Hammill Institute on Disabilities

Journal of Learning Disabilities
2017, Vol. 50(6) 712–723

DOI: 10.1177/0022219416664866
journaloflearningdisabilities.sagepub.com
SAGE

Predictive Validity of Curriculum-Embedded Measures on Outcomes of Kindergarteners Identified as At Risk for Reading Difficulty

Eric L. Oslund, PhD[1], Shanna Hagan-Burke, PhD[1], Deborah C. Simmons, PhD[1], Nathan H. Clemens, PhD[1], Leslie E. Simmons, MEd[1], Aaron B. Taylor, PhD[1], Oi-man Kwok, PhD[1], and Michael D. Coyne, PhD[2]

Abstract

This study examined the predictive validity of formative assessments embedded in a Tier 2 intervention curriculum for kindergarten students identified as at risk for reading difficulty. We examined when (i.e., months during the school year) measures could predict reading outcomes gathered at the end of kindergarten and whether the predictive validity of measures changed across the kindergarten year. Participants consisted of 137 kindergarten students whose reading development was assessed four times from October to February. Measures aligned with content taught in the curriculum and assessed a range of phonologic, alphabetic, and word-reading skills. Results from structural equation modeling indicate that 36.3% to 65.2% of the variance was explained on the latent decoding outcome and 62.0% to 86.8% on the latent phonological outcome across the four time points. Furthermore, the predictive validity of specific skills increased over the kindergarten year, with more complicated tasks (e.g., word segmentation) becoming more predictive at subsequent measurement occasions. Results suggest that curriculum-embedded measures may be viable tools for assessing and predicting reading performance.

Keywords

assessment, predictive validity, early reading, kindergarten

Progress monitoring is a fundamental aspect of responsive intervention, involving the frequent assessment of skills to determine whether a student is responding to instruction. To evaluate students' progress, educators need measures that accurately gauge whether student performance is adequate for reaching meaningful end-of-year goals. This is especially true in the early years. For example, Torgesen (1998) emphasized the importance of early intervention and prevention models to prevent early reading skill deficits from becoming intractable difficulties, thus avoiding the "wait to fail" scenario and its subsequent negative impacts. Progress monitoring allows a teacher to adjust teaching and curriculum using data-driven decisions to improve student outcomes.

Progress monitoring has been the subject of considerable research over the past three decades, particularly with respect to a general outcomes/curriculum-based measurement (CBM) approach. CBM involves the use of brief probes to monitor growth in important academic skills. Typically, CBMs are designed to have comparable difficulty across administrations and measure student progress toward long-term general reading outcomes. They are administered at regular intervals, which allows for monitoring students' benchmark performance (i.e., level) as well as rate (i.e., slope) of improvement. To date, much of the research on CBM has focused on students in Grades 1 through 6 (Reschly, Busch, Betts, Deno, & Long, 2009; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

Comparatively, much less research has examined the technical adequacy of measures used to assess kindergarten students' progress. Several early literacy measures were

[1]Texas A&M University, College Station, USA
[2]University of Connecticut, Storrs, USA

Corresponding Author:
Eric L. Oslund, Department of Elementary and Special Education, Middle Tennessee State University, MTSU Box 69, COE 384, Murfreesboro, TN 37132.
Email: eric.oslund@mtsu.edu

developed under a CBM framework (e.g., fluency-based measures of letter-sound correspondence, phonemic segmentation, and pseudoword reading), and research has examined their reliability and validity in terms of static scores (e.g., see Goffreda & DiPerna, 2010, for a review; Burke, Hagan-Burke, Zou, & Kwok, 2009; Catts, Petscher, Schatschneider, Sittner Bridges, & Mendoza, 2009; Good et al., 2004; Johnson, Jenkins, Petscher, & Catts, 2009; O'Connor & Jenkins, 1999). However, limited attention has been paid to how well these measures function as formative assessments, and even less research has investigated measures specifically embedded within the curriculum of instruction even though they have been recommended as options for monitoring early reading (Gersten et al., 2009). Specifically, Fuchs and Vaughn (2012) noted that whereas schools have incorporated screening assessments as standard components of their response to intervention (RTI) procedures, ongoing progress monitoring has not been widely accepted. The authors identified several important factors that may require further investigation for schools to effectively integrate progress monitoring into real-world practices. These factors include the feasibility, technical adequacy, and decision-making utility of progress-monitoring measures and methods.

Nature of Reading Development and Implications for Assessment

One of the essential characteristics of early reading measures is that they reflect important reading skills (Deno, 1985). Part of the complexity of identifying valid measures for formatively informing kindergarten reading instruction resides in the multiple skills that are learned and their relatively rapid-changing importance throughout the process of learning to read. In kindergarten, children progress through a complex process of integrating phonemic, alphabetic, and orthographic skills in learning to read words. As reading develops, students can struggle with different reading components at different times throughout the kindergarten year (O'Conner, Bocian, Sanchez, & Beach, 2014). Consequently, measures to monitor early reading development involve skills that change in emphasis and importance over the kindergarten year (O'Conner & Jenkins, 1999).

Ehri and McCormick's (1998) phase theory of word-reading acquisition is instructive for understanding this process and the skills that are particularly important for early readers and for monitoring progress. Their theory posits that early in schooling, children progress from a "prealphabetic" phase, in which they demonstrate little to no understanding of the alphabetic system, to a "partial alphabetic" phase in which they begin to associate letters with sounds. Progression into the "full alphabetic" phase takes place when students utilize their understanding of letter-sound correspondence to decode words and, through repeated exposures and decoding opportunities, enter the "consolidated" phase, when they begin to add increasingly larger letter chunks to orthographic memory, thus enabling the rapid reading of words. The authors recommended that measures of these skills also change in accordance with reading progression.

Curriculum-Embedded Measures to Inform Early Reading Intervention

Using student performance to adjust instruction is central to responsive instruction. To assist educators, some intervention programs embed measures at designated points in the curriculum to evaluate whether students have learned what has been taught (e.g., identifying letter names or sounds explicitly taught in lessons). In this article, we refer to these as *curriculum-embedded measures* (CEMs). In their RTI practice guide, Gersten et al. (2009) recommended using mastery checks embedded in the curriculum to monitor the progress of students receiving Tier 2 intervention. CEMs are designed to monitor mastery of skills from a given program or intervention, and they are used to inform instructional decisions (e.g., instruction pacing, student grouping, dosage, reteaching, advancing lessons).

Unlike curriculum-independent measures (e.g., CBM) that assess mastery toward general outcomes (e.g., knowledge of all letter names), CEMs can be considered a form of curriculum-based assessment because they use content drawn from the curriculum of instruction to formatively assess students' mastery of targeted skills and provide information for instructional planning and deciding what to teach (e.g., Blankenship, 1985; Gickling & Thompson, 1985; Howell, Hosp, & Kurns, 2008). While CEMs are typically used to make proximal adjustments to instruction, it would be important to know how performance on CEMs is associated with student outcomes. Unfortunately, limited research has examined CEMs for at-risk kindergarten students. To our knowledge, only one study (Olinghouse, Lambert, & Compton, 2006) investigated and established the predictive validity of a CEM; however, this study involved students in Grades 3 to 5.

Used together, CEMs can be viable complements to CBMs in helping make educational decisions, especially because of their direct alignment with what has been taught. CEMs can be a timely indicator of whether a student is mastering specific content or skills. This is particularly important for early reading when mastery of several prereading skills (e.g., phonemic awareness, letter-sound correspondence) is critical for later success.

A potentially useful feature of CEMs is that they are dynamic in response to the skills and material taught over a period of time. Dynamic indicators are especially important because of the developmental nature of learning to read. As Fuchs and colleagues (2007) point out, static snapshots for at-risk students are not particularly accurate for future academic performance. Moreover, because reading in kindergarten involves multiple skills of increasing complexity, including multiple-skill tests and administering those at different times throughout the year may be the most useful means of detecting reading problems (McCardle, Scarborough, & Catts, 2001). Multiskill assessment batteries, however, are resource intensive and may not be practically feasible. Research is needed that identifies predictors that reflect the dynamic development of reading and are both parsimonious and valid. CEMs may be a potential solution to this issue; in addition to reflecting mastery of targeted skills, the multiskill nature of CEMs may make them especially good predictors of subsequent outcomes.

Study Purpose

In this preliminary investigation, we examined the predictive validity of CEMs. We sought to identify a parsimonious set of measures that predicted year-end outcomes and the points of the year when they were most predictive. Most research to date has examined composites, especially of phonological processing skills (Ritchey & Speece, 2006; Schatschneider, Francis, Carlson, Fletcher, & Foorman, 2004; Speece, Mills, Ritchey, & Hillman, 2003; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). In contrast, we isolated specific early predictors (e.g., identifying first sounds) and examined their relationship to reading outcomes at multiple points in time. In addition, we examined latent outcomes composed of numerous early reading indicators to provide a potentially more accurate picture of the complex processes involved in reading and buffer against the imperfection of any one measure (Gersten et al., 2005).

Finally, this study focused on students in Tier 2 intervention, which few studies have done. Extant research on kindergarten predictors with students at risk for reading difficulties is sparse. Most research has included students representing the full range of achievement, and very few have focused on the subset of students considered to be at risk for academic difficulty (Linklater, O'Conner, & Palardy, 2009; Schatschneider et al., 2004; Torgesen, 1998). Predictive studies for at-risk students are important because of the potentially erroneous predictive applications if results from studies with samples representing the full spectrum of skill levels are overgeneralized to specific populations (Badian, 1995).

Two research questions were addressed:

Research Question 1: What sets of skill-specific CEMs administered during kindergarten are most predictive of end-of-year outcomes for students receiving Tier 2 reading instruction?

Research Question 2: At what time points are skill-specific CEMs predictive of end-of-year outcomes, and how does their predictive validity change across the kindergarten year?

Method

Research Context

Data for this study were collected as part of a larger randomized controlled trial that investigated the effects of an experimental version of the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004) on reading outcomes for kindergarten students. Students were randomly assigned to an experimental or typical practice comparison condition. In this study, only students from the experimental condition were included, as they were the only ones whose progress was measured through CEMs. Students received an average of 102 lessons over 21 weeks, administered daily for 30 minutes within groups of 3 to 5 students. CEMs were administered approximately every 4 weeks, and students were regrouped according to performance to create homogeneous groups at each measurement occasion. Students either repeated or progressed in lessons on the basis of their CEM performance.

Setting and Participants

Students (N = 153) from 10 schools in Florida, 5 in Connecticut, and 2 in Texas participated in the study. School enrollments ranged from 401 to 832 students for Florida, 287 to 739 for Connecticut, and 279 to 889 for Texas. The schoolwide percentage of students eligible for free or reduced-cost lunch ranged from 63% to 92% for Florida, 70% to 82% for Connecticut, and 81% to 82% for Texas. All 17 schools received Title 1 funding. Demographic information about the students is summarized in Table 1.

Students were identified through a two-step screening process. First, schools nominated at-risk students based on school-administered assessments. In the second step, researchers administered standardized assessments of prereading skills, including Letter Naming Fluency (LNF) from the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good & Kaminski, 2002) and the Sound Matching (SM) subtest from the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Only students (N = 153) who had a raw score ≤6 on the LNF and achieved ≤37th percentile on the SM were administered the Rapid Object Naming (RON) subtest from the CTOPP and the Letter Identification subtest from the *Woodcock Reading Mastery Test–Revised/Normative Update* (WRMT-R/NU; Woodcock, 1987/1998). In addition to the cut scores for the LNF and SM, students with standard scores ≤7 or ≤80 on the RON and Letter Identification subtests, respectively, qualified for participation (see Table 1 for pretest scores). Of those who

Table 1. Demographic Variables and Pretest Means and Standard Deviations.

	Participants			
Variable	*n*	%	*M*	*SD*
Gender				
Male	63	46.0		
Female	74	54.0		
Ethnicity				
Asian	0	0		
American Indian or Alaska Native	1	0.7		
Black or African American	40	29.2		
Hispanic or Latino	50	36.5		
White	41	29.9		
Other	5	3.6		
Identified for special education	14	10.2		
English-language learner	22	16.1		
Age			5.44	0.31
Letter Identification[a]			80.65	8.19
Sound Matching[b]			19.71	10.46
Rapid Object Naming[a]			6.08	2.19
Letter Naming Fluency[c]			1.20	1.77

Note. N = 137.
[a]Standard score. [b]Percentile score. [c]Raw score.

Table 2. Curriculum-Embedded Measures Composition and Reliability Estimates.

	Curriculum-embedded measure			
Subtest	1	2	3	4
Letter Names	√	√	√	√
Letter Sounds	√	√	√	√
First Sounds in Words	√	√	√	√
Last Sounds in Words		√	√	√
First Letter Sound		√	√	√
Last Letter Sound		√		√
Whole Word Segmentation			√	√
Cronbach's alpha reliability estimate	.80	.90	.81	.84

completed the pretests, 137 completed the posttests (attrition rate = 12%). Analyses indicated no statistically significant differences on any pretest variables between the students who did and did not complete the study.

Assessment Procedures

The CEMs and reading outcome measures were individually administered by trained research team members. Data collectors participated in 8 hours of training and reached 100% accuracy in delivering and scoring all assessment measures. To ensure accuracy, all measures were double scored by two trained members of the research team. Reading outcome measures were administered within 2 weeks after completion of the intervention; for the majority of the students, this took place in May.

Predictor variables. CEMs were untimed and administered approximately every 4 weeks to monitor student mastery of previously taught content and skills, which were derived directly from the intervention curriculum (i.e., ERI; Pearson/Scott Foresman, 2004). A total of four CEM measurement points were used in the analyses and were collected around the end of October, end of December, middle of January, and middle of February. These measurement occasions corresponded with and were based on completion of a specific part of the intervention curriculum, as opposed to being administered at fixed intervals.

The first CEM assessment battery was composed of three subtests. The Letter Names subtest requires students to name the letters *m*, *p*, *f*, *c*, *t*, *s*, and *d* displayed on a page. The Letter Sounds subtest requires students to provide the letter sound for the letters used in the Letter Names subtest. The First Sounds in Words subtest requires students to provide the first sound of a word presented orally and represented by a picture.

The second CEM battery contained the same three subtests as the first CEM, as well as three additional subtests: Last Sounds in Words, First Letter Sound, and Last Letter Sound. The letters *l*, *a*, *o*, and *r* were added to the Letter Names and Letter Sounds subtests. In the Last Sounds in Words subtest, students are asked to say the last sound in a word spoken orally by the examiner and represented by a picture. In the First Letter Sound subtest, students are given letter tiles *d*, *f*, *l*, *m*, *p*, *r*, *s*, and *t*. Students are then presented with a picture with three blank squares (corresponding to the vowel-consonant-vowel sounds) directly below the image and asked to place the tile representing the first sound of the picture in the first square. In the Last Letter Sound subtest, students are given the same tiles and pictures as in the First Letter Sound subtest; however, now they must place the tile representing the last sound in the last square.

With the exception of the Last Letter Sound subtest, the third CEM battery contained the same subtests as the second CEM with one additional subtest, Whole Word Segmentation, which requires students to segment consonant-vowel-consonant words presented orally into their individual phonemes. The letters *a*, *b*, *c*, *d*, *f*, *l*, *i*, *m*, *n*, *o*, *p*, *s*, and *t* were included in the third CEM to be used with the First Letter Sound and Last Letter Sound subtests.

The fourth CEM assessment included all of the subtests from the first three CEMs. Table 2 summarizes the four CEMs used in our analyses and their estimated reliability in the measured sample based on Cronbach's alpha.

Outcome measures. Seven outcome measures were administered at the end of the intervention.

Blending Words. The Blending Words subtest from the CTOPP was administered to assess students' ability to verbally blend individual sounds into whole words. Internal reliability alpha coefficients ranged from .86 to .89 for children aged 5 through 7 years.

Sound Matching. The SM subtest from the CTOPP is an assessment of phonemic awareness in which students are presented with a target picture and three additional pictures. Students are asked to match one of the three pictures to the target picture based on the first or the last sound. Internal reliability alpha coefficients range from .92 to .93 for children aged 5 through 7 years (Wagner et al., 1999).

Phoneme Segmentation Fluency. In the Phoneme Segmentation Fluency (PSF) subtest from the DIBELS, students are asked to orally produce the individual sounds of a stimulus word presented by the examiner. Students must identify within 1 minute as many individual sounds as possible from words with three to four phonemes. The alternate form reliability of forms given 2 weeks apart for the PSF is .88 (Good et al., 2004).

Nonsense Word Fluency. Nonsense Word Fluency (NWF) from the DIBELS is a measure of letter-sound correspondence and basic decoding skills. From a list of vowel-consonant and consonant-vowel-consonant nonwords, students may name individual letter sounds or read the words as whole units. Students are to name within 1 minute as many letter sounds or words as possible. One-month alternate form reliability is .83 (Good et al., 2004).

Oral reading fluency. Oral reading fluency (ORF) was measured with the "Mac Gets Well" passage (Makar, 1995), which consists of a high percentage of decodable vowel-consonant and consonant-vowel-consonant words. Students are to correctly read within 1 minute as many words as possible in the passage. Internal reliability coefficients reported by Vadasy, Sanders, and Peyton (2008) are .93 for their kindergarten sample.

Word Identification. Word Identification (WI) from the WRMT-R/NU measures students' skills in reading words of increasing difficulty. An item is scored correct if the student is able to read the word with the correct pronunciation. The median split-half reliability coefficient is .97 (Woodcock, 1987/1998).

Word Attack. The Word Attack from the WRMT-R/NU follows the same procedures as the WI but uses pseudowords of increasing difficulty instead of real words. The use of pseudowords requires students to rely on decoding skills to effectively read the words. Median split-half reliability is .87 (Woodcock, 1987/1998).

Data Analyses

Data were analyzed with Mplus 6.12 and SPSS 20. The "maximum likelihood with robust standard errors" estimator was used for the structural equation modeling analyses; this adjusts standard errors by taking into account nonindependent data and uses all available data for estimation. The adjustment of standard errors reduces Type I error, which reduces the likelihood of spurious statistically significant findings. Students were nested within interventionist; thus, the "TYPE=COMPLEX" analysis was used with interventionist as the cluster variable. A confirmatory factor analysis (CFA) was used to validate a two-factor measurement model that summarized kindergarten reading outcomes (described later). Following the CFA analysis, a structural model predicting outcomes from each measurement time point was constructed, resulting in four models. For each model, the pretest RON was entered as a covariate, and a demographic covariate had three dummy-coded variables (i.e., Hispanic, African American, and other ethnicity), with Caucasian students as the reference group.

Results

Descriptive statistics for all CEMs and outcome variables are presented in Table 3, along with correlations between each CEM and reading outcome measure. As illustrated, the CEMs at each administration point demonstrated moderate correlations (M = .45) with the outcome variables, and all were statistically significant.

For our model summarizing kindergarten reading outcomes, the CFA confirmed a two-factor solution with a statistically nonsignificant overall chi-square value: $\chi^2(12) = 9.91$, $p = .624$. Good fit was observed for the estimated measurement model (root mean square error of approximation [RMSEA] = .00, comparative fit index [CFI] = 1.0, standardized root mean square residual [SRMR] = .02). In this model, the phonological awareness latent factor (hereafter, "phonological") was composed of the three phonological awareness measures (Blending Words, SM, PSF), and the decoding latent factor was composed of the four decoding-related outcome variables (NWF, WI, Word Attack, ORF). The residuals for the two fluency measures (NWF, ORF) loading on the decoding outcome were correlated. All loadings on the phonological and decoding latent factors were statistically significant ($p < .01$), and the measured variables were positively related to the latent factors on which they loaded. The decoding and phonological factors were also positively related. The variance explained (R^2) on the measured outcome variables ranged from 41% to 86%.

Table 3. Means, Standard Deviations, and Correlations Between CEMs and Kindergarten Outcomes Measures.

			Spring outcome measures						
Measurement point: CEM predictor	*M*	*SD*	BW	SM	PSF	NWF	ORF	WI	WA
1									
Letter Names	5.52	1.81	.47	.52	.37	.37	.36	.53	.45
Letter Sounds	5.62	1.63	.43	.50	.36	.40	.39	.48	.48
First Sounds in Words	4.65	2.44	.48	.54	.48	.24	.31	.36	.44
2									
Letter Names	10.07	1.81	.45	.45	.47	.34	.33	.54	.40
Letter Sounds	9.85	2.01	.49	.43	.52	.34	.34	.54	.43
First Sounds in Words	8.37	3.23	.53	.52	.63	.29	.42	.53	.55
Last Sound in Words	5.96	3.70	.47	.46	.56	.29	.37	.43	.47
First Letter Sound	8.39	2.85	.58	.57	.64	.39	.44	.65	.61
Last Letter Sound	7.12	3.53	.51	.55	.61	.39	.41	.55	.52
3									
Letter Names	2.49	0.65	.22	.31	.23	.25	.26	.38	.32
Letter Sounds	2.33	0.80	.37	.45	.41	.39	.41	.65	.45
First Sounds in Words	2.14	1.04	.52	.42	.65	.32	.37	.51	.48
Last Sound in Words	1.88	1.14	.31	.45	.53	.28	.34	.38	.34
First Letter Sound	4.10	1.54	.51	.48	.64	.34	.41	.53	.59
Whole Word Segmentation	6.25	2.47	.50	.54	.69	.29	.40	.43	.51
4									
Letter Names	4.28	0.91	.41	.44	.35	.46	.41	.55	.47
Letter Sounds	4.16	1.03	.35	.40	.37	.47	.43	.59	.49
First Sounds in Words	3.92	1.46	.54	.53	.69	.27	.38	.49	.51
Last Sound in Words	3.47	1.77	.37	.44	.47	.30	.28	.41	.45
First Letter Sound	4.25	1.23	.54	.54	.65	.31	.38	.50	.53
Last Letter Sound	3.85	1.53	.45	.54	.60	.38	.39	.45	.56
Whole Word Segmentation	10.58	3.65	.52	.54	.76	.36	.40	.52	.59
M			11.04[a]	9.27[a]	41.03[b]	29.34[b]	11.32[b]	106.39[a]	107.56[a]
SD			2.53	2.36	20.59	15.46	9.42	12.42	10.86

Note. CEM = curriculum-embedded measure; BW = Blending Words; SM = Sound Matching; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = oral reading fluency; WI = Word Identification; WA = Word Attack.
[a]Standard score. [b]Raw score.

October Measurement

This first structural model examined the validity of the individual subtests from the first CEM (Letter Names, Letter Sounds, First Sounds in Words) administered in October, predicting year-end reading skills. The hypothesized model (see Figure 1) includes the standardized coefficients. The chi-square test of model fit was not statistically significant, $\chi^2(47) = 63.45$, $p = .055$, and the model fit the data well (RMSEA = .05, CFI = .98, SRMR = .04). The Letter Names subtest was a statistically significant predictor of both the phonological ($\gamma = 0.52$, $p = .002$) and the decoding ($\gamma = .35$, $p = .037$) factors. The First Sounds in Words subtest was a statistically significant predictor on the phonological factor ($\gamma = .50$, $p < .000$). The Letter Sounds subtest did not reach statistical significance for the phonological or decoding factor. The predictors explained a statistically significant amount of variance on both factors, with 62% of the variance explained on the phonological factor and 36% explained on the decoding factor.

December Measurement

In the second model (see Figure 2), the predictive validity of the six CEM subtests administered in December was examined. The overall chi-square value was not statistically significant, $\chi^2(62) = 69.90$, $p = .229$, and model fit indices indicated good fit (RMSEA = .03, CFI = .99, SRMR = .03). The First Sounds in Words ($\gamma = .21$, $p = .041$), Last Sounds in Words ($\gamma = .21$, $p = .009$), and First Letter Sound ($\gamma = .41$, $p = .003$) were statistically significant predictors of the phonological factor. The First Letter Sound subtest was also a statistically significant predictor on the decoding factor ($\gamma = .54$, $p = .001$). The remaining subtests were not statistically significant predictors on either outcome factor. The total amount of variance explained on the phonological factor was 79% and on the decoding outcome, 55%.

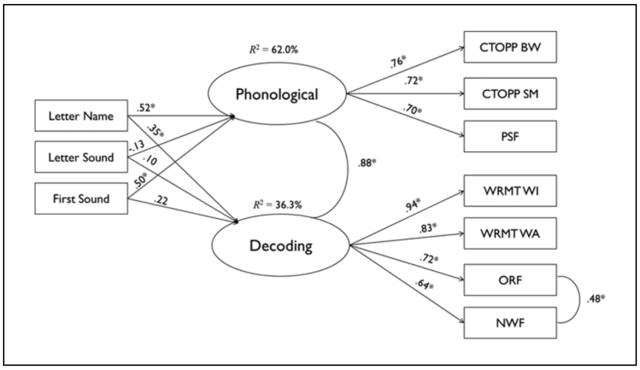

Figure 1. Time 1 model for predictors gathered in October. CTOPP = Comprehensive Test of Phonological Processing; BW = Blending Words; SM = Sound Matching; PSF = Phoneme Segmentation Fluency; WRMT = *Woodcock Reading Mastery Test–Revised/Normative Update*; WI = Word Identification; WA = Word Attack; ORF = oral reading fluency; NWF = Nonsense Word Fluency. *p < .05.

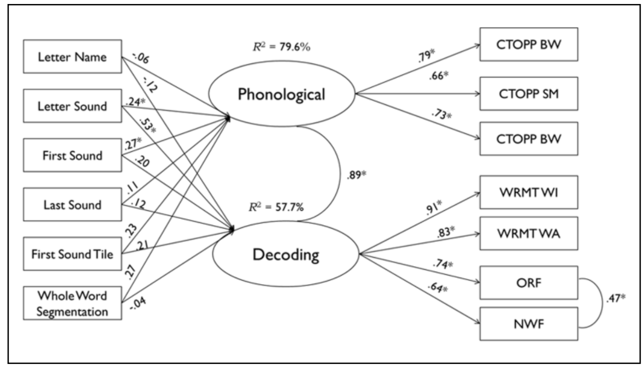

Figure 3. Time 3 model for predictors gathered in January. CTOPP = Comprehensive Test of Phonological Processing; BW = Blending Words; SM = Sound Matching; WRMT = *Woodcock Reading Mastery Test–Revised/Normative Update*; WI = Word Identification; WA = Word Attack; ORF = oral reading fluency; NWF = Nonsense Word Fluency. *p < .05.

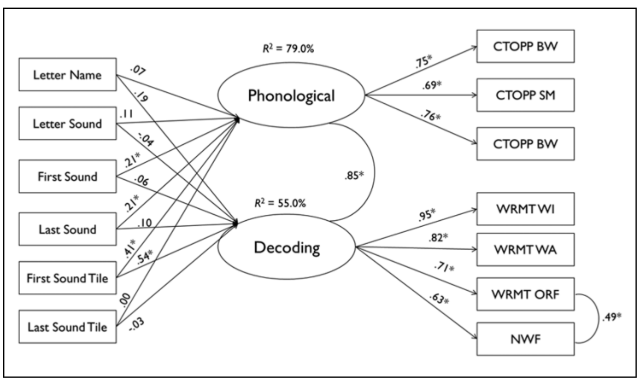

Figure 2. Time 2 model for predictors gathered in December. CTOPP = Comprehensive Test of Phonological Processing; BW = Blending Words; SM = Sound Matching; PSF = Phoneme Segmentation Fluency; WRMT = *Woodcock Reading Mastery Test–Revised/Normative Update*; WI = Word Identification; WA = Word Attack; ORF = oral reading fluency; NWF = Nonsense Word Fluency. *p < .05.

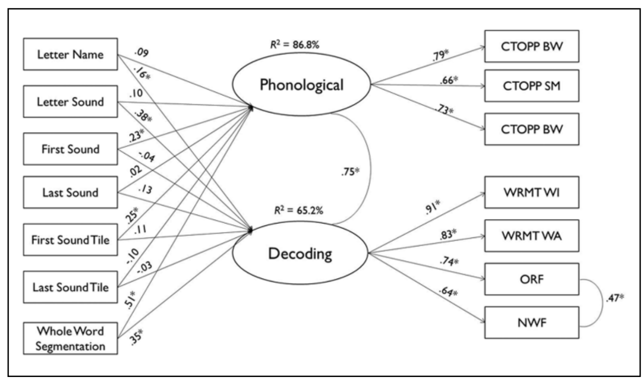

Figure 4. Time 4 model for predictors gathered in February. CTOPP = Comprehensive Test of Phonological Processing; BW = Blending Words; SM = Sound Matching; WRMT = *Woodcock Reading Mastery Test–Revised/Normative Update*; WI = Word Identification; WA = Word Attack; ORF = oral reading fluency; NWF = Nonsense Word Fluency. *p < .05.

January Measurement

The third model (see Figure 3) investigated the predictive validity of the subtests from the January CEM administration. The model fit the data reasonably well based on fit indices, although the overall chi-square value was statistically significant, $\chi^2(62) = 87.26$, $p = .02$ (RMSEA = .06, CFI = .97, and SRMR = .04). The Letter Sounds subtest was a statistically significant predictor for both the phonological ($\gamma = .24$, $p = .045$) and the decoding ($\gamma = .52$, $p < .000$) factors. First Sounds in Words was a statistically significant predictor for the phonological factor ($\gamma = .27$, $p = .002$) and nearly significant for the decoding factor ($\gamma = .20$, $p = .066$). The other CEM subtests were not statistically significant for either factor in this model. The variance explained on the phonological and decoding factors was 80% and 58%, respectively.

February Measurement

The fourth model (see Figure 4) included all the subtests from the previous three CEMs. The chi-square value was statistically significant, $\chi^2(67) = 102.06$, $p = .004$. The fit indices indicated that the model fit the data adequately (RMSEA = .06, CFI = .96, SRMR = .04). Whole Word Segmentation was a statistically significant predictor for

Table 4. Statistically Significant Predictors by Measurement Occasion.

Measure	Measurement 1[a]		Measurement 2[b]		Measurement 3[c]		Measurement 4[d]	
	Phono	Decoding	Phono	Decoding	Phono	Decoding	Phono	Decoding
Letter Names	✓	✓						✓
Letter Sounds					✓	✓		✓
First Sounds in Words	✓		✓		✓		✓	
Last Sounds in Words	—	—	✓					
First Letter Sound	—	—	✓	✓			✓	
Last Letter Sound	—	—			—	—		
Whole Word Segmentation	—	—	—	—			✓	✓

Note. Checkmarks indicate statistically significant predictors for a given measurement period, whereas dashes signify that a measure was not administered during that measurement period. Phono = phonological.
[a]End of October. [b]End of December. [c]Mid-January. [d]Mid-February.

both the phonological ($\gamma = .51$, $p < .000$) and the decoding ($\gamma = .35$, $p < .000$) factors. First Sounds in Words ($\gamma = .23$, $p = .026$) and First Letter Sound ($\gamma = .25$, $p = .014$) were also statistically significant predictors of the phonological factor, whereas Letter Names ($\gamma = .38$, $p < .000$) was a statistically significant predictor for the decoding factor. No other subtests were statistically significant predictors for the phonological and decoding factors alongside the other predictors. The total variance explained was 87% on the phonological factor and 65% on the decoding factor. Table 4 summarizes the CEMs that were administered at the four measurement points, indicating which were predictive at each time point and illustrating how the predictive validity changed across the kindergarten year.

Discussion

This study investigated the predictive validity of CEMs on end-of-kindergarten outcomes within a sample of kindergarten students who were considered to be at risk for reading difficulties and were receiving a systematic Tier 2 reading intervention. The CEMs were derived from a commercial early reading intervention and designed to assess student RTI by measuring mastery of targeted skills taught during designated periods. Measures were administered approximately every 4 weeks, and tasks changed across measurement points to reflect the developmental progression of skills in the reading intervention. We attempted to fill gaps in the research base regarding students in Tier 2 intervention, specifically investigating whether we could identify a parsimonious set of formative measures that were valid and reflected the dynamic development of reading over the kindergarten year. We sought to examine if, in addition to informing instruction, CEMs could predict future reading outcomes.

The results of this study corroborated the relationships of some previously validated skills and extended our understanding of methodologies that may inform the way to measure students' RTI. Findings also indicated that a dynamic set of skills predicted reading outcomes over the four measurement points. In late October, knowledge of letter names, as measured by the CEM Letter Names subtest, was a significant predictor for the phonological awareness and decoding outcomes, making it a parsimonious indicator of multiple outcomes. This finding is consistent with the extensive research that supports letter identification skills as one the most powerful predictors of future literacy skills (Foulin, 2005; Schatschneider et al., 2004). Also predictive at the first time point for the phonological factor was knowledge of the first sound in a word presented orally. The combined power of the three predictors explained 62% of the variance on the phonological outcome and 36% on the decoding outcome. This indicates that the CEMs collected in October provide information that can validly predict student performance at the end of kindergarten.

At the December measurement point, a majority of variance was explained on both outcomes (79% on phonological, 55% on decoding), and a parsimonious set of three predictors emerged. The combined phonemic-alphabetic task—which required a student to isolate the first sound of a word presented orally and select the corresponding letter tile—predicted both outcome factors. This finding may be indicative of students' progress into the partial-alphabetic phase, consistent with Ehri and McCormick's (1998) phase theory, in which students build knowledge of letter-sound correspondences and begin to apply them to print-related activities, thus underscoring the importance of tasks that reflect the progression of early reading skills. The phonemic segmentation task requiring the identification of the first sound of a word presented orally was statistically significant at this measurement point for the phonological outcome, as was the task requiring identification of the last sound. This finding is consistent with that of Linklater et al. (2009), who found that their untimed measure of phoneme segmentation administered in fall was a strong predictor of end-of-kindergarten reading outcomes. As they suggested, their measure outperformed a phoneme segmentation CBM because it was responsive to the instruction that the students received. As in October, the amount of variance explained on both outcomes may give educators confidence that measures provide a valid index of future reading performance and can yield valuable information for making educational decisions.

In January, at the third measurement point, the task that predicted both phonological and decoding outcomes was the Letter Sounds subtest, which measured students' ability to associate a letter with its sound. The ability to associate a letter with its corresponding sound is essential to word reading and related to the partial-alphabetic stage of Ehri and McCormick's phase theory (1998). Furthermore, the ability to isolate the first sounds of words presented orally predicted the phonological outcome and approached significance (p = .066) for the decoding outcome. As with the December measurement, nearly 80% of the variance could be explained on the phonological factor, whereas 58% was explained on the decoding factor.

Whole Word Segmentation was not a significant predictor at the January measurement point, which was unexpected given the validity evidence demonstrated for other phonemic segmentation measures (e.g., Goffreda & DiPerna, 2010). However, this finding may support the notion that the validity of certain predictors may vary per the skill level of the sample. Although students' ability to isolate the first sound of a word presented orally was a significant predictor at each measurement occasion, the Whole Word Segmentation task may have been too difficult at this point of the year. Additionally, performance was highly variable across students, which caused inflated standard errors. Even though the beta weights were relatively large, the imprecision of the estimate, as indicated by high variance, diminished the chance of statistical significance.

On the January measurement, some scores (e.g., Letter Names, First Sounds in Words) were lower than they were in December and February. One possible explanation for this observed decline is that analogous to the "summer slide," where children can regress on reading skills if not actively engaged in reading during academic breaks. The participants were in a daily, intense, and explicit intervention, and an absence of this intervention could explain the regression in scores. Unfortunately, we do not have data on home reading practices over the winter break that may help us better understand the cause or correlates of lower performance following a lack of intervention. Perhaps the age of the participants, their stage in reading development, and the intensity of the intervention increased the "winter slide" as compared with what would be seen among nonstruggling readers.

The final measures that were administered in February explained 87% of the variance on the phonological outcome and 65% on the decoding outcome. With respect to parsimony, the whole-word segmentation task predicted both phonological and decoding outcomes. This finding follows the general pattern across all four models indicating that, concurrent with reading development, tasks that are more complex and require greater integration of reading-related subskills become stronger predictors over time, which is consistent with Ehri and McCormick's phase theory (1998). Furthermore, the pattern held that predictive validity of CEMs strengthened over time, as indicated by greater amounts of variance explained—a factor that may be due to more complex and "reading-like" behaviors that are part of the assessment at later points in time, as well as to closer proximity of the predictor and outcome assessments.

On several occasions, certain CEM predictors, such as letter naming and letter sound identification, were not significant predictors of reading outcomes when included in models with other similar CEM subtests. Their failure to reach statistical significance is likely not due to their lack of relevance as predictors of later reading skills. Rather, it is most likely the result of multicollinearity and the presence of a higher-order skill or task (e.g., First Letter Sound) that subsumed what variance would have been explained individually by the more basic measures (e.g., Letter Names).

Summary and Implications

In this study, we were interested in determining whether a parsimonious set of tasks from a larger set of CEMs could validly predict multivariate latent end-of-kindergarten outcomes for students in Tier 2 intervention and whether those skills changed over time. Few studies have evaluated the validity of such measures specifically with samples identified as at risk for reading difficulties. This study demonstrated results consistent with prior research conducted with students whose performance represented a broad spectrum of reading skills. Overall, findings indicated that a measurement set that prioritizes different tasks over the course of the year (see Table 4) may be important for educators who use formative assessment to inform instructional decisions.

Prior research in the early grades has largely focused on the full range of learners, and findings from this study corroborate that some previously validated measures, such as letter names, letter sounds, and phonological awareness tasks, continue to predict end-of-kindergarten outcomes. In this study, measures of letter sound correspondence were predictive of year-end reading skills across several CEM assessments, which is consistent with other findings showing that letter-sound fluency growth across kindergarten is predictive of year-end reading outcomes (Ritchey & Speece, 2006), as well as recommendations for using measures of letter sounds for monitoring kindergarten reading progress (Fuchs & Fuchs, 2004). Of importance, however, was that measures reflecting greater complexity in terms of the integration of phonological and alphabetic skills (e.g., isolating first or last sounds and matching with a letter tile) or unit of analysis (i.e., whole word versus first sound segmentation) were more strongly associated with reading outcomes over the course of the intervention.

The present findings have practical implications for the frequent assessment of student progress. CEMs are inherently formative assessments given their periodic placement within a curriculum, but the information they provide is often specific to the skills targeted and the ways in which they are interpreted and the decision-making processes using them may look different from common conceptualizations of "progress monitoring." Rate of growth (or lack thereof) relative to a target goal has always been a focal point of data-based decision making with progress monitoring, and recommendations for instructional decision making with progress-monitoring data have long emphasized the visual analysis of graphed data on a time-series basis (Deno, 1985; Stecker, Fuchs, & Fuchs, 2005). Using rate of growth as a decision-making variable requires measures in which the content and difficulty levels are held constant over time, which differs from the changing complexity and difficulty of skills assessed within CEMs.

CEMs may be particularly relevant to informing early reading instruction. Whereas measures of oral reading or text comprehension are appropriate for summarizing general outcomes for students at more advanced stages of reading development, assessing early reading acquisition is challenging given the multiple reading-related subskills (e.g., phonemic awareness, letter knowledge, decoding skills) that represent early literacy. Emergence of these subskills may be closely tied to the sequence at which they are targeted in early reading instruction. Therefore, CEMs that assess the specific skills or specific content (e.g., phoneme blending, individual letter sounds, reading consonant-vowel-consonant words) targeted in the curriculum may be particularly informative for measuring responsiveness to early literacy instruction. Data from the CEMs can then be used to identify when students are not acquiring these skills at the expected rate, thus signaling the need for additional instruction or practice opportunities.

In short, CEMs may be sensitive to students' skill acquisition within a curriculum. Although more work is needed to determine if progress monitoring with CEMs improves monitoring with tools developed under a general outcomes approach, we posit that periodic monitoring with both types of measures may be ideal. That is, CEMs may provide sensitive data on students' discrete skill acquisition, thereby directly informing when review or reteaching is needed and what specific curricular content should be targeted. Periodic assessment with traditional CBM measures, however, may best inform student growth toward important annual objectives and can provide information on the degree to which students are generalizing targeted subskills to overall reading achievement. The lack thereof might inform the need of greater programming for generalization, gaps in students' understanding, or additional instruction to bring skills to mastery. In other words, CBMs might be thought to provide a "wide-angle lens" on students' academic skill development, whereas CEMs can be viewed as a microscope to evaluate improvements in subskills that form the foundation of overall achievement.

Limitations and Future Research Directions

There are several limitations to this study and several avenues for future study. First, the study involved a "medium" sample size (Kline, 2005), which limits power. Second, the results are limited to CEMs from the ERI program (Pearson/Scott Foresman, 2004). Future research should explore CEMs from other interventions and curricula to examine their viability as predictors of early reading outcomes. Studies should also investigate whether teacher-designed mastery checks have similar predictive validity, as not all schools use standardized interventions or curricula. The last measurement point used in this study was from February; research is needed to evaluate the predictive validity of later measurements, particularly if the predictive validity continues to change in the spring. Also, we recognize that students who attain adequate scores are typically transitioned to Tier 1; however, due to the experimental nature of this study, students were retained throughout the entire intervention. Finally, the CEMs in this study were not timed; therefore, it may take longer to administer them than fluency-based measures. The use of untimed CEMs was based on the need to obtain detailed information on students' mastery of targeted skills; examination of fluency-based CEMs is needed to evaluate whether additional benefits to school resources (e.g., teacher time) can be achieved and whether the fluency component may add important information about students' level of skill acquisition and overall proficiency.

Implications and Conclusion

Results of this study provide preliminary evidence supporting the predictive validity of CEMs for kindergarten students at risk for reading difficulties who are receiving tiered intervention. Early intervention is critical, and having reliable and predictive measures allows teachers to make data-informed decisions that potentially lead to better student outcomes. In particular, knowing the changing nature of individual predictors can inform teachers regarding when and which individual skills are most important to monitor. This, in turn, may enable researchers and teachers to develop a more parsimonious set of measures that changes throughout the year according to reading development. Such increased parsimony could reduce demands on often-limited school resources.

Successful RTI models rely on using available data to make instructional adjustments. Having viable CEMs can make teachers more confident about their decisions and potentially improve reading outcomes for kindergarten students struggling with reading. Reliable and valid early reading measures that balance maximum predictive power with feasibility of use will strengthen RTI models and help allocate resources to those who need them most.

Authors' Notes

Eric L. Oslund is now at the Department of Elementary and Special Education, Middle Tennessee State University, Murfreesboro, Tennessee.

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: Preparation of this article was supported in part by Project Early Reading Intervention (Grant R324E060067A). This material does not necessarily represent the policy of the U.S. Department of Education, and neither is it necessarily endorsed by the federal government.

References

Badian, N. A. (1995). Predicting reading ability over the long term: The changing roles of letter naming, phonological awareness and orthographic processing. *Annals of Dyslexia*, *45*, 79–96. doi:10.1007/BF02648213

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children*, *52*, 233–238.

Burke, M. D., Hagan-Burke, S., Zou, Y., & Kwok, O. (2010). A structural equation model using sublexical fluency measures to predict emerging reading ability. *Remedial and Special Education*, *31*, 385–399.

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*, 163–176. doi:10.1177/0022219408326219

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

Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly*, *14*, 135–163. doi:10.1080/1057356980140202

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, *18*, 129–155.

Fuchs, L. S., & Fuchs, D. (2004). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. *Assessment for Effective Intervention*, *29*(4), 25–37.

Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness to intervention: A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children*, *39*(5), 58–63.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*, 195–203. doi:10.1177/0022219412442150

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M. D., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149–164.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades. A practice guide* (NCEE Publication No. 2009-4045). Washington, DC: U.S. Department of Education.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children*, *52*, 205–218.

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.

Good, R. H., III, & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., III, Kaminski, R. A., Shinn, M. R., Bratten, J., Shinn, M., Laimon, D., . . .Flindt, N. (2004). *Technical adequacy of DIBELS: Results of the early childhood research institute on measuring growth and development* (Technical Report No. 7). Eugene: University of Oregon.

Howell, K. W., Hosp, J. L., & Kurns, S. (2008). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 349–362). Bethesda, MD: National Association of School Psychologists.

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. doi:10.1111/j.1540-5826.2009.00291.x

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Linklater, D. L., O'Connor, R. E., & Palardy, G. J. (2009). Kindergarten literacy assessment of English only and English language learner students: An examination of the predictive validity of three phonemic awareness measures. *Journal of School Psychology*, *47*, 369–394. doi:10.1016/j.jsp.2009.08.001

Makar, B. M. (1995). *Primary phonics*. Cambridge, MA: Educators Publishing Service.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research and Practice*, *16*, 230–239. doi:10.1111/0938-8982.00023

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*, 159–197.

O'Connor, R. E., Bocian, K. M., Sanchez, V., & Beach, K. D. (2014). Access to a responsiveness to intervention model: Does beginning intervention in kindergarten matter? *Journal of Learning Disabilities*, *47*, 307–328. doi:10.1177/002221 9412459354

Olinghouse, N. G., Lambert, W., & Compton, D. L. (2006). Monitoring children with reading disabilities' response to phonics intervention: Are there differences between intervention aligned and general skill progress monitoring assessments? *Exceptional Children*, *73*, 90–106.

Pearson/Scott Foresman. (2004). *Scott Foresman sidewalks: Early Reading Intervention*. Glenview, IL: Author.

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. doi:10.1016/j.jsp.2009.07.001

Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology*, *31*, 301–327. doi:10.1016/j.cedpsych.2005.10.001

Schatschneider, C., Francis, D. J., Carlson, C. D., Fletcher, J. M., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282. doi:10.1037/0022-0663.96.2.265

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education*, *36*, 223–233. doi:10.1037//0022-0663.91.3.439

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. doi:10.1002/pits.20113

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator*, *22*, 32–39.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading and Writing: An Interdisciplinary Journal*, *21*, 929–963. doi:10.1007/s11145-008-9119-9

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal*, *21*, 437–480. doi:10.1007/s11145-007-9098-2

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*, 85–120.

Woodcock, R. W. (1987/1998). *Woodcock Reading Mastery Test–Revised/Normative Update*. Bloomington, MN: Pearson Assessments.