Hammill Institute on Disabilities

Learning Disability Quarterly
2015, Vol. 38(1) 3–14

DOI: 10.1177/0731948714524752
ldq.sagepub.com
SAGE

Can Curriculum-Embedded Measures Predict the Later Reading Achievement of Kindergarteners at Risk of Reading Disability?

Eric L. Oslund, PhD[1,2], Deborah C. Simmons, PhD[2], Shanna Hagan-Burke, PhD[2], Oi-Man Kwok, PhD[2], Leslie E. Simmons, MEd[2], Aaron B. Taylor, PhD[2], and Michael D. Coyne, PhD[3]

Abstract

This study examined the changing role and longitudinal predictive validity of curriculum-embedded progress-monitoring measures (CEMs) for kindergarten students receiving Tier 2 intervention and identified as at risk of developing reading difficulties. Multiple measures were examined to determine whether they could predict comprehensive latent first- and second-grade reading outcomes and whether their predictive validity changed concurrent with reading development. CEMs of phonemic, alphabetic, and integrated tasks were given 3 times during the kindergarten year to 299 students. Structural equation modeling indicates that CEMs explained a significant amount of variance on first- (54%–63%) and second-grade (34%–41%) outcomes. The predictive validity of specific measures varied over the kindergarten year with sound and letter identification measures being predictive early and segmenting and word reading becoming important as reading abilities progressed. Findings suggest that CEMs may be viable and helpful tools for making data-driven instructional decisions in a response to intervention framework.

Keywords

CEM, kindergarten, reading difficulties, RTI, data-driven instruction

In Response to Intervention (RTI) models, educators use formative assessments to determine whether an intervention is effective for individual students (Lembke, McMaster, & Stecker, 2010). By design, these assessments need technical adequacy, including predictive validity, to provide information to guide instructional decisions. Measuring student performance to inform instruction is typically accomplished through progress monitoring that involves collecting data about student progress throughout the year. Prior research has demonstrated that the information provided by progress monitoring can be used by teachers to make modifications to their instruction, which leads to improved student outcomes (Stecker, Fuchs, & Fuchs, 2005). In this study, we investigated the predictive validity of curriculum-embedded measures (CEMs) of skills that have been taught during intervention for kindergarten students identified to be at significant risk of developing reading disabilities or difficulties.

In the majority of RTI models, student performance is evaluated using curriculum-based measures (CBMs) that assess progress toward long-term outcomes that students should be able to accomplish after an extended period of instruction (Fuchs, Fuchs, & Compton, 2004). To evaluate progress, CBMs are of comparable difficulty and are administered at regular intervals. Performance is measured in terms of level and slope, and students whose performance falls significantly below identified benchmark levels (e.g., 40 correct words per minute) and who fail to make adequate growth (e.g., 2 words per week) are identified for additional intervention (Fuchs et al., 2004).

As a complement to CBMs, intervention curricula often have embedded measures that focus on short-term skill attainment and assess student mastery of skills that have been explicitly taught (Gersten et al., 2009). Performance is evaluated toward a criterion of performance (e.g., 80%

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

Note: Eric Oslund is CLD's 2013 Outstanding Researcher Award recipient.

Corresponding Author:
Eric L. Oslund, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, College of Education SZB 228, 1912 Speedway, D4900, Austin, TX 78712-1284, USA.
Email: ericoslund@yahoo.com

correct) and students whose performance falls below a designated criterion are identified for further instruction on specific skills. Because the targeted skills and curriculum material taught change over the kindergarten year, CEMs typically are nonequivalent and skills change in accordance with the curriculum. While CBMs serve an important role in progress monitoring in an RTI framework, CEMs may serve a valuable instructional complement. CBMs inform teachers on whether students are making adequate progress toward long-term outcomes, and CEMs provide details on student mastery of recently taught skills. We consider that both serve important functions in determining students' response to intervention.

The *Assisting Students Struggling With Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades* practice guide (RTI Practice Guide) issued by the Institute of Educational Sciences (IES) recommended the use of CEMs to monitor student progress and inform instructional decisions (Gersten et al., 2009). However, few progress-monitoring studies have examined the predictive validity and reliability of CEMs, particularly in kindergarten where reading development for the majority of children occurs rapidly and a single progress-monitoring measure with sufficient predictive validity toward comprehensive reading outcomes is elusive (Bishop, 2003).

Kindergarten is an especially important time for reading intervention. Prior research has established the significance of early intervention to diminish ongoing reading risk and reduce negative, long-term academic outcomes for at-risk students (Al Otaiba & Torgesen, 2007; Gersten & Dimino, 2006). We were able to identify only one study that focused on children receiving Tier 2 intervention that assessed longitudinal reading outcomes (Pennington & Lefly, 2001). An important function of a progress-monitoring measure is its ability to validly predict later reading outcomes. The ability to predict future reading outcomes from measures administered in kindergarten, particularly early enough in the kindergarten year to inform instruction, is critical.

Identifying valid measures that can be used throughout the kindergarten year to predict future reading outcomes has important implications for instructional decisions. Most longitudinal studies of kindergarten reading predictors focused on measures collected at the end of the kindergarten year, which is too late to inform prevention and early intervention efforts in kindergarten. The resulting loss of valuable time may delay early intervention opportunities, placing a student at further risk of reading difficulties and subsequent placement in special education as a student with a learning disability. Clearly, establishing early indicators of future reading outcomes is a critically important feature of an RTI approach.

CEMs that can predict important comprehensive reading outcomes may be an important tool for teachers to use to improve student outcomes. The body of work on monitoring early reading progress highlights a need to establish the changing role of predictors over time, especially during the kindergarten year. Ehri and McCormick's (1998) word learning model suggests that word learning is developmental with easier initial skills (e.g., letter identification) progressing to more difficult skills (e.g., decoding and word reading). This has implications not only for the timing of skills being taught but also for the measurement of those skills. As word reading develops, so should the measures of pre-reading skills in alignment with instruction (Ehri & McCormick, 1998). The purpose of this study was to evaluate the predictive validity of CEMs administered throughout kindergarten to children receiving Tier 2 reading intervention and investigate their dynamic utility for predicting first- and second-grade reading outcomes. We sought to evaluate which skill-specific CEMs, and when, are valid predictors of longitudinal outcomes. Following, we summarize relevant research on longitudinal predictors.

Extant Research Review of Kindergarten Progress-Monitoring Predictors

To identify skills that can inform CEM selection, we first examined prior research that investigated kindergarten predictors of first- and second-grade reading outcomes and have identified promising measures that predicted later reading outcomes. Multiple studies documented that phonological awareness (e.g., first sounds, blending, or segmenting) was predictive of several reading outcomes (i.e., passage comprehension, word reading, and word identification) at the end of first and second grade (Chiappe, Siegel, & Wade-Woolley, 2002; Kirby, Parrila, & Pfeiffer, 2003; Morris, Bloodgood, & Perney, 2003; Schatschneider, Francis, Carlson, Fletcher, & Foorman, 2004). The challenge regarding phonological awareness measures is the number and variety of measures used. For example, there were as few as 1 task and as many as 17 that varied considerably (e.g., phoneme segmentation, phoneme deletion, phoneme identification, etc.). In addition to phonological awareness, kindergarteners' knowledge of letter names was a consistent predictor of longitudinal reading outcomes (Chiappe et al., 2002; Morris, Bloodgood, & Perney, 2003; Schatschneider et al., 2004), as was rapid naming speed (e.g., rapidly naming colors, objects, or letters; Kirby et al., 2003; Schatschneider et al., 2004). Knowledge of letter sounds (Schatschneider et al., 2004) and word concepts (i.e., the ability to identify a target word in a sentence; Morris, Bloodgood, & Perney, 2003) also predicted outcomes gathered in first and second grade. It is important to note that the studies mentioned here examined the predictive ability of measures gathered at one time point when

predicting later outcomes versus the dynamic role of predictors (i.e., how predictors changed over time).

Some studies have focused on the developmental nature of early reading predictors of longitudinal reading outcomes and have examined predictive validity across a number of measurement points in kindergarten. Morris, Bloodgood, Lomax, and Perney (2003) examined the changing role of reading progress measures administered throughout the kindergarten year for predicting first-grade outcomes. They used structural equation modeling and validated a developmental sequence of reading. Their results indicated kindergarteners' alphabet knowledge preceded beginning consonant awareness (gathered in September of kindergarten), followed by concurrent measures of concepts of words in print and spelling with beginning and ending consonants (gathered in February of kindergarten), which were followed by phonemic segmentation abilities (gathered in May of kindergarten). Students' phonemic segmentation abilities then predicted word recognition ability in the fall of first grade, which subsequently predicted reading in context at the end of first grade. Their study highlights the developmental process of reading and the changing predictive validity of measures over time.

Kirby and colleagues (2003) used latent measures of phonological awareness and naming speed and found that they were statistically significant predictors of first- and second-grade reading outcomes. The patterns were reversed for the two latent predictors with phonological awareness becoming a weaker predictor as time passed and naming speed becoming a stronger predictor. In addition, the strength of prediction decreased as the elapsed time between administration of predictive measures and outcome assessments increased. Findings from the Kirby et al. (2003) study indicate that the role and strength of predictors can change over time.

Pennington and Lefly (2001) examined the changing role of predictors for children classified as either high-risk (HR) or low-risk (LR) for developing a reading disability from pre-kindergarten (pre-K) to the end of second grade. They were interested in how well phonologic tasks and letter-name knowledge predicted outcomes in first and second grade. For the LR group, phonological awareness assessed at pre-K, kindergarten, and first grade was the best predictor of second-grade outcomes. For the HR group, they found that the predictive validity of variables changed over time with letter-name knowledge being the best predictor in pre-K and kindergarten and then phonological awareness becoming the strongest predictor in first grade. They concluded that the reason for the differences in predictive validity over time was that the HR group had a developmental shift that had occurred earlier in the LR group. In other words, the predictive validity of measures likely changed and followed a developmental trajectory of reading for both groups but the HR group lagged behind the LR group.

Summary and Research Questions

Prior studies examining kindergarten prediction of longitudinal outcomes suggest that knowledge of letter names and sounds as well as phonemic processing skills are particularly strong predictors of a range of longitudinal reading outcomes. Although the studies reviewed all used some measure of phonemic processing, there was great variability in both the number and type of tasks used and most formed a composite predictor composed of multiple phonemic-related skills. In addition, these predictors were most often used as static measures.

There is evidence suggesting that reasonable predictive power of longitudinal outcomes can be achieved in the kindergarten years. Several studies found that longitudinal prediction can be achieved in the first half of the kindergarten year (Chiappe et al., 2002; Morris, Bloodgood, & Perney, 2003; Schatschneider et al., 2004). In addition, there is some evidence that prediction follows the developmental pattern of reading and that the predictive power of indicators changes over time (Kirby et al., 2003; Pennington & Lefly, 2001). Easier skills such as producing letter names and sounds are early predictors, whereas more difficult phonemic processing skills (e.g., phoneme segmentation, blending, or elision) become more predictive over time. In addition to examining the differential predictive power between the beginning and the end of kindergarten, research needs to examine whether the strength of a predictor changes throughout the kindergarten year.

Establishing the utility of progress-monitoring measures gathered throughout the kindergarten year is important, especially within the context of Tier 2 reading interventions. The authors could not find a single study that examined the longitudinal predictive validity of progress-monitoring measures, including CEMs, for children in a Tier 2 intervention. Although Pennington and Lefly (2001) found that there was a difference in predictors for HR and LR students, their study did not examine children in a Tier 2 intervention. Students are selected for participation in a Tier 2 intervention because of their potential to develop reading disabilities or significant reading difficulties. Therefore, predictive studies are important because few studies have examined these students as a special population or even included them in samples (Torgesen, 1998), which may lead to flawed findings for predictive applications (Badian, 1995).

One way of establishing the utility of indicators of early reading skills is examining predictive validity. The predictive validity of a variable is its ability to explain variance on an outcome variable and is commonly measured by R^2. In addition to the amount of variance explained, statistical

techniques such as structural equation modeling allow the use of latent outcome variables that capture broader elements of reading skills than can be done using a single outcome measure. Having multiple-component outcome variables is important, especially in reading that encompasses several distinct skills (McCardle, Scarborough, & Catts, 2001; Speece, Mills, Ritchey, & Hillman, 2003).

Another need in the research on kindergarten longitudinal reading outcomes is identifying a parsimonious set of predictors, especially phonological predictors. Although letter naming and sounds are established single-skill predictors (Torgesen, 1998), the independent phonological skills most predictive have not been clearly delineated. Phonological processing skills are predictively valid, but which individual skills are valid and when is unknown.

The study makes a unique contribution to the literature by focusing on kindergarten CEMs of specific skills and examining their changing role for predicting longitudinal outcomes in the first and second grade. We hypothesized that the role of predictors would change over time according to reading development with easier skills (e.g., knowledge of letter names and sounds) being predictive early and more difficult skills (e.g., whole word segmentation and word reading) predicting outcomes as the kindergarten year progressed. We also hypothesized that, consistent with previous findings (Kirby et al., 2003; Schatschneider et al., 2004), the amount of variance explained would increase throughout the kindergarten year. The following research questions were addressed:

Research Question 1: Which specific early reading skills measured by CEMs multiple times throughout the kindergarten year are predictive of comprehensive reading outcomes at the end of first and second grade?
Research Question 2: Does their predictive validity change throughout the year following the development of reading?

Method

Research Context

This study examined CEMs from the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004), a kindergarten reading curriculum designed to provide intensive instruction on key early literacy skills (i.e., phonological, alphabetic, decoding, spelling, and sentence reading). The curriculum consists of 126 lessons organized in four parts: (a) Learning Letters and Sounds (42 lessons); (b) Segmenting, Blending, and Integrating (30 lessons); (c) Reading (24 lessons); and (d) Reading Sentences and Storybooks (30 lessons). On average, students completed 112 lessons from the intervention delivered in groups of 3 to 5 for roughly 30 min a day, 5 days a week. Data for this study are from three experimental studies comparing the effects of ERI that used progress-monitoring data to inform instructional decisions.

Students in the current study were children identified in kindergarten as being at risk of developing reading difficulties. Student data from three cohorts spanning 3 years were used in this study. In the first two cohorts, students were assigned to either an ERI treatment condition or typical practice. In the third cohort, students were assigned to either a standard implementation of the ERI program or a condition receiving modified implementation. The modified implementation used the same intervention material as standard implementation; however, the students in the modified version were allowed to either accelerate/repeat lessons and allowed to regroup throughout the year. In the current study, only students from the treatment condition were included in the analyses.

Setting and Participants

Students. Students included 299 kindergarteners from three cohorts across 3 years. In each cohort, students were determined to be at risk of developing reading difficulties at the beginning of the year using school screening and nominations. Those who were recommended by the school and had parental consent were further screened by the researchers to determine eligibility for participation. In each cohort, students were first administered the Letter Naming Fluency (LNF) subtest from Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and Sound Matching (SM) subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). In the first two cohorts, students who had a raw score of six (i.e., 36th percentile) or below on LNF and 37th percentile below on SM qualified for participation. These cut points were used to ensure the selection of students who were at risk of developing reading difficulties and to avoid false negatives. An additional requirement for students in the third cohort was a standard score of 7 (16th percentile) or below on the Rapid Object Naming (RON) from the CTOPP or a standard score of 80 (9th percentile) or below on the Letter Identification (LI) subtest of the Woodcock Reading Mastery Tests–Revised/Normative Update (WRMT-R/NU; Woodcock, 1987/1998). The additional measures with lower cut scores for students in the third cohort were used because findings from the first two cohorts indicated an overidentification (i.e., false positives) of at-risk students. Differences among states and cohorts on pretests measures were controlled for by entering them as covariates in the structural models used in the analyses. A total of 348 students met the criteria for participation at the beginning of kindergarten, and 299 of those students completed kindergarten posttests, which is an attrition rate of 14%. The attrition rate from end of kindergarten to the end of first was 16% and from the end of first grade to the end of second was 45%. However, the loss of 45% of

Table 1. Student Demographics.

Variable	Participants (N = 299)	
	n	%
Gender		
Male	167	44.1
Female	132	55.9
Ethnicity		
Asian	1	0.3
American Indian or Alaska Native	2	0.7
Black or African American	51	17.0
Hispanic or Latino	121	40.5
White	118	39.5
Other	6	2.0
Identified for special education	34	11.4
English language learner	57	19.1
Variable	*M* (*SD*)	
Age	5.47 (0.34)	
Letter ID[a]	85.09 (10.56)	
Sound Matching[b]	22.56 (10.04)	
Rapid Object Naming[a]	7.73 (2.74)	
Letter Naming Fluency[c]	1.20 (1.77)	

[a]Standard score. [b]Percentile Score. [c]Raw Score.

the sample includes the entire Year 3 cohort (n = 103) for which second-grade data were not gathered. The attrition rate of students from the first and second cohorts from end of first to the end of second grade was 8%. Statistical comparisons of students revealed no statistically significant differences on demographic variables between those who remained in the sample and those who attrited. Table 1 summarizes student demographics.

Assessment Procedures

All participating students were administered four pretests (i.e., LNF, SM, RON, and LI) that were used to identify qualifying participants prior to the beginning of the intervention and roughly 6 weeks into their kindergarten school year. Students were removed from their classroom and tested one-on-one by trained assessors who were members of the research team. Assessors received a minimum of 8 hr of training to administer the assessments and were required to achieve 100% accuracy before independently assessing students. All assessment protocols were double scored by two independent research team members. Posttesting procedures were conducted in the same manner and occurred within 2 weeks of the end of intervention.

Predictor variables. Predictors included CEMs measured approximately every 8 weeks with Measurement 1 occurring in the beginning of January, Measurement 2 in the middle of March, and Measurement 3 at the end of April. The measurements were given following the first three of the four curriculum parts. The last CEM following the fourth curriculum part was not used as instruction was not completed prior to the finish of the intervention and posttesting.

The first CEM measures assessed material covered in the first 42 lessons. There were a total of six subtests comprising phonemic, alphabetic, and integrated tasks. The two phonemic subtests (First Sounds and Last Sounds) required the student to provide the first and last sounds of words presented orally. The student was presented a picture, and the examiner spoke the word represented by the picture and then asked the student to provide the first and last sounds of the word. The number of correctly provided first and last sounds was scored separately. The two alphabetic tasks (Letter Names and Letter Sounds) required the students to correctly provide the letter names and letter sound of the letters *m, p, f, c, t, s, d, l, a, o*, and *r*. The final two subtests (First-Letter Sound and Last-Letter Sound) integrated phonemic and alphabetic skills. Students were provided *d, f, l, m, p, r, s*, and *t* letter tiles and a stimulus page containing a picture with three blank boxes below. The student was required to put the tiles representing the first and last sound of the pictured object in his or her appropriate boxes.

The second CEM measure was administered after approximately 72 lessons. The letters *b, i, n, g*, and *u* were measured in addition to the letters from the first CEM. The same subtests and procedures as the first CEM were used. In addition, a new phonemic subtest was introduced that assessed the student's ability to segment whole words. The Whole Word Segmentation subtest required the students to segment vowel–consonant (VC) and consonant–vowel–consonant (CVC) words into their individual sounds. The examiner presented the words orally and students were asked to orally provide the correct constituent sounds.

The third CEM measure was administered following a total of approximately 96 lessons. It included the First Sound and Whole Word Segmentation phonemic subtests and the Letter Names and Letter Sounds alphabetic subtests from the first two CEMs. The letters *j, w, e, z, h*, and *y* were added to the battery. Two additional integrated tasks were added to the third CEM. The Medial Sounds subtest required the student to provide the medial sound in a word presented orally and represented by a picture. Using the same procedures as the First-Letter and Last-Letter Sounds subtests, the student was required to place the letter tile for the mediate sound in a CVC word in the middle box presented on the stimulus page. In the Word Reading subtest, words were presented on a stimulus sheet and students provided the individual sounds for a VC or CVC word and then read the entire word. A response was scored as correct if the word was accurately read. Table 2 summarizes the CEMs by administration point and reports their reliability coefficients in the sample.

Table 2. Composition and Reliability of CEMs by Measurement Point.

CEM subtests	CEM 1 (early January)	CEM 2 (mid-March)	CEM 3 (end of April)
Phonemic			
First Sound	✓	✓	✓
Last Sound	✓	✓	
Whole Word Segmenting		✓	✓
Alphabetic			
Letter Name	✓	✓	✓
Letter Sound	✓	✓	✓
Integrated			
First Sound Tile	✓	✓	✓
Last Sound Tile	✓	✓	✓
Medial Sound Tile			✓
Word Reading			✓
Reliability (Cronbach's α)	.89	.87	.90

Note. CEM = curriculum-embedded measure.

Outcome variables

End of first grade. To measure reading outcomes at the end of first grade, a latent reading construct comprised of six variables related to reading was used. The Word Attack (WA) subtest from the WRMT-R/NU and the Nonsense Word Fluency (NWF) test from DIBELS were used to measure decoding of nonsense/pseudowords. On the NWF test, a student has 1 min to orally produce as many nonsense words or segments as possible. Alternate form reliability when measured 1 month apart is .83 (Good et al., 2004). On WA, students decode pseudowords. Unlike NWF, it is an untimed measure. The split-half reliability of WA is reported as .87 in the technical manual. The Word Identification (WI) subtest from the WRMT-R/NU assessed sight words and also words used infrequently in the English language. Word identification is an untimed measure and has a median split-half reliability of .97 in the technical manual. Oral Reading Fluency (ORF) from DIBELS was used as a measure of fluent reading of connected text. Students earned a correct-words-per-minute score that indicates both the accuracy and fluency of their reading based on reading a passage for 1 min. Alternate form reliability ranges from .89 to .94 as reported by Tindal, Marston, and Deno (1983). The Passage Comprehension (PC) subtest from the WRMT-R/NU measures comprehension and requires a student to correctly provide the missing word in a passage of one to three sentences. The median split-half reliability in first grade is .92 as reported in the technical manual. The Test of Written Spelling–4 (TWS-4; Larsen, Hammill, & Moats, 2005) measures spelling ability by asking the student to write words presented orally. This measure was included because of the strong relation between spelling and early reading (Moats, 2005). It is norm-referenced and Cronbach's alpha as reported in the technical manual is .87 for 6-year-old students.

End of second grade. Reading outcomes at the end of second grade were assessed with a latent reading outcome variable composed of five measures. Measures included the WI, WA, and PC from WRMT-R/NU, ORF from DIBELS, and the TWS-4. The previous section provides task descriptions and reliability estimates.

Data Analyses

Data were analyzed using Mplus 6.12 and SPSS 20. The maximum likelihood with robust standard errors (MLR), which adjust standard errors by accounting for nonindependent data, was the estimation method. In Mplus 6.12, "TYPE = COMPLEX" with interventionist being the cluster variable was used to account for the nested nature of the data. In each model, entry scores from RON from the CTOPP, LI from the WRMT-R/NU, and LNF from DIBELS as well as demographic data were entered as covariates. The demographic variables included three dummy coded variables (Hispanic, African American, and Other ethnicity) with Caucasian as the reference group. For all analyses, we used a two-tailed test of significance at $p < .05$.

An exploratory factor analysis for the end of first grade was conducted using the WI, WA, and PC from the WRMT-R/NU; the ORF and NWF from DIBELS; and the TWS-4. The same process was followed for end of second grade with the exception that NWF was not used as an outcomes measure. Two single factor measurement models composed of all measured outcomes, one for the end of first grade and another for the end of second grade, were confirmed using confirmatory factor analysis (CFA) to estimate model fit. Once a measurement model had been confirmed, a total of six structural models were estimated. The six models were composed of one at each time point in

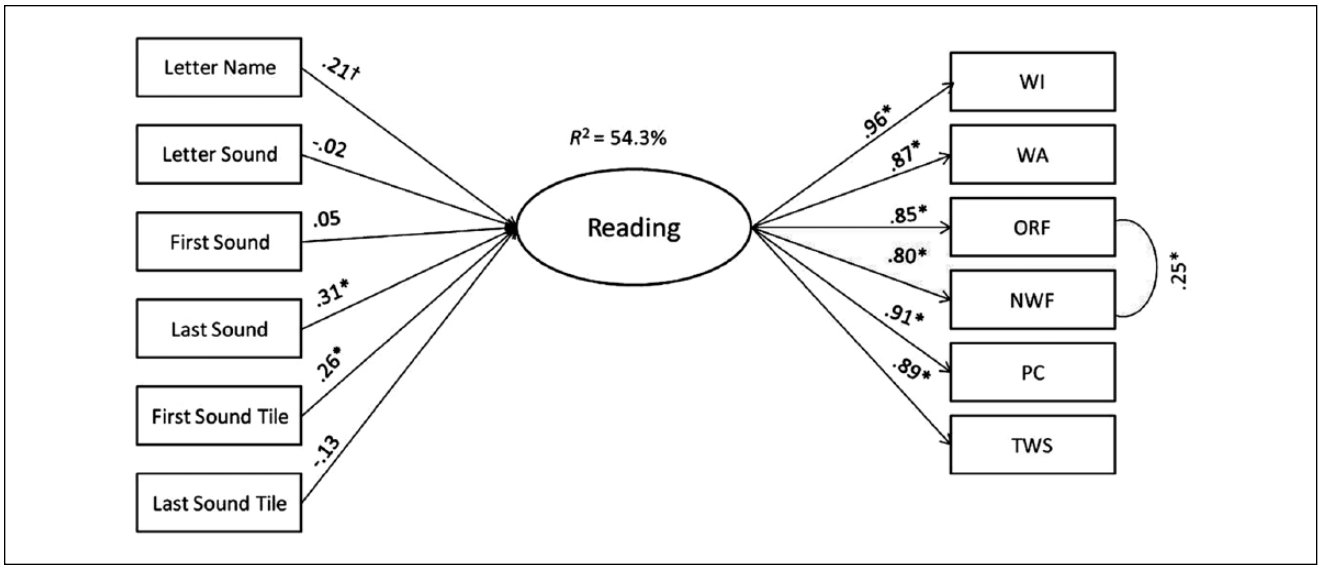

Figure 1. Structural equation model for the January curriculum-embedded measure (CEM) on end-of-first-grade outcomes.
†$p < .10$. *$p < .05$.

kindergarten (three time points) predicting outcomes at the end of first and second grades.

Results

End of First Grade

The exploratory factor analysis conducted for the end-of-first-grade scores indicated a single factor composed of all outcome variables. Analysis of Eigenvalues revealed one Eigenvalue of 5.96 that explained 74.58% of the variance; the next greatest Eigenvalue was 0.62 and explained an additional 7.79% variance. The single factor model was then confirmed using a CFA. The overall chi-square test value, $\chi^2(8) = 16.32$, $p = .038$, was statistically significant. However, fit indices indicated acceptable model fit with root mean square error of approximation (RMSEA) = .06, comparative fit index (CFI) = .99 and standardized root mean square residual (SRMR) = .01. The standardized path coefficients from the measurement model were all positively associated with the factor and statistically significant ($p < .01$). The R^2 for the measured variables, which measured the variance explained, ranged from 65.2% to 91.6%.

The first model (see Figure 1) used predictors from the first CEM (collected early January) to predict outcomes measured at the end of first grade. The chi-square test value, $\chi^2(68) = 136.40$, $p < .001$, was statistically significant; however, model fit indices indicate adequate fit (RMSEA = .06, CFI = .97, SRMR = .03). There were two statistically significant predictors and one predictor that approached significance. The Last-Letter Sounds ($\gamma = .31$, $p < .000$) and First-Letter Sound ($\gamma = .26$, $p = .003$) were statistically and positively related to the reading outcome. The Letter Names test ($\gamma = .21$, $p = .063$) approached significance. A total of 54.3% of the variance could be explained on the latent reading outcome factor.

The second model (see Figure 2) used the subtests from the second CEM (collected mid-March) as predictors. The chi-square test was statistically significant with $\chi^2(73) = 137.02$, $p < .001$. Model fit indices indicated good fit with RMSEA = .05, CFI = .97, and SRMR = .02. There were two statistically significant predictors of the latent reading factor. Letter Sounds ($\gamma = .21$, $p = .001$) and Whole Word Segmentation ($\gamma = .20$, $p = .014$) were positively related to the outcome. There were no other statistically significant predictors. A total of 55.2% of the variance was explained on the latent reading factor.

The final structural model for first grade was calculated using the CEM from late April (see Figure 3). The results of the chi-square test indicated a statistically significant finding with $\chi^2(78) = 133.86$, $p < .001$. Fit indices indicated good model fit with RMSEA = .05, CFI = .97, and SRMR = .02. The Letter Names subtest ($\gamma = .16$, $p = .010$) was statistically significant and positively related to the latent reading outcome. The Word Reading subtest ($\gamma = .32$, $p < .00$) was also positively related to the outcome and statistically significant, whereas the Medial Sounds subtest ($\gamma = .15$, $p = .054$) was approached significance. In total, the predictors were able to explain 62.9% of the variance.

End of Second Grade

An exploratory factor analysis for end-of-second-grade reading outcomes was conducted to examine the factor

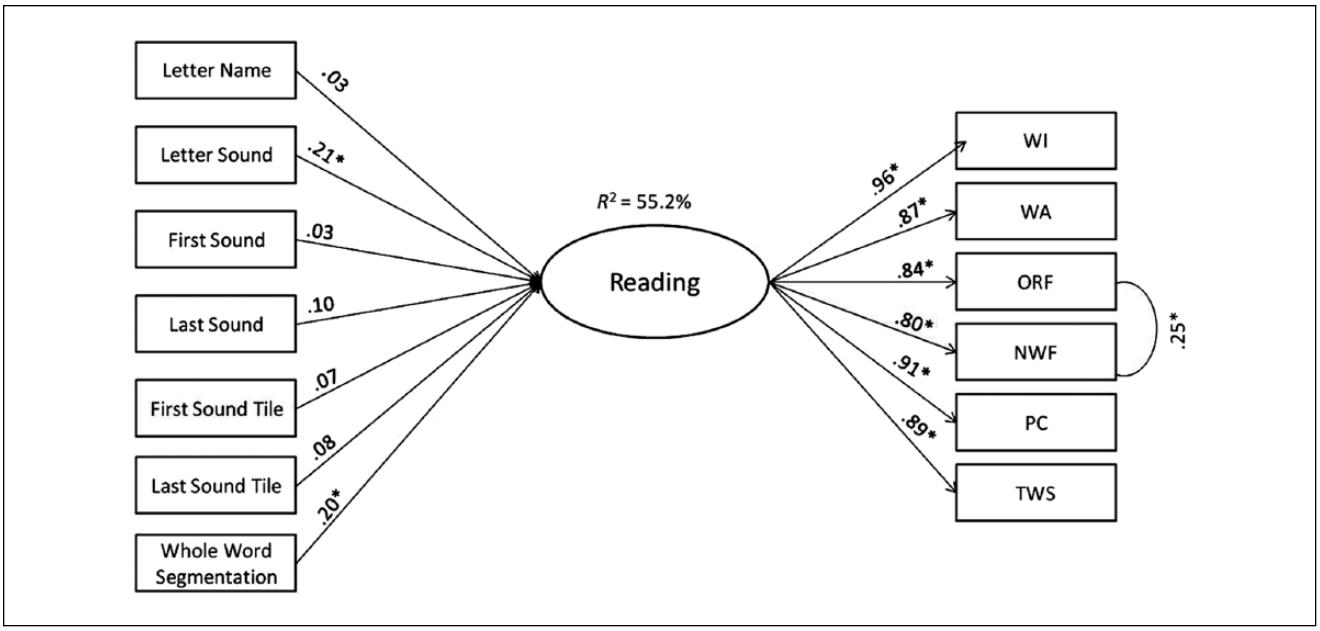

Figure 2. Structural equation model for the March curriculum-embedded measure (CEM) on end-of-first-grade outcomes.
*p < .05.

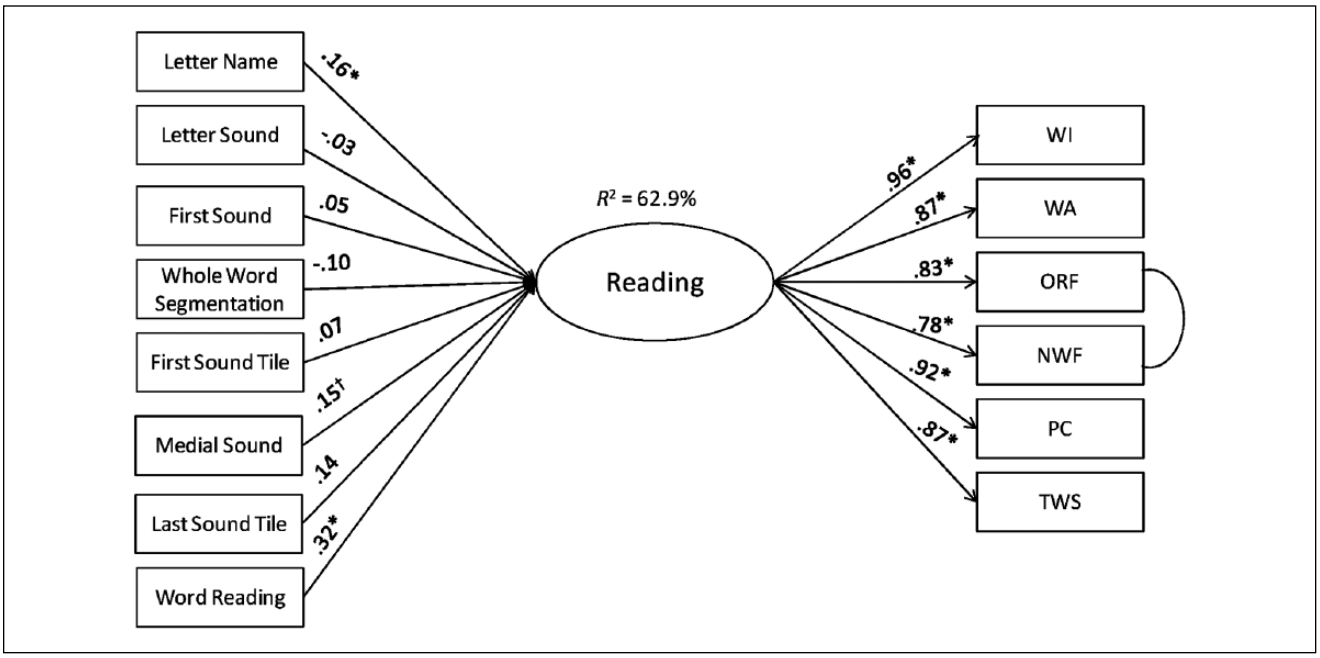

Figure 3. Structural equation model for the April curriculum-embedded measure (CEM) on end-of-first-grade outcomes.
†p < .10. *p < .05.

structure. The first Eigenvalue was 4.08 and explained 81.6% of the variance. The next greatest Eigenvalue was 0.41 and explained 8.1% of the variance. A single factor solution was then confirmed through a CFA. The model fit for the measurement model was good (RMSEA = .03, CFI = 1.00, SRMR = .04) with a nonsignificant chi-square value of 11.80, p = .299. All paths were statistically significant and positively related to the latent reading outcome variable. The variance explained using R^2 values for the measured variables ranged from 65.4% to 96.9%. Figures for the end of second grade will not be presented as there were few significant findings. The exogenous variables used in Figures 1 to 3 were the same predictors used for the end-of-second-grade model.

Table 3. Standardized Beta Weights for Predicting End-of-Second Grade Outcomes.

	Measurement 1 (early January)	Measurement 2 (mid-March)	Measurement 3 (end April)
CEM subtest predictors	β	β	β
Letter Name	.12	.02	.11
Letter Sound	−.07	.04	.21*
First Sound	.15†	.12	.01
Last Sound	.11	−.02	
Whole Word Segmentation		.10	−.06
First Sound with Tile	.09	.19	−.06
Last Sound with Tile	.00	−.01	−.09
Medial Sound			.24†
Word Reading			.17

Note. CEM = curriculum-embedded measure.
†$p < .10$. *$p < .05$.

Results from the first CEM subtests predicting end-of-second-grade outcomes revealed a statistically significant chi-square value, $\chi^2(53) = 96.53$, $p < .001$, with fit indices indicating adequate model fit (RMSEA = .05, CFI = .94, SRMR = .03). Although none of the predictors was statistically significant, the First Sounds subtest ($\gamma = .15$, $p = .060$) approached significance and a total of 34.1% of the variance was explained on the latent reading outcome.

The total variance explained by the second set of CEM subtests was 34.9% with a chi-square value, $\chi^2(57) = 95.04$, $p = .001$. Fit indices indicated good model fit with RMSEA = .05, CFI = .95, and SRMR = .03. There were no statistically significant predictors for the second CEM measure. The final model used the third CEM subtests as predictors with a statistically significant chi-square value, $\chi^2(61) = 111.07$, $p < .001$, with acceptable fit indices (RMSEA = .05, CFI = .94, SRMR = .03). The Letter Sounds subtest was a statistically significant predictor ($\gamma = .21$, $p = .045$), and the total variance explained on the latent reading outcome was 40.6%. Table 3 provides the standardized beta weights for the three CEMs predicting end-of-second-grade outcomes.

Discussion

Measures that can validly inform instructional decisions are central to effective RTI models for children identified as at risk of developing reading disabilities or significant reading difficulties. This study evaluated CEM subtests administered to kindergarten students participating in Tier 2 intervention throughout the kindergarten year and examined their utility for predicting reading outcomes at the end of first and second grades. We were interested in (a) which individual skills were statistically significant predictors of future reading outcomes and (b) whether their predictive power changed over time following reading development. We used structural equation modeling to evaluate the effectiveness of CEM subtests in predicting latent reading outcomes for the end of first and second grades. Findings add to our understanding of the predictive validity of measures that are conducted in conjunction with Tier 2 intervention. In summary, findings indicate that parsimonious sets of CEMs were able to explain reasonably large amounts of variance at the end of first grade but less at the end of second and that their predictive validity changed across the kindergarten year concurrent with reading development, which is consistent with the Ehri and McCormick (1998) model of word reading development. Following, we discuss findings and implications for first- and second-grade outcomes.

Predicting First-Grade Outcomes

Findings indicated that CEMs associated with ERI programs can predict end-of-first-grade outcomes for children at risk of reading disabilities as early as January of the kindergarten year. Specifically, the phonemic tasks that require a student to isolate and produce the last sound of a word presented orally best predicted the reading latent outcome variable. Although prior studies have identified phonological processing tasks administered in the first half of kindergarten as valid predictors of end-of-first-grade outcomes (Chiappe et al., 2002; Scanlon & Vellutino, 1996; Schatschneider et al., 2004), there was considerable range in the skills that were used. For example, phonological processing predictors have been composed of at least six individual tasks, whereas the present study isolated one phonological processing predictor. In addition, a task that integrated alphabetic and phonemic knowledge (i.e., correctly providing the tile representing the first sound in a word presented orally) was a statistically significant predictor of first-grade reading outcomes. Finally, knowledge of letter names in kindergarten approached statistical significance ($p = .063$); letter naming has been identified as a

viable predictor in several prior studies (Chiappe et al., 2002; Scanlon & Vellutino, 1996; Schatschneider et al., 2004). In total, 54.3% of the variance in first-grade reading outcomes was explained by CEMs administered in January of kindergarten, indicating that by midyear, kindergarten CEMs can provide substantial amounts of information vital to making informed instructional decisions.

The CEM administered in March of kindergarten indicated two statistically significant predictors of end-of-first-grade outcomes, and the entire CEM explained 55.2% of the variance on end-of-first-grade reading outcomes. First, the ability to produce the sounds of letters presented orally was a statistically significant predictor. Schatschneider et al. (2004) also found that letter-sound knowledge was a significant predictor of end-of-first-grade outcomes. The second significant predictor from the March CEMs was Whole Word Segmentation, which was the most difficult task measured at this time point. This finding aligns with previous studies (Morris, Bloodgood, & Perney, 2003; Scanlon & Vellutino, 1996) and is also consistent with Ehri and McCormick's (1998) model of reading development where a more difficult pre-reading task (e.g., Whole Word Segmentation) is predictive following reading development.

For the final CEM (administered in April), the Letter Naming and Word Reading subtests were statistically significant predictors of end-of-first-grade outcomes. That Letter Naming was a statistically significant predictor reiterates the strength of letter-name knowledge measured throughout the kindergarten year as a predictor of broad reading outcomes. The Word Reading subtest was the most difficult task in the April CEM and is in agreement with the Ehri and McCormick (1998) model of word reading. In addition, 62.9% of the variance was explained by the April CEM, which is higher than what previous studies have found (Schatschneider et al., 2004). The finding that Whole Word Segmentation was not statistically significant at this time point was unexpected, especially considering it is a more difficult task than Letter Naming. However, the multicollinearity ($r = .66$) between Whole Word Segmentation and Word Reading likely resulted in Word Reading getting the "credit" for the common variance explained by the two variables.

The findings for the end-of-first-grade outcomes support the first hypothesis of this study that predictors would change over time, which aligns with earlier findings (Morris, Bloodgood, Lomax, & Perney, 2003; Schatschneider et al., 2004). In addition, the general pattern was that the predictors achieving statistical significance became the more difficult tasks throughout the year. The second hypothesis that the strength of prediction would increase over time as found by Kirby et al. (2003) was also confirmed; the total variance explained on the outcome measure increased from 54.3% in January to 62.9% in April.

Predicting Second-Grade Outcomes

A primary finding of this study was that there was only one statistically significant predictor of end-of-second-grade reading outcome—knowledge of letter sounds. Catts, Fey, Zhang, and Tomblin (1999) and Hogan, Catts, and Little (2005) found that phonological processing abilities predicted end-of-second-grade outcomes. Another phonemic task (i.e., producing the first sound of words presented orally) administered in January of kindergarten approached statistical significance ($p = .06$), which also aligns with these prior studies. Our observation of fewer statistically significant findings for predicting second-grade outcomes from kindergarten measures concurs with Kirby et al. (2003) who found that predictors' strength decreases the larger the time difference between when the predictors and outcomes are gathered.

For the April CEM administration, the integrated Medial Sounds subtest approached statistical significance ($p = .08$). However, findings are insufficient to support our hypothesis that predictors would change over time. Findings that the amount of variance explained on end-of-second-grade outcomes increased throughout the kindergarten year from 34.1% in January to 40.6% in April support our second hypothesis.

Limitations and Future Directions

Findings from this study must be considered in light of several limitations. The first factor that must be considered is loss of students from the end of first grade to the end of second grade. Second-grade outcomes for the third cohort of students were not collected, which resulted in a sample size that was roughly 55% of the first-grade sample. However, this is not unprecedented in longitudinal studies (Schatschneider et al., 2004). That limitation noted, maximum likelihood estimation methods, such as those used in Mplus, have been demonstrated to handle missing data well (Enders & Bandalos, 2001). Another limitation is that the findings are specific to CEMs from one Tier 2 intervention (Pearson/Scott Foresman, 2004). Future research should examine CEMs from other curricula and as well as other sources of curriculum-derived measures (e.g., teacher-made tests). A third limitation is that variation in the total amount of variance explained at different time points is likely caused by the inclusion of different CEMs at each occasion and the closer relation of later CEMs to the outcome measures, which makes the comparison of variance explained across time points difficult to interpret. An additional limitation is that the third cohort did have one additional criterion measure. However, we controlled for pretest differences on the measures used to identify participants. Finally, we have no information regarding the intensity and type of instruction provided to students in

first and second grades. It is probable that many aspects of instruction (e.g., grouping, delivery, instructional tier, and dosage) differed across children and grade levels. This introduces error variance that may have reduced the predictors' power the greater the time difference between collecting the predictors and the outcomes they were predicting.

Implications for Practice and Conclusion

This study examined the predictive validity and changing roles of measures embedded in an ERI gathered 3 times in the kindergarten year. The ability to predict longitudinal outcomes is important, especially for children who enter kindergarten at risk of developing reading disabilities. Knowing whether a student is likely to continue to be at risk in first grade could allow teachers to make instructional decisions early in kindergarten, which may in turn lead to improved outcomes and mitigate reading difficulties that become more intractable over time, thus preventing the "wait to fail" problem in special education (Compton, Fuchs, Fuchs, & Bryant, 2006).

With respect to implications for practice, findings indicate that measures of students' level of mastery on recently taught skills can serve as useful and valid predictors of end-of-first-grade outcomes. We consider this particularly relevant for teachers who provide Tier 2 intervention as the information they collect as part of instruction can provide an additional indicator of RTI and future risk. Importantly, we found that CEMs can predict reading difficulty early in kindergarten and may inform teachers of students who need more intensive intervention. The finding that single measures can provide valid predictions of performance at different points in time (i.e., Last Sound from the first CEM, Letter Sound and Whole Word Segmentation from the second CEM, and Word Reading from the third CEM), may help streamline measurement processes, or at minimum, increase interventionists' confidence in decisions they make. As a complement to CBMs, CEMs can provide teachers more information about specific areas in which a student is struggling that may improve instructional decision making. Like CBMs, however, single CEMs may serve as indicators of more pervasive reading skill difficulties and may require more diagnostic assessments. For example, the root difficulties of students who fail to master Whole Word Segmentation tasks may be in multiple areas (i.e., phonemic segmentation and letter-sound identification); therefore, for students who fail to master more advanced skills, teachers will need to look at performance on fundamental skills to identify instructional targets. Regarding end of second grade, the utility and validity of CEMs are less clear as only one predictor reached statistical significance. Although a number of measures approached significance, further research is needed to examine the utility and validity of skill-specific CEMs in predicting longitudinal reading outcomes.

Acknowledgments

We thank our testers and data-entry staff for their contributions. We also thank the school districts, administrators, teachers, interventionists, and children for their support and participation in this research.

Authors' Note

Eric L. Oslund is now at the Meadows Center for Preventing Educational Risk, University of Texas at Austin; Aaron B. Taylor is now at Columbus, Ohio. The content and positions in this article do not necessarily represent those of the funding agency.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Simmons is a coauthor of the Early Reading Intervention (ERI) program that was used in this study.

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 in part by Grant R324E06067, National Center for Special Education Research, from the Institute of Education Sciences (IES), U.S. Department of Education to Texas A&M University. The content and positions in this article do not necessarily represent those of the funding agency.

References

Al Otaiba, S., & Torgesen, J. K. (2007). Effects from intensive standardized kindergarten and first-grade interventions for the prevention of reading difficulties. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention the science and practice of assessment and intervention* (pp. 212–222). New York, NY: Springer.

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

Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly*, *26*, 189–200. doi:10.2307/1593651

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*, 331–361. doi:10.1207/s1532799xssr0304_2

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, *6*, 369–400. doi:10.1207/S1532799XSSR0604_04

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. doi:10.1037/0022-0663.98.2.394

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

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5

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

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 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=3

Gersten, R., & Dimino, J. A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly*, *41*, 99–108. doi:10.1598/RRQ.41.1.5

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.

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools*, *36*, 285–293. doi:10.1044/0161-1461(2005/029)

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, *95*, 453–464. doi:10.1037/0022-0663.95.3.452

Larsen, S. C., Hammill, D. D., & Moats, L. C. (2005). *Test of Written Spelling* (4th ed.). Austin, TX: PRO-ED.

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools*, *47*, 22–35. doi:10.1002/pits.20449

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

Moats, L. C. (2005). How spelling supports reading. *American Educator*, *6*, 12–22.

Morris, D., Bloodgood, J., & Perney, J. (2003). Kindergarten predictors of first- and second-grade reading achievement. *Elementary School Journal*, *104*, 93–109. doi:10.1086/499744

Morris, D., Bloodgood, J. W., Lomax, R. G., & Perney, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly*, *38*, 302–328. doi:10.1598/RRQ.38.3.1

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

Pennington, B. F., & Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child Development*, *72*, 816–833. doi:10.1111/1467-8624.00317

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities Research Reviews*, *2*, 54–63. doi:10.1002/(SICI)1098-2779(1996)2:1<54::AID-MRDD9>3.0.CO;2-X

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:10.1177/002246690303600403

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

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Report No. 9). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

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

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

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