	2.60	.80	.69	.60	.53
WIF10	11.98	9.96	0 – 57	2.35	7.62	.78	.65	.36	.53
WIF11	14.47	10.11	0 – 40	.77	−.02	.75	.68	.53	.57
WIF1 SQRT	1.62	.83	0 – 3.87	.40	.67	.57	.54	.28	.45
WIF2 SQRT	1.43	.97	0 – 3.87	.27	−.24	.46	.37	.21 *ns*	.41
WIF3 SQRT	2.03	.92	0 – 5.10	.95	2.47	.59	.55	.30	.48
WIF4 SQRT	2.15	1.11	0 – 4.90	.55	.12	.66	.60	.49	.49
WIF5 SQRT	1.87	1.27	0 – 6.00	1.12	1.58	.57	.44	.12 *ns*	.46
WIF6 SQRT	2.46	1.32	0 – 5.74	.32	−.06	.71	.64	.48	.51
WIF9 SQRT	2.88	1.17	0 – 6.00	.39	.52	.78	.72	.59	.49
WIF10 SQRT	3.21	1.31	0 – 7.55	.62	1.90	.80	.73	.44	.49
WIF11 SQRT	3.50	1.51	0 – 6.32	−.45	.29	.70	.69	.50	.50
ORF	33.25	22.00	3 – 91	.83	−.12	—	.91	.69	.68
SWE	28.10	12.71	3 – 56	.12	−.65		—	.77	.58
PDE	10.94	6.71	0 – 30	.56	−.08			—	.46
Maze	3.50	3.39	0 – 13	.99	.18				—

Note. *N* = 80. Data on the Maze measure were missing for two students. The NWF and WIF progress monitoring data points reflect data for students who were monitored at the specified week. All students were not monitored each week; therefore, *N* for each data point ranges from 54–80 students, and differences in correlations may vary due to varying sample sizes at each point. NWF = Nonsense Word Fluency, WIF = Word Identification Fluency, SQRT = variable following square root transformation, ORF = Oral Reading Fluency, SWE = Sight Word Efficiency, PDE = Phonemic Decoding Efficiency. Numbers following the NWF and WIF variables indicate the week of measurement.

*Pearson correlations of the variables with the spring outcome measures. All correlations were significant at $p < .05$ unless noted.

latent construct of spring reading served as the outcome variable in the LGM analyses of NWF and WIF.

NWF and WIF growth curve models. An LGM approach was chosen for the analyses, as it affords the opportunity to model growth across repeated assessments and to investigate relationships between intercept, slope, and other variables (Duncan et al., 2006). LGMs were specified to model the growth of NWF and WIF, and to investigate the strength at which NWF and WIF slopes were associated with spring reading skills. Since the degree to which slopes were associated with first-grade reading outcomes was of primary interest, not the shape of the NWF and WIF trajectories over time, we did not make a-priori assumptions about the shape of the trajectories but allowed them to be freely estimated (Duncan et al.). In so doing, we fixed the factor loading for the first measurement point at zero and the last measurement point at 1, and allowed all intervening factor loadings to be freely estimated.

Analyses were conducted using Mplus Version 6.1 (Muthén & Muthén, 2010). Model fit was assessed using the CFI, TLI, RMSEA, and SRMR fit indices. Acceptable fit values were considered greater than .95 for CFI and TLI, less than .08 for RMSEA, and less than .08 for SRMR (Browne & Cudeck, 1993; Hu & Bentler, 1999). The R^2 statistic was used as an index of the amount of variance in the spring outcome variable explained by the predictors in the model (e.g., intercept and slope), and was calculated as 1 minus disturbance (residual variance) when the factor variance was set to 1.

Missing data procedures. Modeled parameters were estimated using the full information maximum likelihood (FIML) estimation procedure, which is robust for handling missing data (Arbuckle, 1996; Enders, 2010). The data collection schedule stipulated that data be collected for all Tier 2 students once every two weeks and once per week for all Tier 3 students. Therefore, Tier 2 students had missing data every other week. Because these data were missing due to a planned design element of the data collection model (to increase feasibility of data collection by teachers), and the reason data were missing was not a function of the values that would have been present on those occasions, the missing data points were treated as missing at random (Schafer & Graham, 2002). Other missing data points were the result of student absences, and it was therefore reasonable to believe they were also missing at random. In the final data set used to estimate the growth curve models, the average number of data points per student was 6.94 for NWF (range = 4-9, SD = 2.11), and 6.92 for WIF (range = 3-9, SD = 2.05) across the 11 weeks of the study.

Latent Growth Models – NWF

The NWF growth model, with intercept and slope factors modeled as predictors of spring reading skills, is displayed in Figure 1. This model demonstrated overall acceptable fit to the data ($\chi^2(68) = 110.55$, $p < .01$; CFI = .96, TLI = .95, RMSEA = .09, SRMR = .08). NWF intercept and slope were significantly and positively correlated (.42, $p < .001$), indicating that students with higher initial NWF scores tended to demonstrate higher rates of growth. Both NWF intercept ($\beta = .46$, $p < .01$) and slope ($\beta = .39$, $p < .01$) significantly predicted spring reading, and these predictors explained approximately 51% of the variance in the spring reading latent variable. The slope factor was a reliable summary of the NWF scores across time: the mean R^2 for variance in NWF scores accounted for by the slope factor was .84.

NWF slope as independent predictor. In order to test the independent effects of NWF slope predicting the spring reading outcome variable, a subsequent model was tested that removed NWF intercept as a predictor (i.e., setting the path weight at zero). In this model, the path coefficient for NWF slope predicting spring reading increased to .75 ($p <$ 0.01), and approximately 56% of the variance in spring reading was explained by NWF slope as a predictor. This model demonstrated marginally good fit ($\chi^2(69) = 120.52$, $p < .01$; CFI = .95, TLI = .94, RMSEA = .10, SRMR = .09) and fit significantly worse than the model that included intercept as a predictor ($\Delta\chi^2(1) = 9.97$, $p < .01$); however, poorer fit was expected due to the elimination of an important predictor.

Latent Growth Models –WIF

The WIF growth model predicting the spring latent reading outcome variable (see Figure 2) demonstrated marginally acceptable fit with the data ($\chi^2(68) = 111.43$, $p < .01$; CFI = .95, TLI = .95, RMSEA = .09, SRMR = .09). A positive correlation between intercept and slope (.70, $p < .01$) suggested that higher initial WIF scores were associated with higher WIF slopes. Spring reading was significantly predicted by WIF intercept ($\beta = .27$, $p = .04$) and slope ($\beta = .63$, $p < .01$), and these predictors accounted for approximately 67% of the variance in the spring reading outcome variable. The slope factor was a reliable summary of the WIF scores across time: the mean R^2 for variance in WIF scores accounted for by the slope factor was .77.

WIF slope as independent predictor. To investigate the independent contribution of WIF slope predicting spring reading, the next model removed WIF intercept as a predictor. In this model, the path coefficient for WIF slope predicting spring reading was .85 ($p < .01$), which accounted for 73% of the variance in the spring reading outcome variable. Marginally good fit was observed for this model ($\chi^2(69) = 113.62$, $p < .01$; CFI = .95, TLI = .95, RMSEA = .09, SRMR = .09), which did not differ significantly from the model that included intercept ($\Delta\chi^2(1) = 2.19$, $p = .001$).

LGM summary. The important statistics associated with each model are summarized in Table 2. As illustrated, both NWF and WIF slope were positively and significantly associated with spring reading skills, as were intercepts. However, although NWF explained a good deal of variance in spring reading, most variance in spring reading skills was accounted for by models that included WIF. WIF slope appeared to be more strongly related to reading outcomes than NWF slope based on the stronger relative path weights observed for WIF slope and greater R^2 values when only considering slopes as predictors. In short, results suggested that WIF slope was a stronger predictor of year-end reading skills than NWF slope.

Growth Plots and Latent Class Analysis

As a follow-up to the LGM analyses, and to further explore how slopes discriminated between students according to later reading outcomes, individual students' slopes were plotted and coded as a function of attaining a reading outcome criterion at the end of first grade. These analyses explored NWF and WIF slopes as they might be viewed by teachers collecting progress monitoring data. In practice, several popular web-based CBM programs (e.g., AIMSweb,

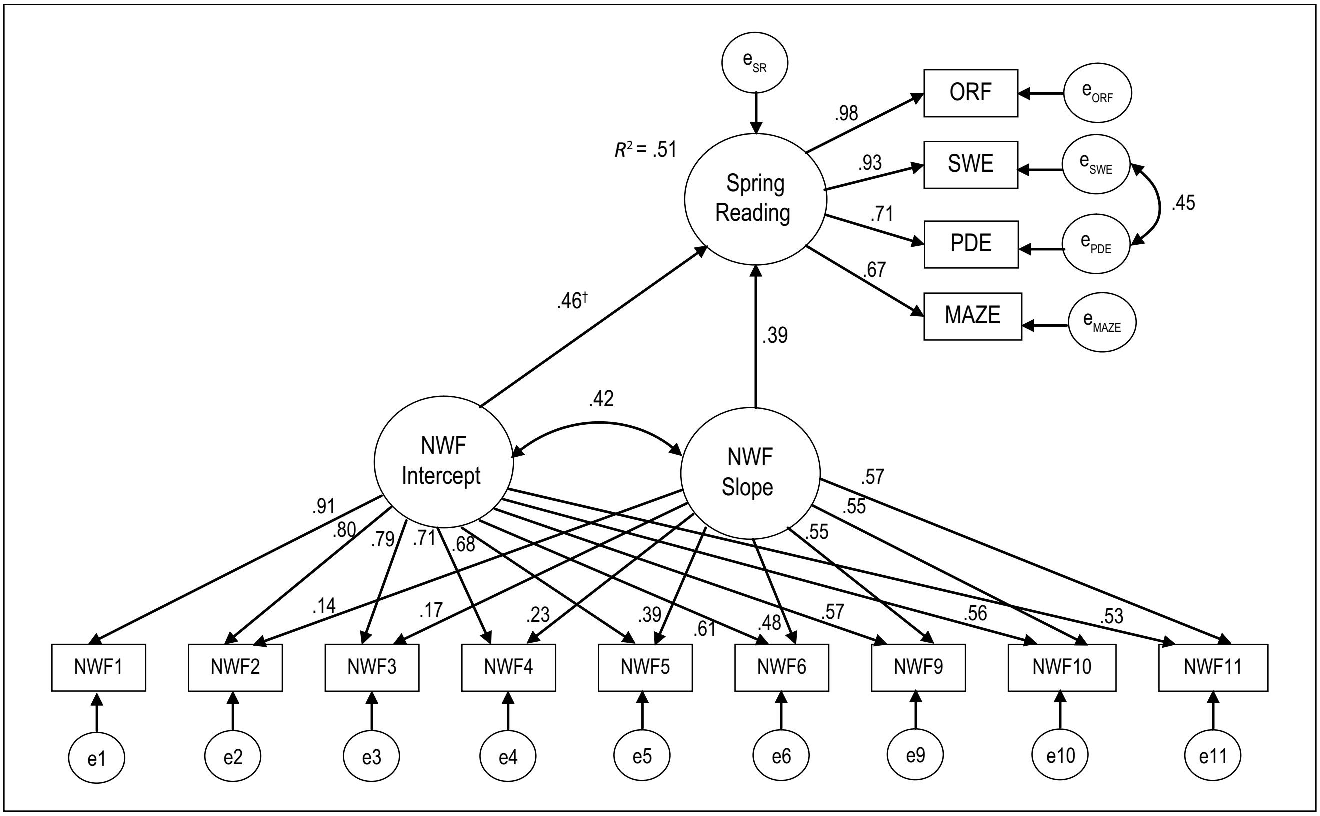

Figure 1. NWF latent growth curve model predicting year-end reading skills. All loadings and coefficients are shown in standardized form. In estimating the model, loadings on the intercept factor were all set to 1 and loadings on the slope factor were set to 0 for NWF1, 1 for NWF11, and freely estimated for all time points in between. Loadings in the figure differ from these values because of standardization of NWF scores, which had differing variances across time. Residual variance values are omitted to make the figure easier to read.

†For the model that investigated the independent contribution of slope in predicting spring reading, the structure coefficient between intercept factor and spring reading was fixed at 0.

DIBELS data system, EasyCBM) summarize student progress using a linear trend line (i.e., straight line). Thus, we fit linear trend lines (ordinary least-squares) to each student's NWF and WIF data. Although this differed from our LGM analyses, in which student trajectories were unconstrained and freely estimated, it was considered important to display student slopes in the manner in which they would most often be viewed by teachers on a regular basis. Linear slopes were a reliable index of student growth, as the median R^2 values were .70 and .72 for NWF and WIF slopes, respectively.

We compared students' slopes according to whether they demonstrated "successful" or "unsuccessful" reading outcomes at the end of first grade. To do so, we dichotomously categorized students using their scores on the reading outcome measures using a latent class analysis (LCA; Thissen, 1989) with a binary latent factor to categorize students, according to relative standing above or below 30th percentile from national norms. The 30th percentile has been recommended as a criterion level for judging student success across measures of achievement (Torgesen, 2000) and has been used in studies to indicate risk status in reading (e.g., Simmons et al., 2008).

Students' scores on each of the outcome measures were first dichotomously coded according to achievement above or below the 30th percentile compared to national normative data for each measure, which were 35 words read correctly for ORF (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002), a standard score of < 93 for SWE and PDE[3], and a score of 5 on Maze (AIMSweb, 2010). Using these cutoffs, the base rates of students with scores falling *below* the 30th percentile on each respective measure was 63% for ORF, 46% for SWE, 53% for PDE, and 70% for Maze. Using Mplus 6.1, the LCA then generated the probability of falling within the upper (above 30th percentile) or lower (below 30th percentile) latent class for each student. If the probability of

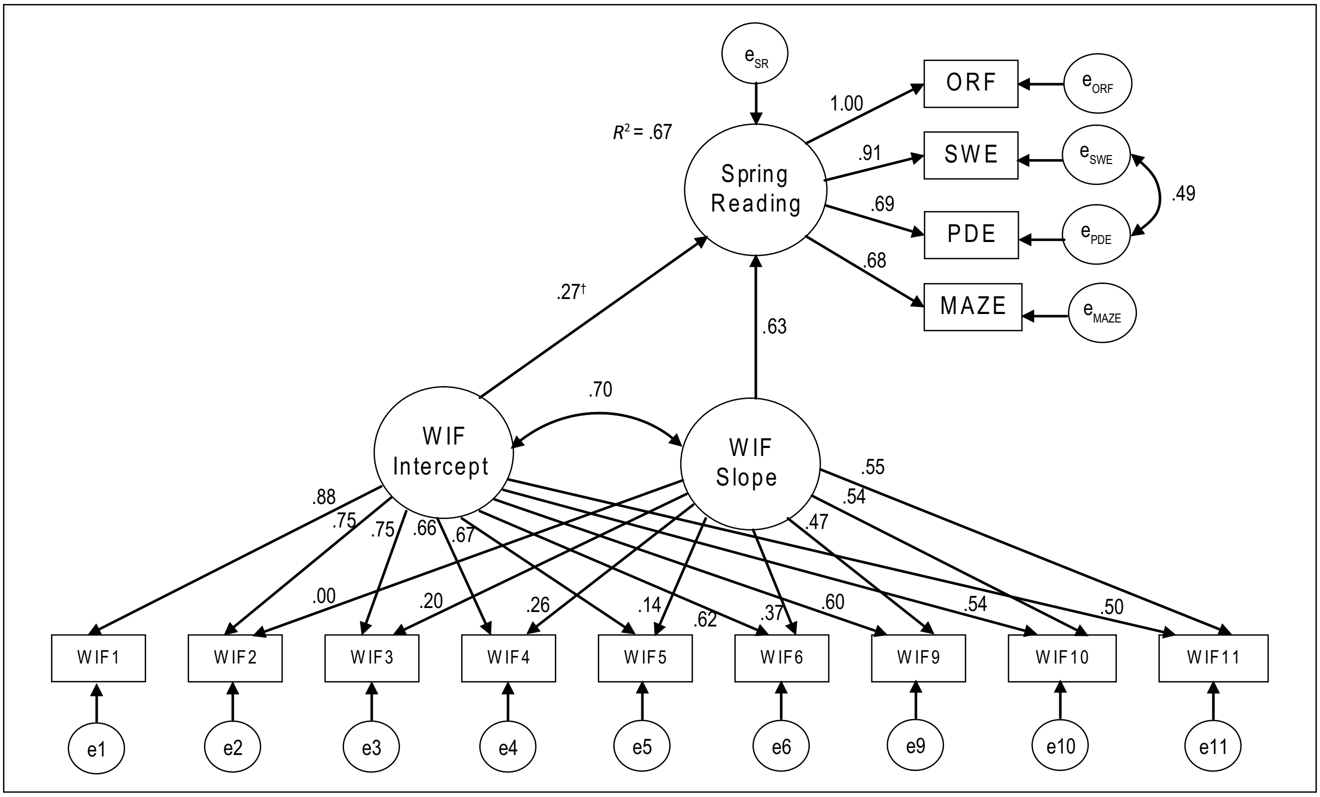

Figure 2. WIF latent growth curve model predicting year-end reading skills. All loadings and coefficients are shown in standardized form. In estimating the model, loadings on the intercept factor were all set to 1 and loadings on the slope factor were set to 0 for WIF1, 1 for WIF11, and freely estimated for all time points in between. Loadings in the figure differ from these values because of standardization of WIF scores, which had differing variances across time. Residual variance values are omitted to make the figure easier to read. WIF scores were all square root transformed prior to estimation of the model to reduce their skewness.
†TFor the model that investigated the independent contribution of slope in predicting spring reading, the structure coefficient between intercept factor and spring reading was fixed at 0.

Table 2. Summary of Latent Growth Model Path Coefficients, R^2 Values, and Model Fit Statistics for NWF and WIF Growth Models Predicting the Spring Reading Latent Outcome Variable.

				Model Fit			
Model	Intercept β	Slope β	Spring Reading R^2	CFI	TLI	RMSEA	SRMR
NWF intercept and slope	.46**	.39**	.51	.96	.95	.08	.07
WIF intercept and slope	.27*	.63**	.67	.95	.95	.08	.07
NWF slope only	—	.75**	.33	.95	.94	.09	.08
WIF slope only	—	.85**	.67	.95	.95	.09	.09

Note. NWF = Nonsense Word Fluency; WIF = Word Identification Fluency.
* $p < .05$
** $p < .01$

being in the upper group was greater than 0.5, students were categorized as achieving "successful" reading outcomes; the remaining students were categorized in the "unsuccessful" outcomes group. Among the 80 students in the sample, 39 students (48.8%) were classified in the successful outcomes group, and the remaining 41 (51.2%) students were classified in the unsuccessful outcomes group.[4] Spaghetti plots were then created with the corresponding linear slopes for each student, which were color-coded such that students who demonstrated successful achievement were depicted with black lines, and students with unsuccessful achievement depicted with grey lines. This dichotomous distinction

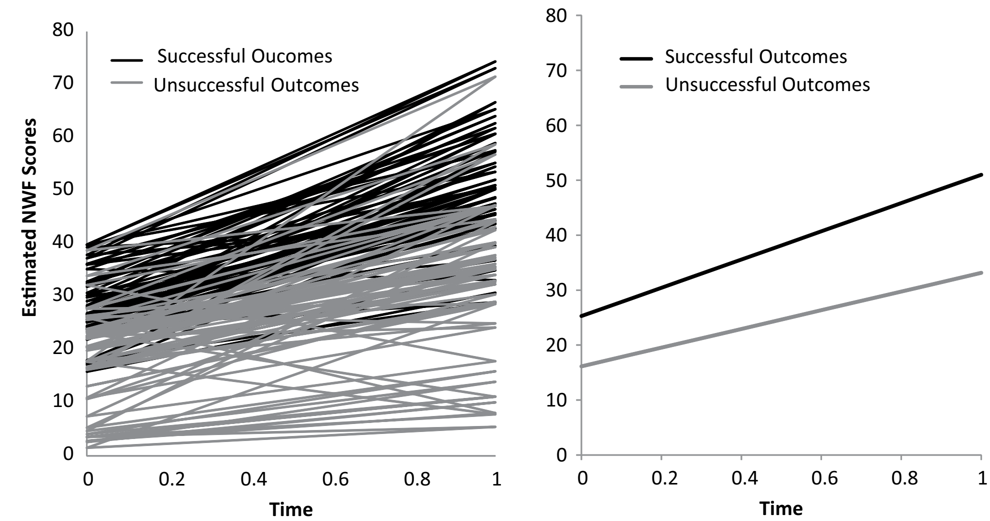

Figure 3. Students' estimated NWF slopes (left panel) and average estimated NWF slopes (right panel) coded according to year-end reading outcomes.

permitted exploration of the degree to which NWF and WIF slopes discriminated according to year-end outcomes.

The NWF spaghetti plot is displayed in Figure 3 (left panel). Figure 3 also includes the average slopes for students categorized according to year-end achievement (right panel). Students in the two groups appear to be more clearly distinguished in terms of initial level; that is, year-end group membership appears to be more clearly distinguished by their initial NWF scores as opposed to slope. This is consistent with the growth model results indicating that intercept was a somewhat stronger predictor of reading outcomes than slope. A cross-hatched pattern suggests that students who would later be considered above 30th percentile on the reading outcome measures tended to display slopes that were similar to those of students who would score below the 30th percentile at year end. In other words, there does not seem to be a clear pattern whereby slopes discriminate between the two groups. In addition, several students demonstrated very high rates of growth but later scored *below* the 30th percentile on the criterion outcome measure at the end of the year. Those students had some of the highest slopes of the group, but their rapid rate of improvement did not translate into successful reading outcomes.

In comparison, WIF individual student slopes (see Figure 4, left panel) displayed more of a fan-shaped pattern. Students who later achieved successful outcomes tended to display higher slopes than students who would later demonstrate unsuccessful outcomes. The average WIF slopes displayed by the two groups (Figure 4, right panel) also appear to show a somewhat larger discrepancy than they did for NWF slope.

Group differences were further tested using *t* tests and effect size calculations (Cohen's *d*; 1988), which are summarized in Table 3. For these analyses, slope was converted to rate of improvement (ROI) by dividing the slope estimate by the total number of weeks students were monitored (11), yielding the number of sounds (NWF) or words (WIF) gained per minute per week. As indicated in Table 3, group differences were greatest for WIF slope. Students with successful outcomes on the year-end reading outcomes variable demonstrated significantly higher intercept and slope than students with unsuccessful outcomes on both NWF and WIF. However, effect sizes indicated a greater difference in slopes between the two groups for WIF than for NWF. WIF slope (ROI) for the students with successful outcomes was more than 2.69 times greater than the ROI for the students with unsuccessful outcomes. By comparison, students with successful outcomes demonstrated an average NWF ROI of 1.51 times greater than the students with unsuccessful outcomes. In summary, the groups differed on both measures with regard to both initial level and slope, but group differences were most pronounced for WIF slope.

Discussion

This study compared NWF and WIF as measures for monitoring the reading progress of at-risk first-grade students, specifically, the degree to which slope was associated with reading skills at the end of first grade. Both NWF and WIF fall slopes were positively and significantly associated with year-end reading outcomes, which extends prior work

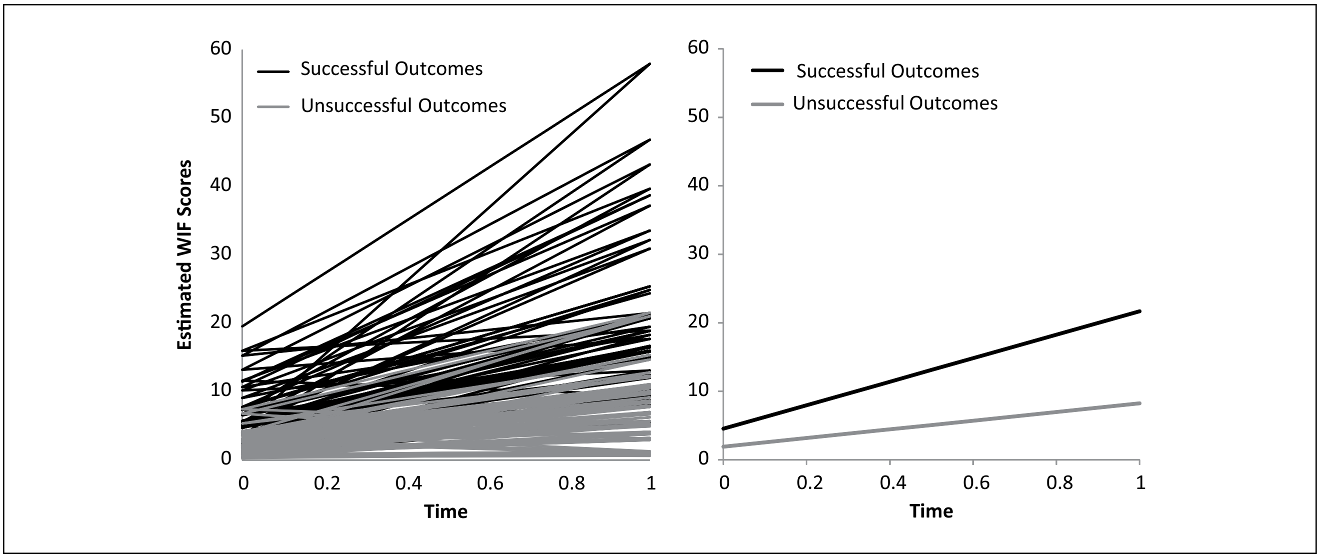

Figure 4. Students' estimated WIF slopes (left panel) and average estimated WIF slopes (right panel) coded according to year-end reading outcomes.

Table 3. Comparison of Group Differences on NWF and WIF Intercept and Slope.

	Below 30th Percentile	Above 30th Percentile				
Variable	*M(SD)*	*M(SD)*	*t*	*df*	*p*	*d*
NWF Intercept	16.14 (8.75)	25.27 (6.10)	5.38	78	< .01	1.21
NWF ROI (slope)	1.55 (0.69)	2.34 (0.71)	5.00	78	< .01	1.12
WIF Intercept	1.91 (1.26)	4.53 (3.38)	4.64	78	< .01	1.03
WIF ROI (slope)	0.58 (0.32)	1.56 (0.79)	7.37	78	< .01	1.65

Note. Below 30th percentile = students whose year-end reading achievement on a latent outcome variable was below the 30th percentile (*n* = 41); Above 30th Percentile = students whose reading achievement on a latent outcome variable was above the 30th percentile (*n* = 39). NWF = Nonsense Word Fluency; WIF = Word Identification Fluency. ROI = weekly rate of improvement, which reflects the number of sounds/words gained per minute per week on the respective measures. NWF scores are the number of correct letter sounds per minute, and WIF scores are the number of words read correctly in one minute.

that has demonstrated the validity of NWF and WIF slope, particularly for low-achieving students (Harn et al., 2009; Compton et al. 2006; 2010; Cummings et al., 2011; Zumeta et al., 2011) . However, consistent with Fuchs et al. (2004), WIF slope was more strongly associated with these outcomes than NWF slope; it accounted for more variance in the spring outcome variable than NWF slope both when slope was used as the sole predictor and when initial level was accounted for. Further, differences in WIF slope were more pronounced over time between students considered to be successful and unsuccessful readers at the end of the year.

A primary purpose of CBM is to monitor student progress for evaluating and improving instruction (Deno, 2003; Deno & Mirkin, 1977). Inherent in this idea is that student progress as reflected by the CBM measure is meaningful; that is, that growth is indicative of overall skill acquisition. Slope is important for instructional decision-making (Jenkins & Terjeson, 2011), and a key assumption for CBM in reading is that an increasing pattern of scores is indicative of growth in overall reading proficiency.

The results of the present study suggest that although both NWF and WIF slope were associated with later reading skills, WIF may provide a clearer index of reading development and more defensible data on which to signal the need for instructional changes for at-risk students, and may more clearly differentiate students who are not on pace to meet year end reading goals.

Validity of NWF and WIF Slopes as Indices of Reading Development

The development of reading skills provides a context for understanding the relevance of NWF and WIF as progress monitoring tools. Automatic word recognition skills, critical for skilled reading, result from the development of several interdependent pre- and early-literacy skills, a process that has been described in several seminal works (e.g., Adams, 1990; Ehri, 2002; Perfetti, 1984, 1985; Share, 1995, 2008; Snow, Burns, & Griffin, 1998). This process is not made possible by memorizing word shapes or entire word forms, but through a connection-forming process in which an understanding of letter-sound relationships is used to ultimately link letter strings to pronunciations in memory. Combined with basic skills in phonemic awareness (particularly that words have beginning, middle, and ending sounds, and that sounds can be blended to form a word), letter-sound knowledge can be utilized to begin to decode printed words, thus providing the basis for a powerful self-teaching mechanism in which unfamiliar words can be deciphered using available alphabetic information (i.e., "phonological recoding;" Share, 1995, 2008).

Over time, reading increasingly becomes a process of "unitization" or "modularization" (Adams, 1990; Ehri, 2002) in which letter sounds are unitized into larger and larger chunks and stored in long-term memory. Subsequently, words are processed on an increasingly more holistic basis where letter combinations are linked to pronunciations in memory and read very quickly. The alphabetic code forms the all-important foundation for this process by providing the mechanism on which unfamiliar words can be deciphered, and the mapping system to quickly add words to the learner's orthographic repertoire. According to Perfetti (1984), "the identification of words is mediated by the perception of letters" (p. 46). Quite simply, beginning readers process individual letters to decode words, and as skills develop, more advanced readers process words as units of letters and match them to pronunciations in memory (Ehri).

With this process in mind, NWF can be considered sensitive to skills with the alphabetic principle, and growth is suggestive of developing mastery of letter-sound associations—skills that are critical for decoding, and ultimately, word reading. This CLS scoring method (the scoring approach that has been the most widely studied and used across available NWF measures) is likely to be sensitive to students' fluency with identifying letter sounds within the context of pseudowords and beginning skills in decoding. Alphabetic knowledge is truly vital in the development of word reading skills, and NWF provides a system to assess growth with this critical foundational skill.

Although NWF may provide a good index of knowledge and fluency with letter-sound correspondence, it may yield ambiguous data on developing word reading skills. For example, one student may achieve a score of 35 on NWF by reading only letter sounds, and another student may achieve the very same score by reading words as whole units. The latter student is demonstrating more sophisticated skills in word reading, and is likely a stronger reader. In a similar way, two students may demonstrate rapid rates of growth; however, one student may be building fluency in simply naming letter sounds, while the other may be developing fluency in reading the NWF words as whole units. Although the growth rates of both students are equal and high, the skills demonstrated by the two students are clearly dissimilar, as slope for the latter is likely to be associated more closely with overall skills in word reading. Further complicating this picture is evidence that students can change their patterns of responding to NWF stimuli over time (Ritchey, 2008), something that is not apparent in a simple NWF CLS score. As observed in this study, high growth on NWF was not always associated with year-end reading success, thus presenting potentially misleading information about overall reading acquisition.

In contrast, WIF might be viewed as an index of more consolidated reading skills; a measure with the potential to assess general orthographic knowledge, and word reading skills of a more overall variety. As discussed earlier, rapid and accurate word reading is facilitated by early foundational skills in the alphabetic principle and phonological awareness that drive decoding (Aaron et al., 1999; Perfetti, 1985), and students become successful readers by consolidating individual letters sounds into larger units (Ehri, 2002). Even when reading phonologically irregular words, skills in the alphabetic principle are still utilized since many phonetically irregular words consist of a high percentage of letters that correspond to their most common sounds. The important point is that fluency of word reading, in which students are asked to read a list of words of various lengths, various spelling patterns, and varying degree of phonetic regularity, is reflective of a host of early precursor skills that are critical for early reading. Certainly, WIF would not provide information directly specific to students' understanding of the alphabetic principle as clearly as a measure like NWF. However, WIF might be considered a more comprehensive measure and able to assess skills somewhat more "downstream", as growth in word recognition speed is indicative of a student's proficiency with several foundational literacy skills and is clearly predictive of later reading achievement. As pointed out by a reviewer of an initial version of this paper, slope of WIF *is* early reading acquisition almost by definition.

Recent studies have demonstrated that scoring NWF according to the number of items read as partial or whole words can explain additional variance in reading outcomes over NWF CLS (Cummings et al., 2011; Harn et al., 2008; but see Ritchey, 2008). Some have postulated that students' unitizing strategies on NWF are indicative of their position in the phases of reading development proposed by Ehri (2002), suggesting that changes in NWF response patterns, from naming letter sounds to reading word segments or whole units, may signify transition between reading phases and that students who demonstrate a higher degree of unitization are functioning at full or consolidated phases of word reading (Cummings et al.; Harn et al.). The latest version of the DIBELS tools, DIBELS Next, has revised the NWF measure to include a "whole words read" (WWR) score, which is scored in addition to the CLS score.

However, are obtaining these nuanced data provided by NWF WWR, which requires a deeper level of analysis by teachers, an efficient use of time? Certainly if fluency in letter sound correspondence or reading brief pseudowords is an instructional target, one could argue that it is. CBM was not originally intended to provide diagnostic information but to provide an efficient and repeatable index of overall skills that yields clear and easily interpretable information regarding student progress toward general outcomes (Deno & Mirkin, 1977; Fuchs & Deno, 1991; Stecker et al., 2005). If periodic diagnostic information on students' decoding skills is desired, a different measure containing stimuli that varies beyond the VC or CVC pattern found on NWF would be desirable by providing a more complete picture of students' decoding strengths and weaknesses. In contrast to the multiple scoring strategies available for NWF, simplicity may favor WIF. Although the level at which a student demonstrates a unitizing strategy on the NWF task might provide suggestive evidence of functioning at a full or consolidated phase, word reading skills on WIF would identify these skills more unambiguously. In general, WIF may be a simpler, clearer, and more efficient option for obtaining a brief snapshot of students' progress toward critical first grade reading objectives.

Limitations and Future Directions

The results of this study must be considered in light of several limitations. First, conclusions regarding the validity of NWF slope are limited to the CLS scoring approach. Although CLS is presently the only score that is entered when using NWF within AIMSweb, and still factors prominently in the DIBELS Next version of NWF, future research should investigate the use of the WWR method for scoring NWF and whether WWR slope is similar or superior to CLS or other early reading progress monitoring measures.

Second, a study of validity conducted in an educational context when the predictor and criterion measures are separated in time will face inherent challenges due to the intervening nature of instruction. Slope was our variable of interest which captures students' growth over time (and accounts to some degree of the effects of instruction), however instructional variables that occurred between the progress monitoring period and the end of the school year may have contributed to weaker relationships between slope and year-end outcomes. Although movement out of tiered intervention was minimal within this sample, specific information was not available on instructional changes or the degree to which NWF and WIF were used for these decisions. Factors that temper these concerns are that NWF and WIF were compared within the same sample of students, progress monitoring data for both measures were available to all teachers, and students were exposed to the same RTI model and interventions across schools, thus any weakening of the relationship between slope and year-end outcomes due to intervening instructional variables might be expected for both measures.

Third, the assessment battery used to assess reading outcomes was weighted heavily toward measures of word recognition and reading fluency. Although Maze can be considered a measure of literal comprehension, the battery did not assess comprehension more completely and cannot be truly deemed an assessment of reading "outcomes." However, our outcome battery might be considered indicative of reading skills of primary importance for acquisition in first grade; namely, the acquisition of skills in word recognition, decoding, and reading connected text. Comprehension assessment in first grade is problematic, as it is difficult to determine whether errors in comprehension are due to inaccurate or inefficient word reading rather than the result of actual comprehension difficulties. Future research might examine the degree to which slope on the measures investigated here is associated with reading comprehension skills (or difficulties) measured in later grades.

Fourth, although the LGM analyses modeled growth without a hypothesized trajectory shape, the second set of analyses displayed student slopes that were "forced" into linear trajectories. Although our reasons for using linear slope were based on the way that teachers would most likely view trend lines from progress monitoring software (e.g., AIMSweb), students' slopes in the second set of analyses may not have been modeled in ways that best represented the true shape of their trajectories.

Additional technical questions remain regarding progress monitoring for early readers. Jenkins and Terjeson (2011) demonstrated that, for CBM-R with students in grades 2-6, collecting fewer data points across a period of time (e.g., once every eight weeks) may be sufficient for prompting instructional changes while maximizing teachers' time for instruction, although other research suggests the need for more data points in order to ensure slope reliability (Christ, 2006). Subsequent research might investigate the minimum number of data points needed for instructional decision-making with NWF and WIF, as well as the sensitivity of NWF and WIF slopes to instructional changes, which is an important factor to consider when selecting a measure in addition to slope validity. Additionally, Zumeta et al. (2011) found that WIF lists that sampled broadly from a larger corpus of high frequency words demonstrated greater slope validity than more narrowly sampled WIF lists. The WIF measure used in this study was consistent with Zumeta et al.'s narrow sampling approach, thus, future research might investigate the strength of broadly-sampled WIF lists relative to other measures.

Further research is needed regarding the implications of slope for educational decision-making, particularly in early grades. Al Otaiba et al. (2011), studying students' response to kindergarten instruction, found that first-grade reading skills were predicted primarily by end-of-kindergarten reading status. Growth across kindergarten attenuated these relationships, such that more rapid growth across kindergarten was often associated with lower reading skills the end of first grade, as the students who had to grow rapidly in order to achieve successful kindergarten outcomes were still at risk by the end of first grade. Longitudinal investigations are needed to fully understand the degree to which slope is indicative of later reading outcomes, including when slope is important and for which reading skills.

Finally, questions pertaining to measure selection for early reading progress monitoring are relevant. Our goal in this study was to compare two available options for early first grade. Although recent research has focused on NWF in first grade (Cummings et al., 2011; Fein et al., 2010; Good et al., 2009; Harn et al., 2009), it should be acknowledged that NWF was designed as a screening and progress monitoring option across kindergarten *and* first grade, and as reviewed earlier, has demonstrated validity across these time periods. Measures relevant across these early grade levels may be advantageous for evaluating longitudinal growth. WIF measures might also be used across kindergarten (Lai et al., 2010), but may lack sensitivity in kindergarten compared to measures that assess alphabetic understanding more specifically. WIF is similar to CBM-R, and the number of high frequency words may make WIF an attractive option for the beginning of first grade, however, CBM-R could certainly be an option for monitoring first grade progress. Other tools might include reading passages consisting of a high percentage of phonetically regular words (Shinn, 2009), or computer-adaptive assessments (e.g., McCarthy & Christ, 2011). Certainly this is an important area of future study, and the selection of progress monitoring measures should be based on several factors including the skills targeted in instruction, technical properties, ease of use, and expected outcome skills (e.g., word recognition, reading comprehension).

Implications and Conclusions

When monitoring the progress of students with or at risk for reading difficulties, teachers must be confident that the slope of progress reflected by a progress monitoring measure is indicative of overall reading skills development, and given the number of early literacy measures presently available, educators may face uncertainty regarding the strongest measures for this purpose in early first grade. This study found that when monitoring the reading progress of at-risk first-graders, WIF slope may be more indicative of later reading skills than NWF slope. NWF slope was indeed associated with later reading outcomes; however, WIF slope was a more powerful predictor of year-end reading skills and more clearly discriminated between students who experienced successful versus unsuccessful outcomes. For teachers making important instructional decisions, and when student responsiveness to instruction is of key importance for educational decision-making, relying on the measure in which slope is more clearly related to overall outcomes may be paramount.

Acknowledgments

We thank Kirsten McBride for editorial assistance, as well as the reviewers for their constructive feedback on earlier versions of this manuscript. This study was based in part on data collected in the doctoral dissertation of the first author.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the U.S. Department of Education, Office of Special Education Programs, grant H326M050001. This material does not necessarily represent the policy or positions of the U.S. Department of Education, nor is the material necessarily endorsed by the federal government.

Notes

1. This RTI model involved placing students, based on screening assessments that indicated higher level of academic need, directly into tiers in which more intensive instructional services were provided. This model differs from RTI models in which students must demonstrate failure to respond to instruction at less intensive tiers prior to being assigned to more intensive tiers. In addition to direct placement, students could also move to more or less intensive tiers based on progress, or lack thereof.
2. Teachers monitored progress using CLS as the primary scoring metric for NWF, and it was the only score retained in the progress monitoring datasets.
3. Students' scores on the SWE and PDE measures were converted to grade-based standard scores, since word reading skills are more related to instruction as opposed to age. Grade norms for the beginning of second grade were used because first grade norms resulted in an overestimation of the achievement of the overall first grade population from which the present sample was drawn, possibly a result of the TOWRE sampling and data smoothing (J.K. Torgesen, personal communication, March 15, 2010). Using norms from the beginning of second grade were considered a more accurate estimate of achievement given that the TOWRE was administered during the last two week of first grade.

4. This should not be construed as a median split; the 50% breakdown of the data is the result of the percentages of students scoring below the 30th percentile across the four measures.

References

Aaron, P. G., Joshi, R. M., Ayotollah, M., Ellsbury, A., Henderson, J., & Lindsey, K. (1999). Decoding and sight-word naming: Are they independent components of word recognition skill? *Reading and Writing, 11,* 89-127.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

AIMSweb (2010). *AIMSweb national norms table: MAZE comprehension.* Retrieved from https://aimsweb.pearson.com/.

Al Otaiba, S., Folsum, J. S., Schatschneider, C., Wanzek, J., Greulich, L., Meadows, J., & Connor, C. M. (2011). Predicting first-grade reading performance from kindergarten response to tier 1 instruction. *Exceptional Children, 77,* 453-470.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Thousand Oaks, CA: Sage.

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42,* 163-176.

Christ, T.J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35,* 128-133.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Clemens, N.H., Shapiro, E.S., & Thoemmes, F.J. (2011). Improving the efficacy of first grade reading screening: An investigation of Word Identification Fluency with other early literacy indicators. *School Psychology Quarterly, 26,* 231-244.

Compton, D. L., & Carlisle, J. F. (1994). Speed of word recognition as a distinguishing characteristic of reading disabilities. *Educational Psychology Review, 6,* 115-140.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327-340.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98,* 394-409.

Cummings, K. D., Dewey, E. N., Latimer, R. J., & Good, R. H. (2011). Pathways to word reading and decoding: The roles of automaticity and accuracy. *School Psychology Review, 40,* 284-295.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37,* 184-192.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Minneapolis: University of Minnesota.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling concepts, issues, and applications*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Ehri, L. C. (2002). Phases of acquisition in learning to read words and implications for teaching. In R. Stainthorp & P. Tomlinson (Eds.), *Learning and teaching reading* (pp. 7-28). Leicester, UK: British Psychology Society.

Ehri, L.C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing, 7,* 295-326

Enders, C. K. (2010). *Applied missing data analysis* (1st ed.). New York, NY: The Guilford Press.

Fien, H., Park, Y., Baker, S. K., Mercier-Smith, J. L., Stoolmiller, M., & Kame'enui, E. J. (2010). An examination of the relation of Nonsense Word Fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review, 39,* 631-653.

Fletcher, J.M, Lyon, G.R., Fuchs, L.S., & Barnes, M.A. (2007). *Learning disabilities: From identification to intervention.* New York, NY: Guilford.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488-500.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21,* 449-460.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199-208.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71,* 7-21.

Fuchs, L.S ., Fuchs, D., & Hamlett, C. L. (1989). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43-52.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20-28.

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39,* 463-483.

Goffreda, C. T., DiPerna, J. C., & Pedersen, J. A. (2009). Preventative screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools, 46,* 539-552.

Good, R. H., Baker, S. K., & Peyton, J. A. (2009). Making sense of nonsense word fluency: Determining adequate progress in early first-grade reading. *Reading and Writing Quarterly, 25,* 33-56.

Good, R. H., & Kaminski, R. A. (Eds.). (2007). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., Smith, S., & Flindt, N. (2004). *Technical adequacy and decision making utility of DIBELS* (Technical Report No. 7). Eugene: University of Oregon.

Good, R. H., Wallin, J. U., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). *System-wide percentile ranks for DIBELS benchmark assessment* (Technical Report No. 9). Eugene: University of Oregon.

Gough, P.B. (1996). How children learn to read and why they fail. *Annals of Dyslexia, 46,* 3-20.

Greene, V. E., & Enfield, M. E. (1997). *Project read.* Bloomington, MN: Language Circle Enterprises.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41,* 143-157.

Hilt-Panahon, A., Shapiro, E.S., Clemens, N.H., & Gischlar, K.L. (2011). The structure and content of the RTI model. In E.S. Shapiro, N. Zigmond, T. Wallace, & D. Marston (eds.), *Models for implementing response to intervention: Tools, outcomes, and implications* (pp. 11-45). New York: Guilford Press.

Houghton Mifflin. (2005). *Houghton Mifflin reading.* Boston, MA: Author.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. *Learning Disabilities Research and Practice, 18,* 237-245.

Jenkins, J.R., Hudson, R.F., & Johnson, E.S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36,* 582-600.Jenkins, J., & Terjeson, K.J. (2011). Monitoring reading growth: Goal setting, measurement frequency, and methods of evaluation. *Learning Disabilities Research and Practice, 26,* 28-35.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). Can we improve the accuracy of screening instruments? *Learning Disabilities Research and Practice, 24,* 174-185.

Juel, C., Paratore, J. R., Simmons, D., & Vaughn, S. (2008). *My sidewalks on reading street: Intensive reading intervention.* Glenview, IL: Pearson Scott Foresman.

Kim, Y. S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102,* 652-667.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 62*, 293-323.

Lai, C.F., Nese, J.F.T., Jamgochian, E.M., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM primary-level reading measures (grades K-1), 2009-2010 version (Technical Report #1003).* University of Oregon: Behavioral Research and Teaching.

McCarthy, A. & Christ, T. J. (2011, February). *Computer Based Assessment System for Reading (CBAS-R)*, Paper presented at the annual conference for the National Association of School Psychologists, San Francisco, CA.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

Nathan, R. G., & Stanovich, K.E. (1991). The causes and consequences of differences in reading fluency. *Theory Into Practice, 30*, 176-184.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Institute for Literacy.

Perfetti, C. A. (1984). Reading acquisition and beyond: Decoding includes cognition. *American Journal of Education, 93,* 40-60.

Perfetti, C. A. (1985). *Reading ability.* New York, NY: Oxford University Press.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*, 427-469.

Ritchey, K. D. (2008). Assessing letter-sound knowledge: A comparison of letter sound fluency and nonsense word fluency. *Exceptional Children, 74,* 487-506.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147-177.

Share, D. L. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition, 55,* 151-218.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. *Advances in Child Development and Behavior, 36,* 31-82.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the Maze task for curriculum-based measurement of reading growth. *Journal of Special Education, 34,* 164-172.

Shinn, M.R. (2009, February). *Advances in early literacy assessment beyond DIBELS standard tools.* Paper presented at the annual convention of the National Association of School Psychologists, Boston, MA.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459-479.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. *Focus on Exceptional Children, 24,* 1-20.

Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook: Administration and scoring of Reading Maze for use in general outcome measurement.* Bloomington, MN: NCS Pearson. Available: https:// aimsweb.pearson.com/.

Simmons, D. C., Coyne, M. D., Kwok, O., McDonaugh, S., Harn, B., & Kame'enui, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41,* 158-173.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Speece, D.L., Schatschneider, C., Silverman, R., Case, L.P., Cooper, D.H., & Jacobs, D.M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal, 111,* 585-607.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819.

Thissen, D. (1989). Review: Latent class analysis and latent class models. *Journal of Educational Measurement, 26*(1), 98-100.

Torgesen, J. K. (2000). Individual differences in response to early intervention in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15,* 55-64.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review, 32,* 5-17.

Zumeta, R.O., Compton, D.L., & Fuchs, L.S. (2011). Using word identification fluency to monitor first-grade reading development. *Exceptional Children, 78,* 201-220.