AN EXAMINATION OF THE PREDICTIVE VALIDITY OF CURRICULUM-EMBEDDED MEASURES FOR KINDERGARTEN STUDENTS

A Dissertation

by

ERIC LARS OSLUND

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Educational Psychology

An Examination of the Predictive Validity of

Curriculum-Embedded Measures for Kindergarten Students

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF CURRICULUM-EMBEDDED MEASURES FOR KINDERGARTEN STUDENTS

A Dissertation

by

ERIC LARS OSLUND

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Shanna Hagan-Burke
	Deborah C. Simmons
Committee Members,	Oi-Man Kwok
	Aaron B. Taylor
Head of Department,	Victor Willson

August 2012

Major Subject: Educational Psychology

ABSTRACT

An Examination of the Predictive Validity of

Curriculum-Embedded Measures for Kindergarten Students.

(August 2012)

Eric Lars Oslund, B.S., California Polytechnic State University;

M.S., West Texas A&M University

Co-Chairs of Advisory Committee: Dr. Shanna Hagan-Burke
Dr. Deborah C. Simmons

The purpose of the present research was to examine the predictive validity of curriculum-embedded mastery-check measures (CEMs) for kindergarten students in Tier 2 intervention. Two studies examined the predictive validity, parsimony, and changing role of CEMs using a structural equation modeling (SEM) framework. Study 1 examined the ability of CEMs gathered throughout the kindergarten year to predict end-of-kindergarten latent reading outcomes. Study 2 examined the ability of kindergarten CEMs to predict end-of-first and end-of-second grade latent reading outcomes.

Study 1 used SEM with two latent outcomes (i.e., phonemic and decoding) composed of diverse measures of early reading skills gathered at the end of kindergarten. Findings indicated moderate to large effects, as measured by variance explained, for CEMs on predicting phonemic and decoding outcomes. For CEMs gathered at four time points throughout the kindergarten year, a parsimonious set of subtests emerged. In addition, the role of CEMs changed throughout the year as predictors reaching statistical significance were increasingly difficult. Findings indicated that an increased amount of variance could be explained on the outcomes measures as the year progressed.

Study 2 used one latent reading outcome factor gathered at the end of first and second grades. Findings for the end of first grade indicated that parsimonious sets of predictors from CEMs administered at three times during the kindergarten year predicted end-of-first grade outcomes. Additionally, the role of indicators changed during the year and the amount of variance explained increased from the first to third CEM. Results for the end of second grade indicated the variance explained on the outcome measure increased from the first CEM to the third CEM. When considering near-significant results, a pattern emerged demonstrating parsimonious subsets of indicators that changed during the kindergarten year.

Findings from both studies provided support for the predictive validity of CEMs gathered during kindergarten for students in Tier 2 intervention. Results from both studies demonstrated statistically significant subsets of predictors that emerged and changed during the kindergarten year congruent with reading development, which can be useful for informing educational decisions.

DEDICATION

To my wife, Allison, and son, Owen. Allison – your beauty, intelligence, strength and support are more than I could ever ask for. Owen – not a day has passed since you were born where you didn't make me smile and laugh. You bring more joy to your parents that you will ever know.

To my parents – you laid the foundation for any success I have ever had. Dad – you taught me to respect others and the value of hard work, often without saying a word. Mom – you taught me the importance of being compassionate and kind. Your hard work and success in raising a family is evident in your children's love and appreciation for you and all you have done.

ACKNOWLEDGMENTS

I would like to thank my co-chairs, Dr. Hagan-Burke and Dr. Simmons. Your support, dedication, and generosity over the past four years made this possible. You are tremendous researchers and even better mentors and have provided a most excellent model for me to follow. Thank you for all you have done.

I would also like to thank my committee members, Dr. Kwok and Dr. Taylor. You sharing your knowledge of all things statistical has had a great impact on me and helped shape my current interests and future endeavors. Thank you for always being approachable and helping me understand the analyses that made this dissertation possible.

Thanks to my cohorts – Rose, Melissa, Stacey, Leigh Ann, and Jaime. You each contributed to my education and helped me in more ways than you know. I thank you not only for being friendly, but also for being my friends. We have gone from asking "what on earth am I doing here" to "I think I may be able to pull this off" to "I can do this" to "I'm done" in four short years. I wish you the best in the future and look forward to following your important work.

Thank you to my parents for working so hard to provide me with so many opportunities throughout my life. You sacrificed and put your children first at every turn and a simple "thank you" does not come close to my appreciation and love for you. Thank you to my brothers and sister for your unwavering loyalty and love. Thank you for reminding me to be humble and relax. I could not be prouder of you as individuals and consider it my greatest honor to be your brother. Thank you to my in-laws for your love, support, and words of encouragement, which mean so much to me and are greatly appreciated.

Thank you to ERI research team, especially Leslie Simmons and Minjung Kim, for your extraordinarily hard work and assistance over the past four years. Thank you to the Department of Educational Psychology staff. Your responsiveness and guidance are greatly appreciated. Thank you to Dr. Clemens for filling in during my defense.

I saved the best and most important for last. Thank you Allison – your belief in me motivated me, your encouragement sustained me, and your support made life not only possible, but enjoyable. This accomplishment is due more to your hard work and dedication than it is to mine. Finally, thank you God, from whom all blessings flow, none of which I deserve and all of which I greatly appreciate.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Measurement and Intervention

Early intervention for students identified at risk for academic difficulties is vitally important. In particular, reading difficulties are much more difficult to remediate the longer a student struggles to read (Al Otaiba & Torgesen, 2007). Research of longitudinal outcomes has demonstrated that students who struggle to read at the end of first grade almost never achieve average reading ability by the end of elementary school (Al Otaiba & Torgesen, 2007). Failure to intervene early and change reading development trajectories can lead to poor outcomes even beyond high school graduation.

Response-to-intervention (RTI) provides methods to prevent reading failure by administering early intervention based on student need (Gersten & Dimino, 2006). The typical RTI model includes universal screening, multi-tiered instruction, progress monitoring, and data-driven decision making (Gersten et al., 2009; Mellard, McKnight, & Woods, 2009). In the general RTI process, as described by Fuchs, Mock, Morgan and Young (2003), all students receive effective general education and their progress is monitored. Students who fail to respond to the general curriculum are moved into a second tier where more intense instruction is provided, typically in a small group setting and several times per week. Students' progress continues to be monitored and those who do not respond move into a third tier of instruction that consists of more intensive intervention that may include greater dosage, smaller groups, individualized

---

This dissertation follows the style of *Journal of Educational Psychology*.

interventions, and possibly placement in special education (Gersten et al., 2009). The RTI model proposed by Fuchs et al. (2003) focuses on data-driven instructional decisions based on progress-monitoring data to improve student outcomes.

Early and continuing assessment of academic progress and the resulting instructional decisions are critically important to RTI and effective early intervention. Utilizing progress monitoring to make data-based decisions is especially important for young children, as it helps prevent a “wait to fail” model by providing effective instruction to a student who enters school at risk for reading difficulty (Compton, Fuchs, Fuchs, & Bryant, 2006). Because longitudinal findings have shown children with early reading risk (Al Otaiba & Torgesen, 2007) rarely catch up to their average-performing peers, failing to intervene early can cause the “wait to fail” scenario to become a larger “continue to fail” problem.

Curriculum-Embedded Progress-Monitoring Measures

Curriculum-embedded measures (CEMs) are derived directly from the curriculum and assess students’ mastery of the material taught. CEMs are in contrast to curriculum-based measures, which measure general student outcomes. While curriculum-based measures provide important information regarding whether students are on track to attain end-of-grade outcomes (Fuchs, 2004), the author proposes that CEMs can fill an important gap in understanding student response to intervention. Mellard, McKnight, and Woods (2009) stated that progress-monitoring measures “should be particularly sensitive to the effects of the intervention and thus should be similar to the tasks on which the learner has received the instruction” (p. 188). Because

CEMs are based on the curriculum, they provide data on how well the student is mastering the specific content taught. The distinction between measuring general outcomes and curriculum-specific outcomes is important in the context of RTI, where intervention decisions and planning are informed by progress-monitoring measures. CEMs tied directly to the intervention provide data about the specific needs of the student and can help inform decisions regarding how to refine or modify interventions to best serve the individual student. It is especially to important to know the student's mastery of material in early reading interventions because beginning reading skills (e.g., letter naming or letter-sound knowledge) are the foundation for future reading success or failure (Gersten et al., 2009).

Another advantage of CEMs is that they are dynamic and change in accordance with the curriculum being taught. The dynamic characteristic is important because the development of word reading is itself dynamic according the to the Ehri and McCormick model of learning to read (Ehri & McCormick, 1998). To gauge the response to a developmental progression of reading instructions requires measures of student progression to likewise be dynamic and concurrent with the intervention.

Measurement Considerations

As is the case with any measure used to make educational decisions, CEMs need to be technically adequate and demonstrate reliability and validity. The ability to validly predict future outcomes for early readers is vital because knowing the likely result of progression, especially if it is a poor outcome, can help teachers alter the course of reading development through informed decision making, and thus improve student outcomes. Predictive validity uses current indicators (e.g., student performance on a particular task such as letter naming) to predict later outcomes (e.g., word identification).

The process of reading requires the mastery and application of several pre-reading skills (e.g., alphabet knowledge, phonemic blending and segmentation, decoding) (Kim, Petscher, Foorman, & Zhou, 2010). A central premise of CEMs is that formative performance on taught skills will predict later reading outcomes and that predictors of reading outcomes should align with the skills taught previous to the assessment. While skills are sequential in development, they are not mutually exclusive; rather, they fall on a continuum and often overlap. For this reason, assessments used to predict outcomes should not only be dynamic, but also measure and reflect the multiple skills that contribute to reading outcomes.

Outcome measures should encompass multiple skills. As noted above, the reading process integrates several skills. When predicting reading outcomes, measures should reflect the complexity and broadness of reading (McCardle, Scarborough, & Catts, 2001; Speece, Mills, Ritchey, & Hillman, 2003). Most early reading studies focus on one outcome, assessing one skill at a time. While it is true that some outcome measures ostensibly include the subcomponent skills of reading (e.g., oral reading fluency subsumes decoding skills), a latent outcome composed of individual tasks provides more information about broad student performance than does a measure assumed to include multiple reading skills. The use of latent variables composed of several crucial reading outcomes provides a more comprehensive and accurate understanding of overall reading ability.

Central to predictive validity is the issue of parsimony. While many individual skills may predict reading outcomes, it is important to consider school resources. By determining measures that most strongly predict or account for the most explained variance, we may be able to streamline the formative assessment process and protect resources. Conceivably, to identify a parsimonious set of predictors will require that predictors of reading outcomes would change as well. Understanding the changing role of predictors is important practically, as it may allow teachers to administer fewer measures based on predictors' power at any given administration point. Because of the rapid development of early reading skills during kindergarten, knowing which skills predict outcomes for the end of kindergarten and beyond allows teachers to focus on the most relevant skills at appropriate times with the fewest measures, and therefore potentially lower costs.

Dissertation Purpose

The purpose of the research discussed in this dissertation was to evaluate the validity of CEMs gathered during the kindergarten year to predict reading outcomes gathered at the end of kindergarten, first grade, and second grade. Empirical support for CEMs used for prediction in kindergarten, particularly predictors gathered during the kindergarten year as opposed to the end of the year, is lacking. Gersten et al. (2009) recommended using CEMs in the Institute of Educational Sciences (IES) practice guide *Assisting Students Struggling with Reading: Response to Intervention and Multi-tier Intervention in the Primary Grades*; however, the authors found limited empirical evidence for their recommendation. In fact, panel members stated that no studies had used CEMs or any other form of progress monitoring as independent (i.e., predictor) variables and concluded, “no inferences can be drawn about its effectiveness based on the research reviewed” (p. 46).

In addition to limited evidence supporting CEMs gathered in kindergarten, there are no studies that have evaluated progress-monitoring measures for students in a Tier 2 intervention. This population warrants special attention because of their heightened risk status and because of the positive impact effective interventions can have when data-driven decisions are applied. The failure to include these students or examine them as an individual population can lead to misinformed decisions, which may have deleterious effects on student outcomes for an already vulnerable group.

Early interventions are critical, as are data-driven decisions made early enough in kindergarten to improve reading outcomes. As important as is it to have viable kindergarten predictors, identifying valid early measures has not been accomplished (Speece et al., 2003), and the ability to predict which children will have the most serious reading difficulties is still far from perfect (McCardle et al., 2001). The absence of measures of student progress in Tier 2 interventions that can validly predict student outcomes substantiates the need for additional research in assessment development (Boscardin, Muthén, Francis, & Baker, 2008).

This dissertation presents two studies that evaluated the validity of kindergarten predictors and their changing role over the course of the kindergarten year, as based on Ehri and McCormick’s (1998) model of word learning, recommendations of IES, and extant research. The first study focused on the predictive validity and parsimony of

CEMs from a Tier 2 intervention gathered four times in the kindergarten year on end-of-kindergarten reading outcomes. The second study evaluated the effectiveness of CEMs gathered three times in the kindergarten year on predicting longitudinal outcomes gathered at the end of first and second grades. Both studies included students receiving a Tier 2 intervention and used latent reading outcomes in a structural equation modeling framework.

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF KINDERGARTEN CURRICULUM-EMBEDDED MEASURES FOR END-OF-KINDERGARTEN OUTCOMES

Response-to-intervention (RTI) is a prominent method that links instruction and assessment to align intervention with student needs. A fundamental purpose of RTI is to provide preventative instruction for students identified at risk of academic difficulties (Lembke, McMaster, & Stecker, 2010). At the core of RTI, research-based instruction is provided and student progress is monitored to provide data to inform further instructional decisions (Seethhaler & Fuchs, 2010). Typically, RTI provides tiers or levels of instruction that are intensified as indicated by student performance on formative assessments.

Since it was formally authorized by congress in 2004 in the Individuals with Disabilities Education Improvement Act, RTI has gained increasing use in schools (Mellard et al., 2009). Despite this widespread use, several components of RTI warrant further research. In particular, there is need for further research establishing the validity of measures to monitor student progress in Tier 2 intervention (R. Gersten et al., 2009). To address this issue, the author designed a study to examine the validity of curriculum-embedded measures used to monitor the progress of students participating in a Tier 2 reading intervention in kindergarten. In this study, the author examined whether curriculum-embedded measures administered at four points during the year could validly predict end-of-kindergarten achievement of students identified as at risk of reading difficulty, and which measures explained the most variance at respective measurement points. My study of curriculum-embedded measures is based on several related literature sources and agency recommendations including research on progress monitoring, research on predictors of early reading, and research-guided recommendations by the research panel that authored the *Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades* practice guide (Gersten et al., 2009; hereafter used as "RTI practice guide") commissioned by the Institute of Education Sciences (IES).

Research on Progress Monitoring

Progress monitoring is an integral part of RTI in reading, and involves frequent measurement of performance to determine whether a student is responding to intervention. To evaluate students' progress, educators need measures that accurately gauge whether student performance is adequate to reach established end-of-year benchmarks. This is especially important early, as Torgesen (1998) highlighted. Torgesen indicated that by intervening early, we move from a remediation model to a prevention model, thus avoiding the "wait to fail" scenario that can have negative impacts on at-risk students. As important as progress-monitoring measures are for making instructional decisions, there is limited research examining their predictive validity, particularly in kindergarten. In the RTI practice guide, Gersten et al. (2009) reviewed 11 studies that met the criteria for inclusion; only three of the studies reported using mastery checks or progress-monitoring measures for informing Tier 2 instructional decisions.

While progress-monitoring studies in the context of Tier 2 interventions are limited, prior research shows that progress monitoring provides instructors data to inform educational decisions about a student's progress and further educational needs (Gersten & Dimino, 2006; Lembke et al., 2010). Fuchs, Deno, and Mirkin (1984) found that progress monitoring gives teachers a more realistic understanding about student performance, which has implications for instructional decisions. Progress monitoring has been shown to positively impact student achievement (Fuchs et al., 1984; Marston et al., 2007; Stecker, Fuchs, & Fuchs, 2005). In their review of the impact of progress monitoring on student achievement, Stecker, Fuchs, and Fuchs (2005) found that student achievement increased when teachers modified instruction based on progress monitoring data.

One of the essential requirements of progress-monitoring measures is that they validly predict future reading outcomes. Concerning predictive validity in reading, Gersten et al. (2009) stated, "Predictive validity is an index of how well the measure provides accurate information on future reading performance of students—and thus is critical" (p. 14). Predictive validity is the ability of a measure to predict outcomes on a dependent variable. Predictive validity is typically measured through R^2, which is a measure of how much variance an independent variable (i.e., the predictor) explains on an outcome variable (i.e., dependent variable). While there is research that has examined the predictive validity of particular measures administered at given points in time, few studies have examined the predictive validity of progress-monitoring measures in kindergarten, and valid measures are elusive (Bishop, 2003; Scarborough, 1998).

Research on Predictors of Early Reading

Part of the complexity of identifying valid progress-monitoring measures in kindergarten resides in the multiple skills that are learned and their changing importance and complexity throughout the process of learning to read. In kindergarten, children progress through multiple phases of word learning (Ehri & McCormick, 1998) that involve knowledge and integration of phonemic, alphabetic, and orthographic skills. Ehri and McCormick proposed five phases of word learning, three of which are particularly important for early readers and for monitoring progress.

In their model, children begin learning to read in a pre-alphabetic stage and have little to no understanding of the alphabetic system or letter knowledge (i.e., knowing letter names and letter sounds) (Ehri & McCormick, 1998). Children in this stage "need to acquire letter knowledge and phonemic awareness, and they need to engage in activities that strengthen this knowledge and incorporate it into their literacy activities" (p. 143). The next stage in Ehri and McCormick's (1998) model is the partial-alphabetic phase. In this phase, children begin to associate letters composing words into their associated sounds. Their knowledge of letter names and sounds increases in this stage, allowing them to identify not only letters or sounds in isolation, but also in the context of a word (e.g., identifying the first or last sound of a word). Further knowledge of the association among letters and their sounds is developed in the third phase, known as the full-alphabetic stage. Knowing the associations enables the student to complete more complex reading tasks such as segmenting and decoding novel words.

At issue is how to measure progress of multiple skills through methods that are not only technically adequate and meaningful for informing instruction, but that also follow the development of reading skills. Although research identifying valid progress-monitoring measures for students in Tier 2 intervention is limited, prior research has established several useful early indicators. Letter identification, also referred to as letter naming, has shown strong predictive validity of reading outcomes for kindergarten children in numerous studies (Gersten & Dimino, 2006; Scarborough, 1998; Speece et al., 2003). The ability to identify the correspondence between a letter and its sound, also known as letter-sound knowledge, is also a viable predictor (Scarborough, 1998; Speece et al., 2003). Additionally, phonological awareness and phonological processing measures have demonstrated predictive ability (Gersten & Dimino, 2006; Speece et al., 2003). Gersten et al. (2009) recommend using phonemic measures, particularly phoneme segmentation measures. The RTI Action Network suggests using phoneme segmentation fluency, rapid-letter naming, and letter-sound fluency (LSF; Fuchs, n.d.).

Speece, Mills, Ritchey, and Hillman (2003) examined the early reading performance of 39 children in a part-time kindergarten class with 13% ranked with low literacy by their teachers. In addition to several other predictor variables, phonological awareness, letter names, and letter-sound knowledge were measured in April. Phonological awareness was assessed by the blending and elision measures from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). In the letter names and sound test, children were asked to name the 10 most difficult letter names and sounds. The outcome measure was the letter word identification (LWI) subtest from the Woodcock Johnson Psycho-Educational Battery – Revised (WJ-R; Woodcock & Johnson, 1989) administered in June. The LWI subtest requires the child to identify letters and words in isolation. Phonological awareness and letter name identification accounted for 50% of the variance on the LWI task, with phonological awareness having the strongest Beta weight.

In a study examining the predictive validity of measures taken in January and February of kindergarten, Ritchey (2004) examined end-of-kindergarten reading performance in 92 kindergarten children. She included measures of letter name and letter-sound knowledge as well as a phonological composite composed of blending and elision measures and the rapid automatized naming of colors subtest taken from the CTOPP. The end-of-kindergarten performance on fluency measures was predicted by the initial level of performance in each skill. Additionally, knowledge of letter names predicted May performance on the letter naming fluency (LNF) task from Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). Letter-sound knowledge predicted both LNF and nonsense word fluency (NWF) measures taken in May. The phonological awareness composite score predicted outcomes on the phonemic segmentation task (PSF), also taken in May.

Further, Ritchey (2004) found that growth on measures of LSF, LNF, and NWF taken every three weeks starting in January until May was quadratic. The LNF and LSF growth slowed over the time period while the rate of the NWF growth increased. These findings support the idea that reading skills are developmental and that predictive ability changes during the course of reading instruction.

Vellutino, Scanlon, Zhang, and Schatschneider (2008) used RTI measures taken in December, March, and June of kindergarten to predict end-of-kindergarten word reading outcomes for children at risk of reading difficulty who were receiving an intervention. The RTI measures were described as "experimental tests assessing print concepts, phoneme segmentation, phoneme blending, word identification, letter-sound knowledge, letter-sound decoding, and spelling" (p. 449). Approximately half of the children were randomly assigned to an intervention condition and received small group instruction 30 minutes per day, twice a week. The comparison group did not receive any supplementary instruction. They found that children can be identified for reading difficulties at the beginning of kindergarten and that potential reading difficulties may be prevented in most children if they are provided with intervention at the onset of kindergarten. Additionally, growth in RTI measures explained significant variance in both growth and level of performance on word-level skills. RTI measures outperformed screening measures taken in the beginning of kindergarten in classifying children still at risk at the beginning of first grade. They concluded that children should be identified and given supplementary instruction/intervention at the beginning of kindergarten. They also concluded that educators may be able to develop RTI measures with predictive validity for end-of-kindergarten outcomes as well as being able to identify those children at the onset of first grade who still need supplementary instruction.

Curriculum-Embedded Measures to Monitor Progress of Students in Tier 2 Intervention

Gersten et al. (2009) recommended using mastery checks embedded in the curriculum to monitor student progress of students who were not making adequate progress in Tier 2. A form of progress monitoring, curriculum-embedded measures (CEMs) assess mastery of skills from the taught program or intervention. Student performance on these measures can provide teachers with valuable information about specific skills with which a student is struggling and assist teachers in making data-driven decisions. CEMs have the potential to serve the dual purposes of informing instructional decisions and predicting student outcomes, which can make them a valuable tool in early intervention.

Another feature of CEMs is that they are dynamic, as they change based on the skills and material taught over an interval of time. Although the need for progress-monitoring measures is critical, currently there is a paucity of research that has examined the predictive validity of CEMs in kindergarten. Dynamic indicators are important because of the developmental nature of learning to read. Establishing the predictive validity of CEMs is an important step in establishing their potential for informing instructional decisions and predicting later reading outcomes.

As important as CEMs are for progress monitoring, there are significant gaps in the research base regarding them. To this author's knowledge, only one study has examined and established the predictive validity of a CEM (Olinghouse, Lambert, & Compton, 2006). There have been no studies that examine the predictive validity of

CEMs taken throughout the kindergarten year on end-of-kindergarten outcomes and beyond. Additionally, there are no predictive studies on CEMs or any other type of progress-monitoring measures that comprise only students in Tier 2 intervention. It is important to examine these students separately from students not receiving intervention, as the predictive validity likely differs between the two groups. Although RTI is widespread and the RTI practice guide suggests using CEMs as progress monitoring measures, there is a lack of research establishing the reliability and predictive validity of CEMs taken throughout the kindergarten year to predict outcomes at the end of kindergarten and beyond.

Study Purpose

Research has established the need for multifaceted independent variables (predictors) that are parsimonious and sensitive to changes in skill over the kindergarten year. Using multi-skill measures that encompass several valid predictors improves the accuracy of prediction above what any single measure can provide. Multivariate approaches involving correlated but not identical predictors increase the precision with which reading difficulties can be predicted and provide the most promise in early prediction research (McCardle et al., 2001; Scarborough, 1998). Because reading skills develop over time and a student may perform poorly on a task at one time point and better the next, including multiple-skill tests and administering those at different times throughout the year may be the most useful means of detecting reading problems (McCardle et al., 2001).

Research indicates a need for more comprehensive outcome measures of early reading skills that capture the broad and rapidly changing range of pre-reading and early reading skills (McCardle et al., 2001; Speece et al., 2003). Multiple reading indicators that form latent variables could provide a potentially more accurate picture of the complex processes involved in reading. Multiple measures help buffer against the imperfection of any one measure, allowing for a more accurate assessment of performance (Gersten et al., 2005). Some students may struggle with decoding but not with segmenting, while for others the opposite might be true. Measuring only a single outcome does not give the entire picture of the reading process. So far, studies have mainly focused on using one type of outcome at a time (e.g., fluency, comprehension, or decoding) and few have examined broader latent constructs. To best identify the type of intervention a student needs and then serve the student accordingly requires using broad measures of reading outcomes that include more than just one measure of a specific skill. Examining the impact predictors have on latent reading constructs composed of manifold measures is another area in need of research.

Although limited, there is evidence that the predictive validity of specific skills changes over time and follows a developmental trajectory, with easier tasks being most predictive early and more difficult tasks becoming more predictive with further learning. The changing role of predictors should be expected because reading is a developmental process; however, little research has examined how predictors change over the course of the kindergarten year.

As Kim, Petscher, Foorman, and Zhou (2010) asserted, knowledge of the alphabet and the phonemic structure of language are building blocks of early literacy. Each study reviewed above used composite phonological awareness measures and, while phonological awareness tasks have been shown to have predictive validity, it is less clear what specific tasks are most predictive as phonological awareness is typically measured by several tasks that form a larger composite. As Torgesen (1998) illustrated, phonological awareness can be measured by as many as 20 tasks or more. It would be hard for a school to administer 20 phonemic tasks alone, not to mention additional alphabetic measures. Parsimony is an important consideration for in-school use, as resources are limited and the list of available measures is long. For this reason, a current research need is to identify a set of predictors that is both parsimonious and has acceptable predictive validity, especially considering there is always a trade-off between parsimony and maximum prediction.

Another need in kindergarten research is examining students in Tier 2 interventions and the role of CEMs as predictors. Three studies examined measures predicting outcomes measured at the end of the kindergarten year, and only one included progress-monitoring measures with students receiving Tier 2 intervention (Vellutino et al., 2008). Establishing the predictive validity of the CEMs in a Tier 2 intervention setting is important considering their use in informing instructional decisions.

The current study provides unique contributions to the study of kindergarten predictors based on the needs described above. It isolates specific early predictors (e.g., identifying first sounds); most research to date has examined large composites, especially of phonemic processing skills (Ritchey, 2004; Schatschneider, Francis, Carlson, Fletcher, & Foorman, 2004; Speece et al., 2003; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). This study examines latent outcomes composed of numerous early reading indicators; no previous studies have examined latent reading outcomes. This study also examines the changing role of predictors over the course of the kindergarten year following the developmental nature of reading using CEMs derived from an explicitly and systematically taught intervention. So far, no studies have examined the changing role of CEMs during kindergarten. Finally, this study examines only students in Tier 2 intervention, which only one other study has done. This study answers the following research questions:

1. What skill-specific progress-monitoring measures embedded in the curriculum and administered during the kindergarten year are most predictive of broad end-of-year outcomes for children receiving Tier 2 supplementary instruction?
2. At what time points can skill-specific CEMs be predictive of end-of-year outcomes and does the predictive validity of skill-specific CEMs change following a developmental progression?

Method

Research Context

Data for this study were collected as part of a larger randomized control 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 control condition. This study examines data from the students in the experimental condition as they were administered CEMs to monitor their progress. Students received an average of 102 lessons over 21 weeks. The lessons were 30 minutes daily and given to groups of 3-5 students. The intervention provided direct and explicit reading instruction in four parts. Curriculum-embedded measures were administered approximately every four weeks and students were regrouped based on performance to create homogeneous groups at each measurement occasion. Students were allowed to either repeat or accelerate lessons based on performance. While the data from curriculum-embedded measures were used to adjust instruction in this study, the predictive validity of the measures had not been established.

Setting and Participants

Students from ten schools from Florida, five from Connecticut, and two from Texas participated in the study. School enrollments ranged from 401-832 for Florida, 287-739 for Connecticut, and 279-889 for Texas. The percent of students eligible for free or reduced-cost lunches ranged from 63-92% for Florida, 70-82% for Connecticut, and 81-82% for Texas. All schools received Title 1 funding.

At the beginning of the year, 156 students were determined to be at risk for reading difficulties and therefore qualified for participation. Students with parental consent were identified through a screening process that included informal school recommendations and standardized testing. Schools nominated children based on school-administered assessments. Researchers then administered standardized assessments of pre-reading skills including letter naming fluency from DIBELS (Good & Kaminski, 2002) and sound matching (SM) from a subtest of CTOPP (Wagner, Torgesen, & Rashotte, 1999). Beginning criteria required the student to score at or below a raw score of 6 on LNF and at or below the 37th percentile on SM. Students who met those criteria were then administered the rapid object naming (RON) subtest from CTOPP and the letter identification (Letter ID) from the Woodcock Reading Mastery Test – Revised/Normative Update (WRMT-R/NU; Woodcock, 1998). Those students scoring standard scores of 7 or 80 on RON and Letter ID, respectively, qualified for participation.

Of those who completed pretests, 137 also completed posttests, which is an attrition rate of 12%. Analyses indicated no statistically significant differences between the students who completed the study and those who attrited on any demographic variables. Student demographics are provided in Table 1.

Assessment Procedures

Pretest measures assessed students' early reading skills. The pretest measures were administered prior to the start of the intervention and approximately six weeks after the school year began. 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 pretest measures. The measures were double-scored by two trained members of the research team to ensure accuracy. The posttesting procedures were the same as the pretesting and conducted within two weeks after completion of the intervention.

Predictor variables. CEMs are curriculum-dependent measures administered approximately every four weeks to monitor student mastery of previously taught content and skills.

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

	Participants (N = 137)	
Variable	N	(%)
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
Variable	Mean (*SD*)	
Age	5.44 (0.31)	
Letter ID[a]	80.65 (8.19)	
Sound matching[a]	19.71 (10.46)	
Rapid object naming[a]	6.08 (2.19)	
LNF[b]	1.20 (1.77)	

Note. [a]Standard score. [b]Raw score.

The intervention on which CEMs were based is the Early Reading Intervention (Pearson/Scott Foresman, 2004) that has four parts. CEMs were given at the middle and end of each part for a total of 8 measurement points (measurements 1-8). Because student groups could repeat or accelerate lessons based on their mastery of content, at time point 5 (i.e., mid-part 3), the lesson number different student groups were being taught began to overlap with subsequent measurement points. Therefore, only measurements 1-4 were used in the analyses. Measurements 1, 2, 3, and 4 were collected at approximately the end of October, end of December, middle of January, and middle of February, respectively.

Each CEM sampled previously taught skills and knowledge and was designed to reflect the developmental progression of kindergarten reading skills. CEMs are multiple-skill measures designed to measure mastery of previously taught content. The first CEM was composed of three individual subtests. The Letter Names subtest requires the student to name the letters *m, p, f, c, t, s,* and *d* displayed on a page. The Letter Sound subtest requires the students to provide the letter sound for the same letters used in the Letter Name test. The First Sounds in Words test requires the student to provide the first sound of a word presented orally and represented by a picture. The examiner presents a picture of an object (e.g., cat) and says the word represented by the picture. The student is then asked to say the first sound in the word.

The second CEM contains the same three subtests as the first CEM and includes three additional subtests. The letters *l, a, o,* and *r* are added. In the Last Sounds in Words subtest, the student is presented a word orally and a picture representing the word. The student is asked to say the last sound in the word. In the First Letter-Sound subtest, the student is given letter tiles *d, f, l, m, p, r, s,* and *t*. The student is then presented a picture of an object and below the picture are three blank squares. The student is asked to place the tile representing the first sound of the picture into the first square. In the Last Letter-Sound subtest, the student is given the same tiles and pictures as in the First Letter-Sound subtest; however, the student must place the tile representing the last sound of the letter in the last square.

With the exception of the Last Letter-Sound subtest, the third CEM contains the same subtests from the second CEM and includes one additional subtest. The Whole Word Segmentation subtest requires the student to segment three consonant-vowel-consonant (CVC) words presented orally into its individual phonemes. The letters *a, b, c, d, f, l, i, m, n, o, p, s,* and *t* were included in the third CEM. The fourth CEM includes all of the subtests from the first three CEMs. Table 2 summarizes the four CEMs used in the analyses and their estimated reliability in the measured sample using Cronbach's alpha.

Table 2 *Curriculum-Embedded Measures, Composition and Reliability Estimates*

Subtests	ERI Mastery-Check			
	1	2	3	4
Letter Names	√	√	√	√
Letter Sounds	√	√	√	√
First Sounds in Words	√	√	√	√
Last Sounds in Words		√	√	√
First Letter-Sound with Letter Tiles		√	√	√
Last Letter-Sound with Letter Tiles		√		√
Whole Word Segmentation (no picture display)			√	√
Cronbach's Alpha Reliability Estimate	.80	.90	.81	.84

Outcome measures. A total of seven outcome measures were administered at the end of the intervention.

Blending Words. The Blending Words (BW) subtest from CTOPP was administered to assess the student's ability to verbally blend individual sounds into whole words. Internal reliability alpha coefficients ranged from 0.86 to 0.89 for children ages 5 through 7 years old.

Sound Matching. The Sound Matching (SM) subtest from CTOPP was used to assess phonemic awareness. In SM, the student is presented a target picture and three additional pictures. The student is asked to match one of the three pictures to the target picture based on first or last sound. Internal reliability alpha coefficients range from 0.92 to 0.93 for children aged 5 through 7 years old (Wagner et al., 1999).

Phoneme Segmentation Fluency. In the Phoneme Segmentation Fluency (PSF) from DIBELS, a student is asked to orally produce the individual sounds of a stimulus word presented by the examiner. The student has one minute to identify as many individual sounds as possible from words with three to four phonemes. The alternate form reliability of forms given two weeks apart for PSF is 0.88 (Good et al., 2004).

Nonsense Word Fluency. Nonsense Word Fluency (NWF) from DIBELS assesses a student's ability to decode vowel-consonant (VC) and consonant-vowel-consonant (CVC) nonwords. The student is given one minute to correctly decode as many nonwords as possible. Alternate form reliability with a one-month interval is 0.83 (Good et al., 2004).

Oral Reading Fluency. Oral Reading Fluency (ORF) was measured using the "Mac Gets Well" passage (Makar, 1995). The student has one minute to correctly read as many words as possible in the passage. Internal reliability coefficients reported by Vadasy, Sanders, and Peyton (2008) were 0.93 for their kindergarten sample.

Word Identification. The Word Identification (WI) from the WRMT-R/NU measures the student's ability to read words of increasing difficulty and words used less frequently in the English language. An item is scored correct if the student is able to read the word with the correct pronunciation. In WI, the student may rely on both sight-word recognition and word decoding skills to correctly read the words. The median split-half reliability coefficient is 0.97 and concurrent validity is not reported.

Word Attack. The Word Attack (WA) from the WRMT-R/NU uses the same procedures as WI but uses pseudowords. The use of pseudowords requires the student to rely more on word-reading skills to effectively decode the words. Median split-half reliability is 0.87 and concurrent validity is not reported.

Data Analyses

Data were analyzed using Mplus 6.12 and SPSS 20. The maximum likelihood with robust standard errors (MLR) was used for the structural equation modeling analyses. The MLR estimator adjusts standard errors by taking into account non-independent data and uses all available data for estimation. Students were nested in interventionist and the “TYPE=COMPLEX” analysis was used with interventionist being the cluster variable.

A confirmatory factor analysis (CFA) was used to validate the two-factor measurement model. Three measured variables loaded on the phonemic latent variable and four loaded on the decoding latent outcome. Following the CFA analysis, a structural equation modeling (SEM) model predicting outcomes from each measurement time point was constructed resulting in a total of four models.

For each model, entry-level RON and demographic data were entered as covariates. Demographic data had three dummy coded variables (i.e., Hispanic, African-American, and other ethnicity) with Caucasian students as the reference group.

Results

The CFA confirmed a two-factor solution with an overall chi-square value of χ^2 (12) = 19.36, $p = 0.013$, which is not statistically significant. Fit indices showed good fit of the estimated measurement model (RMSEA = 0.00, CFI = 1.0, SRMR = 0.02). The phonemic latent factor was composed of the phonemic awareness measures (i.e., BW, SM, and PSF) and the decoding latent factor was composed of the decoding related measured outcome variables (i.e., NWF, WI, WA, ORF). All loadings on the phonemic and decoding latent factors were statistically significant ($p < 0.01$) and the measured variables were positively related to the latent factors on which they loaded. The decoding and phonemic factors were also positively related with each other. The variance explained (i.e., R^2) on the measured outcome variables ranged from 41-86%. Figure 1 shows the hypothesized measurement model with standardized factor loadings.

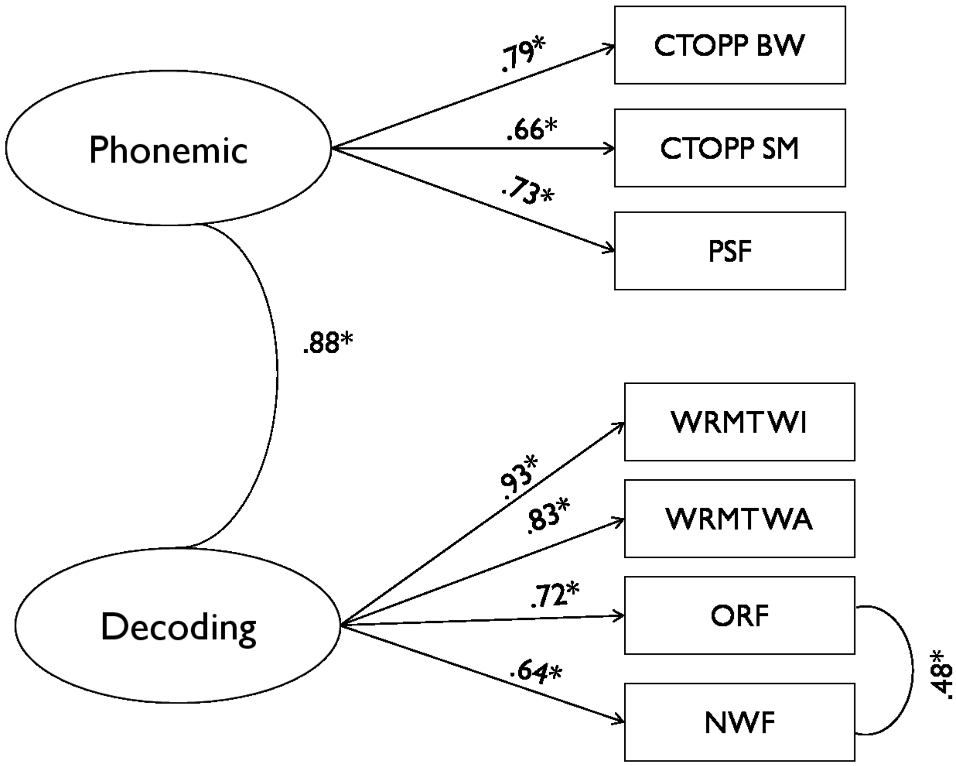

Figure 1. Measurement model. * p < 0.05.

The first model examined the predictive power of the individual subtests (i.e., Letter Names, Letter Sounds, and First Sounds in Words) from the first CEM, which was administered in October. The hypothesized model is presented in Figure 2 and includes the standardized coefficients. The chi-square test of model fit was statistically significant with χ^2 (47) = 63.45, p = 0.055. Although the chi-square test was statistically significant, fit indices (RMSEA = 0.05, CFI = 0.98, SRMR = 0.04) indicated the hypothesized model fit the data well. The Letter Names subtest was a statistically significant predictor both on the phonemic (γ = 0.52, p = 0.002) and decoding factors

($\gamma = 0.35$, $p = 0.037$). The First Sounds in Words subtest was a statistically significant predictor on the phonemic factor ($\gamma = 0.50$, $p < 0.000$). The Letter Sounds subtest did not reach statistical significance for the phonemic or decoding factors. The predictors explained a statistically significant amount of variance on both factors, with 62.0% of the variance explained on the phonemic factor and 36.3% explained on the decoding factor.

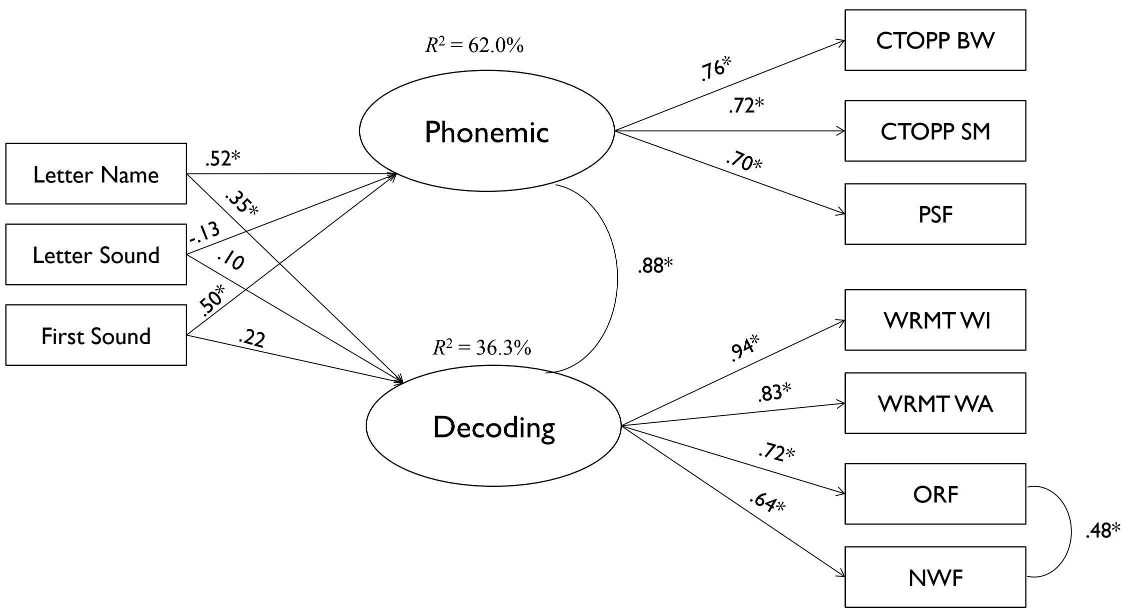

Figure 2. Time 1 model. * p < 0.05.

In the second model, the predictive validity of the six CEM subtests administered in December was examined. The overall chi-square value was statistically significant at χ^2 (62) = 69.90, $p = 0.229$; however, model fit indices indicated good fit of the model to the data (RMSEA = 0.03, CFI = 0.99, SRMR = 0.03). The First Sounds in Words (γ = 0.21, *p* = 0.041), Last Sounds in Words (γ = 0.21, *p* = 0.009) and First Letter-Sound (γ = 0.41, *p* = 0.003) were statistically significant predictors on the phonemic factor. The First Letter-Sound subtest was also a statistically significant predictor on the decoding factor (γ = 0.54, *p* = 0.001). The Letter Name, Letter Sound, and Last Letter-Sound subtests were not statistically significant predictors on either outcome factor. The total amount of variance explained on the phonemic factor was 79.0%, and 55.0% of the variance was explained on the decoding outcome. The model with standardized coefficients is presented in Figure 3.

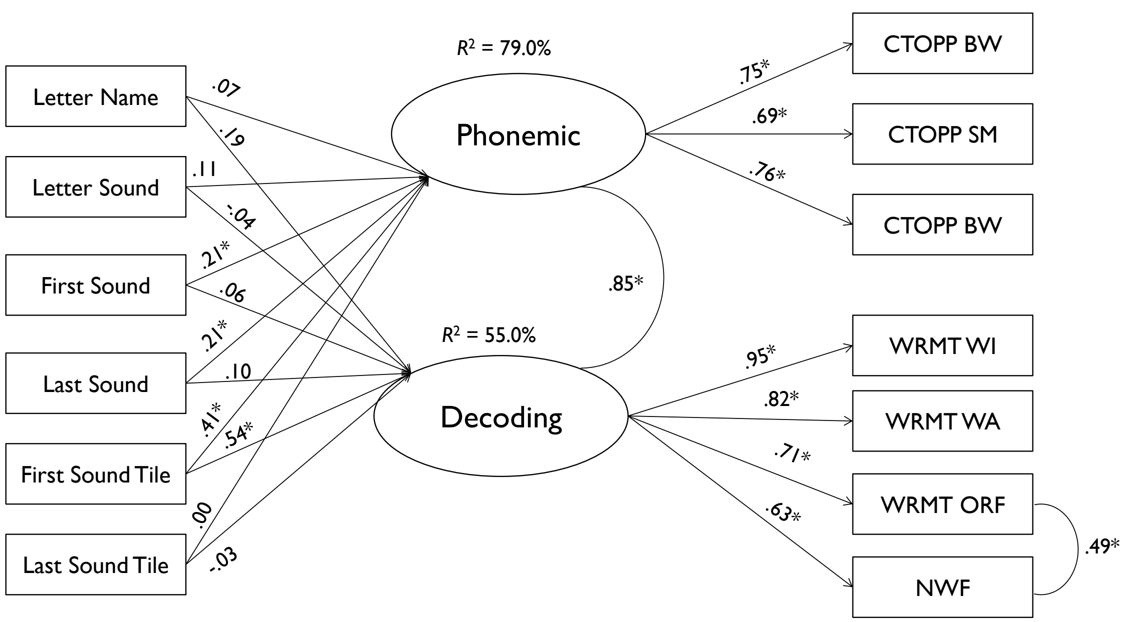

Figure 3. Time 2 model. * p < 0.05.

In the third model, the subtests from the January administration of the CEM were examined for predictive validity. The subtests included were Letter Name, Letter Sound, First Sounds in Words, Last Sounds in Words, First Letter-Sound and Whole Word Segmentation. The model fit the data reasonably well based on fit indices, although the overall chi-square value was statistically significant, $\chi^2(62) = 87.26$, $p = 0.02$; RMSEA = 0.06, CFI = 0.97 and SRMR = 0.04. The Letter Sound subtest was a statistically significant predictor both for the phonemic ($\gamma = 0.24$, $p = 0.045$) and the decoding factors ($\gamma = 0.53$, $p < 0.000$). First Letter-Sound was a statistically significant predictor for the phonemic factor ($\gamma = 0.27$, $p = 0.002$) and nearly significant for the decoding factor ($\gamma = 0.20$, $p = 0.066$). The other CEM subtests were not statistically significant for either factor. The variance explained on the phonemic and decoding factors was 79.6% and 57.7%, respectively. The third model is displayed in Figure 4.

The fourth model (see Figure 5) included all the subtests from the previous three CEMs, and the chi-square value of $\chi^2(67) = 102.06$, $p = 0.004$ was statistically significant. The fit indices indicated the model fit the data adequately (RMSEA = 0.06, CFI = 0.96 and SRMR = 0.04).

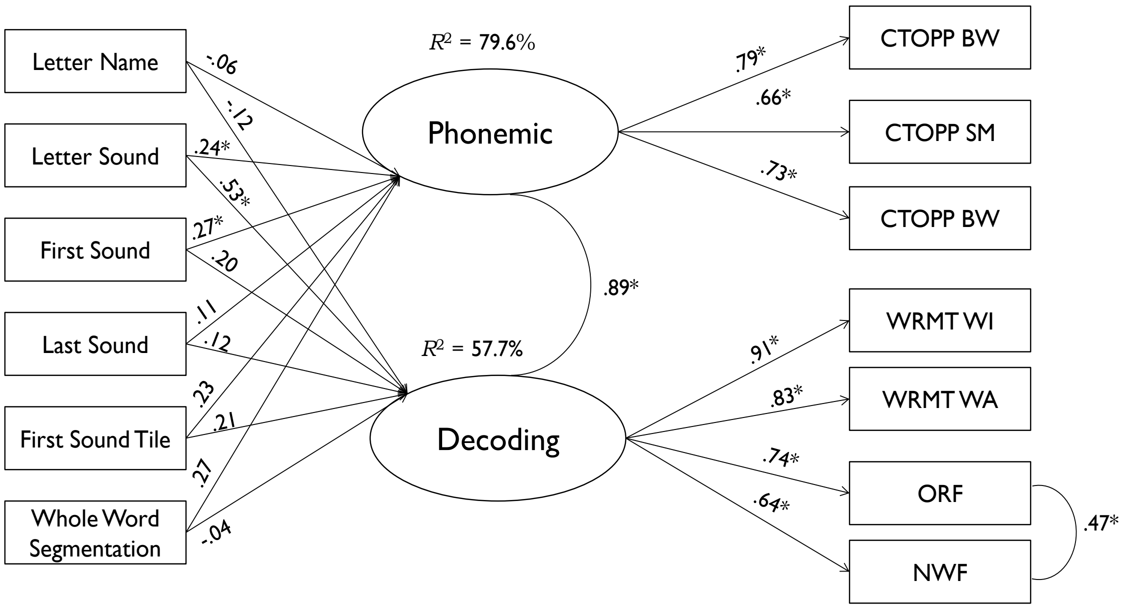

Figure 4. Time 3 model. * p < 0.05.

Whole Word Segmentation was a statistically significant predictor for the phonemic ($\gamma = 0.51$, $p < 0.000$) and decoding ($\gamma = 0.35$, $p < 0.000$) factors. In addition to Whole Word Segmentation, First Sounds in Words ($\gamma = 0.23$, $p = 0.026$) and First Letter-Sound ($\gamma = 0.25$, $p = 0.014$) were statistically significant predictors of the phonemic factor. Letter Names ($\gamma = 0.38$, $p < 0.000$) was also a statistically significant predictor for the decoding factor. No other subtests were statistically significant predictors for the phonemic or decoding factors. The total variance explained was 86.8% on the phonemic factor and 65.2% on the decoding factor.

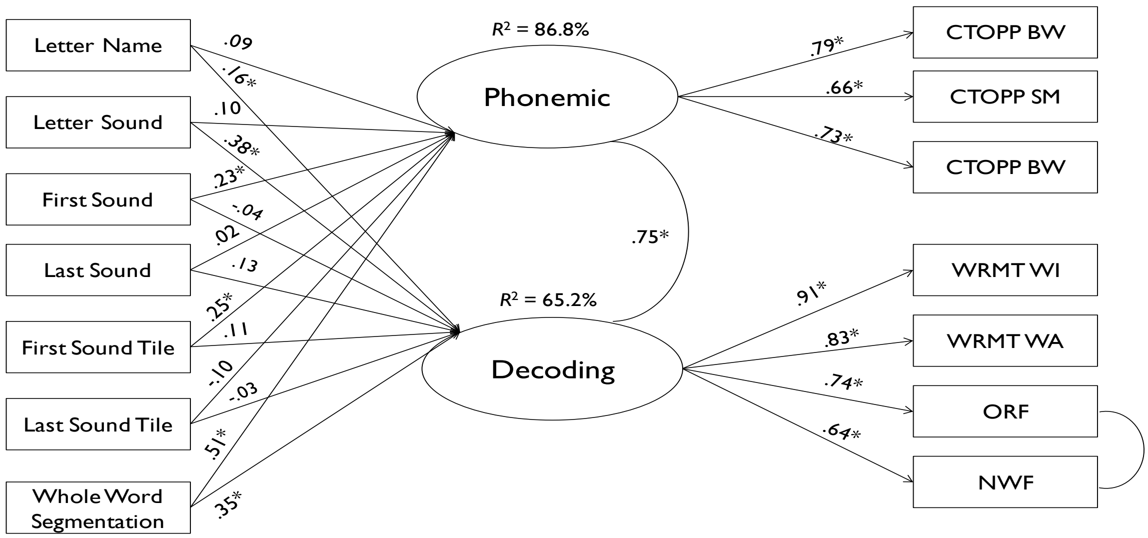

Figure 5. Time 4 model. * $p < 0.05$.

Discussion

The RTI practice guide for struggling readers in the primary grades recommends using mastery measures from intervention programs to monitor performance of students in Tier 2 intervention. To date, limited research has investigated the utility of this recommendation. In this study, the author examined the predictive validity of progress-monitoring CEMs on end-of-kindergarten outcomes. The CEMs were derived from the curriculum and designed to assess student response to intervention through measuring mastery of taught skills. Measures were administered approximately every four weeks, and tasks were modified across measurement points to reflect the developmental progression of skills in the reading intervention. Although there are a few studies examining predictors of end-of-kindergarten outcomes, studies examining the validity of curriculum-embedded mastery measures for students receiving Tier 2 intervention are limited. While research has shown that predictive validity changes throughout the course of the year following a developmental trajectory (i.e., easier tasks being predictive early on and more difficult tasks becoming predictive subsequently) (Ritchey, 2004; Schatschneider et al., 2004), to the author's knowledge the predictive validity of progress-monitoring measures has not been established with students participating in Tier 2 intervention. The author was interested in whether specific tasks that measured previously taught skills could validly predict end-of-kindergarten outcomes and whether their validity changed over time.

Findings indicated a dynamic set of tasks that predicted reading outcomes over the four measurement points. At the first time point, knowledge of letters names as measured by the CEM Letter Names subtest was a significant predictor for both the phonemic and decoding outcomes. This finding aligns with Ehri and McCormick's (1998) first stage in reading development (i.e., pre-alphabetic stage), where students are acquiring letter name knowledge. Also predictive at the first time point for the phonemic factor was knowledge of the first sound in a word presented orally. The combined predictive power of these three predictors used at the first time point explained 62% of the variance on the phonemic outcome and 36% on the decoding outcome. This indicates that the set of three measures collected in October provides valuable information that can validity predict student performance at the end of kindergarten.

At the December measurement, a majority of variance was explained on both outcomes (i.e., 79.0% on phonemic outcome and 55.0% on decoding) and a parsimonious set of three predictors emerged. The combined phonemic/alphabetic task that 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 corroborates Ehri and McCormick's (1998) model and the importance of progress-monitoring 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 the December measurement for the phonemic outcome, as was the task requiring identification of the last sound. It is not surprising that two phonemic tasks were able to predict the phonemic outcome. As in October, the amount of variance explained on both outcomes may give educators valuable information in making educational decisions.

In January, at the third progress monitoring point, the task that predicted both phonemic and decoding outcomes was the ability to associate a sound with the letter. This skill is essential to word reading and related to the partial-alphabetic stage of the Ehri and McCormick model. Further, the ability to isolate the first sounds of words presented orally predicted the phonemic outcome and was nearly significant (p = 0.066) for the decoding outcome. Whole word segmentation was the most difficult subtest measured in January and was not a significant predictor, which was unexpected. However, a caveat to this finding is that the path coefficient on the phonemic factor was the second highest beta weight of all predictors. Given that this skill was newly introduced at the time of measurement, performance was highly variable, 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. As with the December measurement, nearly 80% of the variance could be explained on the phonemic factor, and 57.7% was explained on the decoding factor.

The final measurement, administered in February, explained 86.8% of variance on the phonemic outcome and 65.2% on the decoding outcomes. Whole word segmentation predicted both phonemic and decoding outcomes. This follows the general pattern of findings across all four models that tasks that are more difficult became stronger predictors over time concurrent with reading development. The pattern also holds that predictive validity of CEMs strengthens over time, as indicated by higher amounts of variance explained. The ability to isolate first sound of a word presented orally was a significant predictor at each measurement occasion.

In summary, the author was interested in whether a parsimonious set of tasks from a larger predictor set could validly predict multivariate latent end-of-kindergarten outcomes and whether those skills changed over time. Table 3 summarizes the predictors' beta weights and statistical significance at each measurement point. One predictor per time point was able to predict both outcomes and can be considered the most parsimonious. In October, it was knowledge of letter names. In December, it was identifying the first sound using a tile. By January, it was knowledge of letter sounds and in February, it was segmenting whole words. When including indicators that predicted one of the outcomes, parsimony was also achieved. In October, the First Sound subtest predicted the phonemic outcome. The First Sound and Last Sound subtests predicted the phonemic outcome in December. In January, the First Sound subtest predicted the phonemic outcome and was nearly significant for the decoding outcome. For the February measurement, the First Sound and First Sound with Tile tasks predicted the phonemic outcome and the Letter Name and Letter Sound tasks predicted the decoding outcome.

Overall, findings demonstrate the CEMs can serve as valid predictors of end-of-kindergarten outcomes. Moreover, findings illustrate the changing topography of progress-monitoring measures in kindergarten. The ability of CEMs to explain over 60% of the variance on the phonemic factor in October and a majority of variance on both outcomes from December forward provides evidence that CEMs can provide valuable and valid data to inform data-based decisions. The predictive validity of individual subtests changed throughout the year, which suggests more attention may be warranted to different skills, depending on the time in the year the measures are administered, and that fewer measures may be needed to inform instructional decisions.

Table 3 *Standardized Beta Weights for Latent Outcomes by Measurement Point*

	Measurement 1 (End October)		Measurement 2 (Middle December)		Measurement 3 (Middle January)		Measurement 4 (Middle February)	
Predictors	Phonemic	Decoding	Phonemic	Decoding	Phonemic	Decoding	Phonemic	Decoding
R^2	62.0%	36.3%	79.0%	55.0%	79.6%	57.7%	86.8%	65.2%
Standardized Beta Weight								
Letter Name	.52*	.35*	.07	.19	-.06	-.12	.09	.16*
Letter Sound	-.13	.10	.11	-.04	.24*	.53*	.10	.38*
First Sound	.50*	.22	.21*	.06	.27*	.20†	.23*	-.04
Last Sound			.21*	.10	.11	.12	.02	.13
Whole Word Segmentation					.27	-.04	.51*	.35*
First Sound with Tile			.41*	.54*	.23	.21	.25*	.11
Last Sound with Tile			.00	-.03			-.10	-.03

Note. * $p < 0.05$. $^{\dagger}p < .10$

Limitations and Future Research Directions

There are several limitations to this study. First, the study involved a small sample, which limits power and the ability to find statistically significant results. There were several coefficients that approached significance that may have been statistically significant had the sample size been larger. Another limitation is that 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. The field should also research 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. Finally, the CEMs in this study were not fluency based. Examination of CEMs that include fluency measures are needed.

Implications and Conclusion

Results of this study provide preliminary evidence supporting the predictive validity of CEMs for kindergarten students at risk for reading difficulties receiving Tier 2 intervention. Early intervention is critical and having reliable and predictive progress-monitoring measures can enable teachers to make data-informed decisions that potentially lead to better student outcomes. The ability to explain relatively large amounts of variance on broad reading outcomes by October and a majority of variance by December allows for early adjustments to be made that could increase the trajectory of response toward better reading outcomes. In addition to predictive validity, knowing the changing nature of individual predictors can inform teachers regarding when and which individual skills are most important to monitor, which may enable researchers and teachers to develop a more parsimonious set of measures that changes throughout the year based on reading development. This increased parsimony could reduce burdens on often-limited school resources.

Successful RTI models rely on using available data to make instructional adjustments, and having viable CEMs can give teachers more confidence in 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 in the allocation of resources to those who need them most.

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF KINDERGARTEN CURRICULUM-EMBEDDED MEASURES FOR END-OF-FIRST AND END-OF-SECOND GRADE OUTCOMES

In 2004, congress reauthorized the Individual with Disabilities Education Act, which, in addition to special education placement, authorized schools to use response-to-intervention (RTI) methods for preventing academic difficulties. The most common RTI approach involves tiered levels of instruction of increasing intensity to best match instruction to student needs. By doing so, a key purpose of RTI is to provide prevention-oriented intervention for students identified at risk for reading difficulties (Lembke et al., 2010).

To maximize the potential of RTI as a preventative practice, intervention must be linked to and informed by formative assessments of student performance (Gersten & Dimino, 2006; Lembke et al., 2010). Measuring student performance to inform instruction is typically accomplished through progress monitoring. Depending on a student's instructional level and needs (i.e., Tier 1, Tier 2 or Tier 3), progress may be monitored as frequently as weekly or as little as three times a year (Gersten & Dimino, 2006; Jenkins, Graff, & Miglioretti, 2009; Mellard et al., 2009). These data can then be used to make instructional decisions that better meet the needs of particular students. Prior research has demonstrated that the information provided by progress monitoring can be used by teachers to make modifications to instruction, which leads to improved student outcomes (Stecker et al., 2005).

The *Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades* practice guide (hereafter used as "RTI practice guide") issued by the Institute of Educational Sciences recommends the use of curriculum-embedded measures (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; rather, most have focused on general outcome measures such as curriculum-based measures (CBMs). While CBMs serve an important role in progress monitoring and as an RTI framework, they do not provide detailed information about direct mastery of skills taught and progress within a particular curriculum. Because CEMs are directly derived from the material taught, they provide more specific information regarding a student's response to intervention and corresponding instructional needs. Student performance data on CEMs can be used as the basis for instructional modifications (e.g., curricular adjustments, instructional pacing, student regrouping, amount of intervention time) to better ensure prevention and mitigation of reading difficulties for students identified as at risk for reading problems.

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). However, few studies examined the utility of CEMs for monitoring kindergarten reading progress; even fewer focused on children receiving Tier 2 intervention or assessed longitudinal 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. By making adjustments and intervening early, the "wait to fail" problem in special education can be avoided (Compton et al., 2006).

Longitudinal studies are needed to identify methods and tools that accurately identify reading risk and provide an empirical basis for intervention (McCardle et al., 2001). 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 focus 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 for reading difficulties. Clearly, establishing early indicators of future reading outcomes is a critical important feature of an RTI approach. Following is a brief summary of the extant literature on longitudinal predictors.

Research Review of Kindergarten Predictors

Predictors of End-of-First Grade Outcomes

Several researchers have focused on a variety of early reading predictors gathered in kindergarten on end-of-first grade outcomes. In their study of kindergarten reading predictors, Scanlon and Vellutino (1996) assessed approximately 1,400 children in kindergarten and again in first grade. During the first half of the kindergarten year, students were administered 25 different measures from six broad sets (i.e., linguistic, math, memory, conceptual development, executive function and reading). Midway through first grade, children were administered the word identification (WI) subtest from the Woodcock Reading Mastery Tests-Revised (WRMT-R; Woodcock, 1987). The authors analyzed each set of predictor variables independently. From the reading set of kindergarten predictors, the ability to name letters in kindergarten was the strongest predictor of first-grade reading skill; letter naming alone accounted for 35.2% of the variance explained on WI. From the linguistic set of kindergarten measures, phoneme segmentation was the best predictor, accounting for 18.5% of the variance compared to 4.6% explained by the second best predictor in linguistic set. The authors concluded that it is possible to identify students at risk for developing reading difficulty early in their kindergarten year.

Schatschneider et al. (2004) found that kindergarten measures of phonological awareness and letter name and letter-sound knowledge predicted longitudinal reading outcomes among 540 students who were assessed four times between October and April of the kindergarten year. Findings indicated that phonological awareness, letter sounds, and rapid automatized naming (RAN) of letters were the best predictors across different outcomes (i.e., letter-word identification, passage comprehension, and word reading efficiency) measured at different times (i.e., end of first and second grades). Letter naming measured at the beginning of kindergarten, in addition to phonological awareness, letter sounds, and RAN, was also a significant predictor of reading outcomes. There was the tendency for measures of letter naming to explain more variance than measures of letter sounds. The authors found that assessments given at the beginning of kindergarten were nearly as predictive of first-grade outcomes as those given at the end of kindergarten.

Chiappe, Siegel, and Wade-Woolley (2002) examined the predictive validity of measures administered in October and November of kindergarten on outcome measures administered in March and April of first grade. In their sample of 727 native English speakers, they found that letter identification and phonological processing explained 30.5% of the variance on a word identification task. Although the two measures had overlapping explained variance, both explained a significant amount of unique variance with letter identification explaining more than phonological processing. It should be noted that phonological processing was measured by six different measures and forced into the regression analysis as a single step. The study indicated that statistically significantly variance on end-of-first grade measures could be explained by measures administered early in kindergarten.

Morris, Bloodgood, Lomax, and Perney (2003) examined 102 kindergarten students who were assessed in September, February, and May of the kindergarten year and again in October and May of first grade. The assessments included an alphabet knowledge task where children were asked to name 15 letters. A beginning consonant awareness measure included two tasks; one required the student to provide the first consonant sound in a word presented orally and the second required the student to identify pictures that started with the same consonant as a target word. A phoneme segmentation task required the student to correctly segment at the phoneme level and use blocks to indicate how many phonemes were contained in the word. Additional tasks included concepts of words in text, spelling with beginning and ending consonants, word recognition, and contextual reading. The authors used structural equation modeling and validated a developmental sequence of reading. They confirmed a model where all paths were statistically significant. Their results indicated that in the kindergarten year, alphabet knowledge preceded beginning consonant awareness, followed by concurrent measures of concepts of words in print and spelling with beginning and ending consonants, which predicted phonemic segmentation abilities. Phonemic segmentation abilities at the end of kindergarten then predicted word recognition ability, which subsequently predicted reading in context. The study highlighted the developmental process of reading and the changing predictive validity of measures over time.

The studies examining predictors of end-of-first grade outcomes have indicated several useful predictors. In particular, letter naming/identification and phonological processing (e.g., phoneme segmentation and phonological awareness) have demonstrated predictive validity. Although limited, research also indicates that the predictive power of indicators changes over time and follows the developmental process of reading.

Predictors of End-of-Second Grade Outcomes

In addition to examining the predictive validity of kindergarten measures for end-of-first grade outcomes, researchers have examined predictors of end-of-second grade outcomes. Catts, Fey, Zhang, and Toblin (1999) investigated the predictive validity of measures administered in kindergarten when predicting second-grade outcomes of 604 students. The students were assessed on phonological awareness through a phoneme deletion task, a rapid naming task, and measures of oral language. Phonological awareness explained unique variances on both reading comprehension and word identification measures administered in the second grade, as did rapid naming and oral language.

Hogan, Catts, and Little (2005) examined the amount of variance kindergarten measures could explain on word reading outcomes in second grade with a sample of 570 students. The students were administered a syllable/phoneme deletion task to assess their phonological awareness and a letter identification task in kindergarten. The students were assessed again in the second grade with measures of word reading and word decoding as outcome measures. Phonological awareness and letter identification predicted outcomes on the word reading measure.

Pennington and Lefly (2001) examined 124 children classified as either high- or low-risk for developing a reading disability from prekindergarten to the end of second grade. They tested whether there were differences between the groups in how well five variables predicted outcomes. The predictors included four phonological components comprising 17 separate tasks and a measure of letter name knowledge. The four outcomes were letter-word identification, nonword reading, a spelling task and a measure of reading speed and accuracy. For the low-risk (LR) group, they found that phonological awareness was the best predictor of second-grade outcomes when taken at prekindergarten, kindergarten, and first grades. For the high-risk (HR) group, they found that the predictive validity of variables changed over time, with letter name knowledge being the best predictor in prekindergarten 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 "the HR group, therefore, underwent a developmental shift by Time 3 that had mostly occurred by Time 1 in the LR group" (p. 828). In other words, the predictive validity of measures likely changed and followed a developmental trajectory of reading for both groups but the high-risk group lagged behind the low-risk group.

Kirby, Pfeiffer, and Parrila (2003) followed 161 students from kindergarten until fifth grade. They administered four measures of phonological awareness and two measures of naming speed during the kindergarten year. The phonological measures included a sound isolation task requiring the student to identify first, last, and middle sounds of words. There was also a phoneme elision measure and two blending measures; one requiring blending at the onset and rime level and another blending at the phoneme level. The naming speed tasks required the student to name colors and pictures. The predictors were factor analyzed and two factors emerged: a phonological awareness factor composed of the four phonological awareness tasks and another factor derived from the two naming speed tasks. The outcome measures were the WI, word attack (WA), and passage comprehension (PC) subtests from the Woodcock Reading Mastery Tests – Revised/Normative Update (WRMT-R/NU; Woodcock, 1998).

The pattern of findings for phonological awareness across the multiple outcomes was fairly consistent. Phonological awareness was a significant predictor on both the WI and WA measures in the first and second grades; it was a significant predictor for PC in first grade (Kirby et al., 2003). Additionally, the strength of prediction decreased as the time between prediction and outcome assessments increased. Naming speed was a significant predictor for PC and WI from kindergarten to fifth grade; it was a significant predictor for WA in third, fourth, and fifth grades. The patterns were opposite for the two latent predictors, with phonological awareness becoming a weaker predictor as time passed and naming speed becoming a stronger predictor. The Kirby et al. (2003) study reiterates previous findings that valid indicators of reading can be measured in kindergarten and that the role and strength of predictors change over time.

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 can validly predict a range of reading outcomes. While the studies reviewed all used some measure of phonemic processing, there was great variability in both the number and type of tasks used. The predictor measures in the studies employed as few as one task and as many as seventeen that varied considerably (e.g., phoneme segmentation, phoneme blending, phoneme deletion, phoneme identification, etc.). Most formed a composite predictor comprised of multiple phonemic related skills.

Additionally, there is evidence suggesting 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. There is some evidence that prediction follows the developmental pattern of reading. Easier skills such as producing letter names and sounds are early predictors, while more difficult phonemic processing skills (e.g., phoneme segmentation, blending, or elision) become more predictive over time.

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 the measurement of those skills. As word reading develops, so should the measures of pre-reading skills in alignment with instruction (Ehri & McCormick, 1998). However, the authors review identified only one study that examined the changing strength of predictors; Schatschneider et al. (2004) found that the predictive validity of early reading measures changed from beginning of kindergarten to end of kindergarten for outcomes measured in first and second grades. In addition to examining the differential predictive power between the beginning and end of kindergarten, research needs to examine if 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 author could not find a single study that examined the longitudinal predictive validity of progress-monitoring measures, including CEMs. In addition, the author review did not identify any studies examining children in Tier 2 intervention. While Pennington and Lefly (2001) found that there was a difference in predictors for high- and low-risk students, their study did not examine children in Tier 2 intervention. Students in Tier 2 intervention are receiving that type of intervention because of their heightened risk status. Therefore, predictive studies for this population are important because few studies have examined these students as a special population or 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 by 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, which 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 because it encompasses several distinct skills (McCardle et al., 2001; Speece et al., 2003).

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

The purpose of this study was to evaluate CEMs administered throughout kindergarten to children receiving Tier 2 reading intervention and to investigate their utility for predicting first- and second-grade reading outcomes. The study makes a unique contribution to the literature by focusing on CEMs of specific skills and examining their changing role for predicting longitudinal outcomes. It was hypothesized that the role of predictors would change over time according to reading development, with easier skills being predictive early and more difficult skills predicting outcomes as the kindergarten year progressed. It was also hypothesized that the amount of variance explained would increase throughout the kindergarten year. The following research question was addressed:

1. Which specific early reading skills measured by CEMs multiple times throughout the kindergarten year are predictive of comprehensive longitudinal reading outcomes, and 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 3-5 for roughly 30 minutes a day, five days a week. Data for this study are from three experimental studies comparing the effects of early reading interventions that used progress-monitoring data to inform instructional decisions.

Students in the current study were children identified in kindergarten as being at risk for developing reading difficulties. Student data from three cohorts spanning three 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.

Setting and Participants

Schools and interventionists. A total of 23 schools in three states participated in this study: 11 schools were in Florida, 8 in Connecticut and 4 in Texas. The range of students qualifying for free or reduced-cost lunches in these schools was 72% to 81% in Texas, 3% to 81% in Connecticut, and 31 to 81% in Florida. Students were nested under 70 interventionists with 18 interventionists in Texas, 23 in Connecticut and 29 in Florida. Two of the interventionists were male and 68 were female. One interventionist had an associate's degree, one a doctorate, three a high school diploma, five an educational specialist, 23 held a bachelor's degree, and 36 held a master's degree. Three interventionists were African-American, six were Hispanic, and 61 were white. The average teaching experience was 11.03 years (*SD* = 9.14). Sixty-one interventionists were certified teachers and nine were paraprofessionals. Interventionists were given two days of professional development training covering the ERI materials and implementation. The first day focused on parts one and two of the intervention and took place before the intervention started. The second day of training occurred following the conclusion of the second part of the intervention and covered parts three and four.

Students. Students included 299 kindergarteners from three cohorts across three years. In each cohort, students were determined to be at risk for 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 where further screened by the researchers to determine eligibility for participation. In each cohort, students were first administered the letter naming fluency (LNF) subtest from DIBELS (Good & Kaminski, 2002) and sound matching (SM) subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999).

In the first two cohorts, students who had a raw score of 6 or below on LNF and 37^{th} percentile or below on SM qualified for participation. An additional requirement for students in the third cohort was a standard score of 7 or below on the rapid object naming (RON) subtest from the CTOPP or a standard score of 80 or below on the letter identification (Letter ID) subtest of the Woodcock Reading Mastery Tests-Revised/Normative Update (WRMT-R/NU; Woodcock, 1998). Differences among states and cohorts on pretest 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 represents an attrition rate of 14%. The attrition rate from end of kindergarten to the end of first grade was 16%, and from the end of first grade to the end of second grade was 45%. However, the loss of 45% of 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 grade 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 4 summarizes student demographics.

Assessment Procedures

All participating students were administered four pretests (i.e., LNF, SM, RON, Letter ID) prior to the beginning of the intervention and roughly six weeks into their kindergarten school year. Students were removed from their classrooms and tested one-on-one by trained assessors who were members of the research team. Assessors received a minimum of 8 hours 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 two weeks of the end of intervention.

Predictor variables. Predictors included CEMs measured approximately every eight weeks with measurement 1 occurring in the beginning of January, measurement 2 in middle of March, and measurement 3 in end of April. The measurements were given following the first three of four curriculum parts.

Table 4 *Student Demographics*

Variable	Participants (N = 299) N	(%)
Gender		
Male	167	44.1
Female	143	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	Mean (*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)	
LNF[c]	1.20 (1.77)	

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

The last CEM following the fourth curriculum part was not used as instruction on that part was not completed prior to the finish of the intervention and posttesting.

The first CEM assessed material covered in the first 42 lessons. There was a total of six subtests comprising phonemic, alphabetic, and integrated tasks. The two phonemic subtests (First Sounds and Last Sounds) required a student to provide the first and last sounds of words presented orally. The student was presented with a picture and the examiner spoke the word represented by the picture and then asked the student to provide the first and last sound 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 appropriate boxes.

The second CEM 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 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 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 medial 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 5 summarizes the CEMs by administration point and reports their reliability coefficients in the sample.

Table 5 *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 Alpha):*	.89	.87	.90

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 WA subtest from the WRMT-R/NU and the NFW test from DIBELS were used to measure decoding of nonsense/pseudowords. On the NWF test, a student has 1 minute to orally produce as many nonsense words or segments as possible. Alternate form reliability when measured one month apart is 0.83 (Good et al., 2004). On WA, students decode pseudowords. Unlike NWF, it is not fluency based. The split-half reliability of WA is reported as 0.87 in the technical manual. The WI subtest from the WRMT-R/NU assessed sight words and also words used infrequently in the English language. Word identification is not fluency based and has a median split-half reliability of 0.97 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 minute. Alternate form reliability ranges from 0.89 to 0.94 as reported by Tindal, Marston, and Deno (1983). The 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 0.92 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. It is norm-referenced and Cronbach's alpha as reported in the technical manual is 0.87 for six-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 adjusts standard errors by accounting for non-independent 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 white as the reference group.

An exploratory factor analysis (EFA) 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 outcome 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) in order to estimate model fit. Once a measurement model had been confirmed using CFA, a total of six structural models were estimated. The six models were composed of one at each time point in kindergarten (three time points) predicting outcomes at the end of first and second grades.

Results

End of First Grade

The EFA conducted for the end-of-first grade scores indicated a single factor composed of all outcome variables. Analysis of Eigenvalues revealed one Eigenvalue with a value 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 of χ^2 (8) = 16.32, p = .038 was statistically significant. However, fit indices indicated acceptable model fit with RMSEA = 0.06, CFI = 0.99 and SRMR = 0.01. Figure 6 shows the measurement model including standardized path coefficients, which were all positively associated with the factor and statistically significant ($p < 0.01$). The R^2 for the outcome variables, which measured the variance explained, ranged from 65.2% to 91.6%.

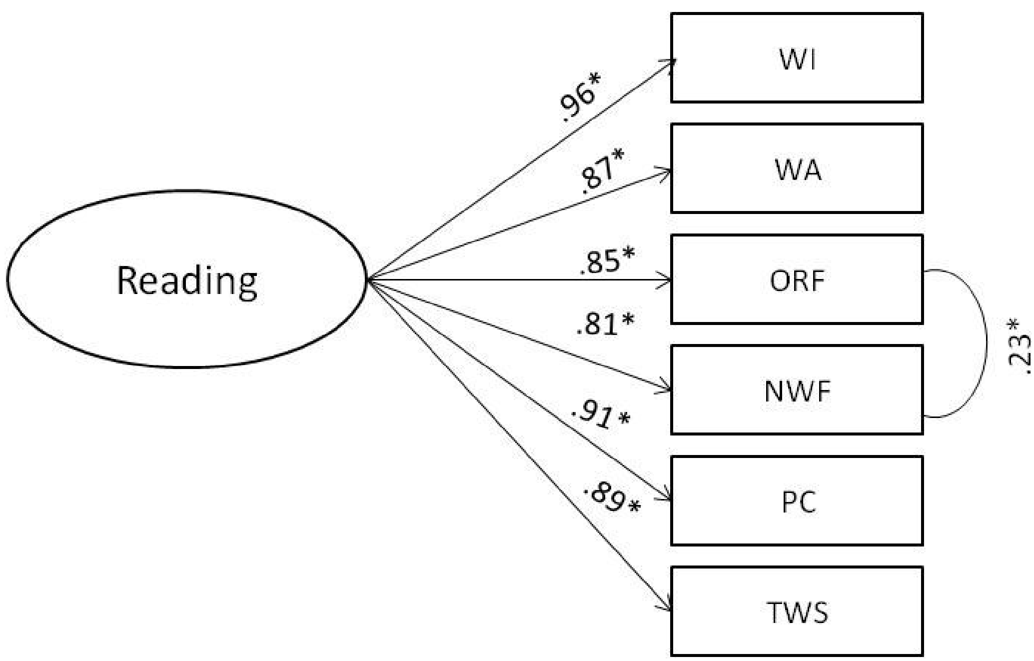

Figure 6. End-of-first grade measurement model. * $p < 0.05$.

The first model (see Figure 7) used predictors from the first CEM (collected early January) to predict outcomes measured at the end of first grade. The chi-square test value of χ^2 (68) = 136.40, $p < 0.001$ was statistically significant; however, model fit indices indicate adequate fit (RMSEA = 0.06, CFI = 0.97, SRMR = 0.03). There were two statistically significant predictors and one near-significant predictor. The Last Sounds ($\gamma = 0.31$, $p < 0.000$) and First Letter-Sound ($\gamma = 0.26$, $p = 0.003$) were statistically and positively related to the reading outcome. The Letter Names test ($\gamma = .21$, $p = 0.063$) was nearly significant. A total of 54.3% of the variance could be explained on the latent reading outcome factor.

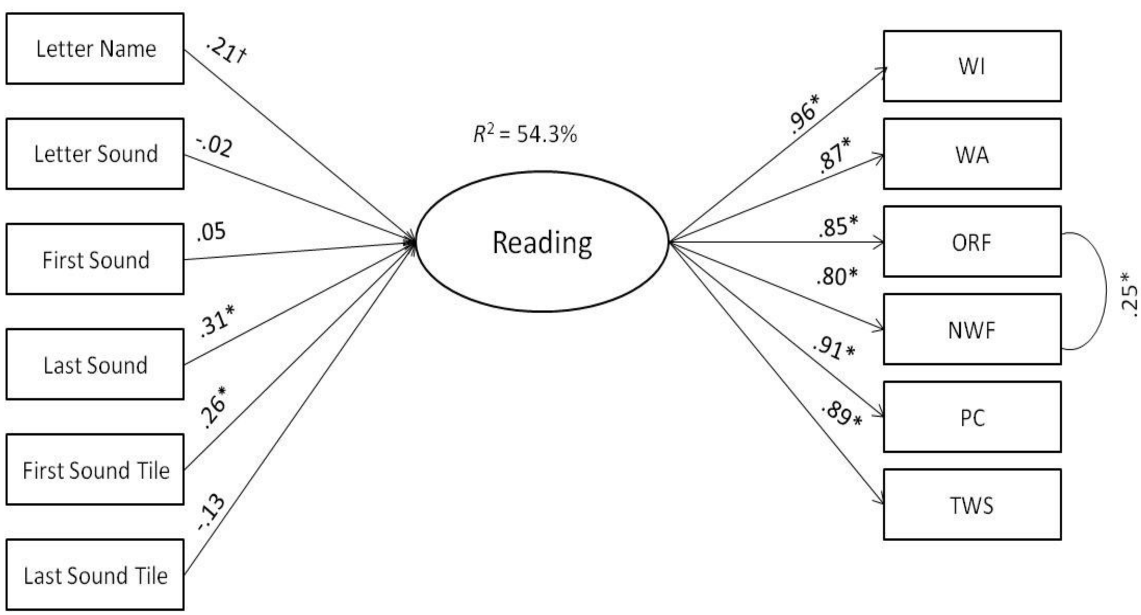

Figure 7. Structural equation model for curriculum-embedded measure 1 on end-of-first outcomes. * $p < 0.05$. † $p < 0.10$.

The second model (see Figure 8) used the subtests from the second CEM (collected mid-March) as predictors. The chi-square test was statistically significant with χ^2 (73) = 137.02, $p < 0.001$). Model fit indices indicated good fit with RMSEA = 0.05, CFI = 0.97 and SRMS = 0.02. There were two statistically significant predictors of the latent reading factor. Letter Sounds ($\gamma = 0.21$, $p = 0.001$) and Whole Word Segmentation ($\gamma = 0.20$, $p = 0.014$) 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.

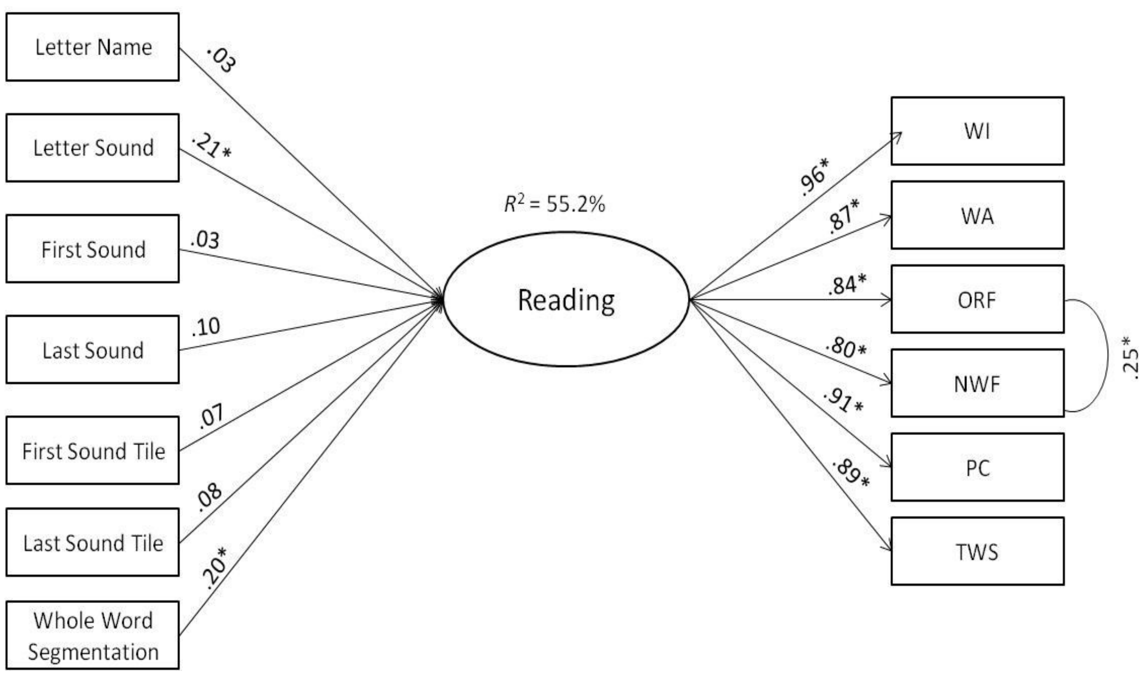

Figure 8. Structural equation model for curriculum-embedded measure 2 on end-of-first outcomes. * $p < 0.05$.

The final structural model for first grade was calculated using the CEM from late April (see Figure 9). The results of the chi-square test indicated a statistically significant finding with χ^2 (78) = 133.86, $p < 0.001$. Fit indices indicated good model fit with RMSEA = 0.05, CFI = 0.97 and SRMR = 0.02. The Letter Names subtest ($\gamma = 0.16$, $p = 0.010$) was statistically significant and positively related to the latent reading outcome. The Word Reading subtest ($\gamma = 0.32$, $p < 0.00$) was also positively related to the outcome and statistically significant while the Medial Sounds subtest ($\gamma = 0.15$, $p = 0.054$) was nearly significant. In total, the predictors were able to explain 62.9% of the variance.

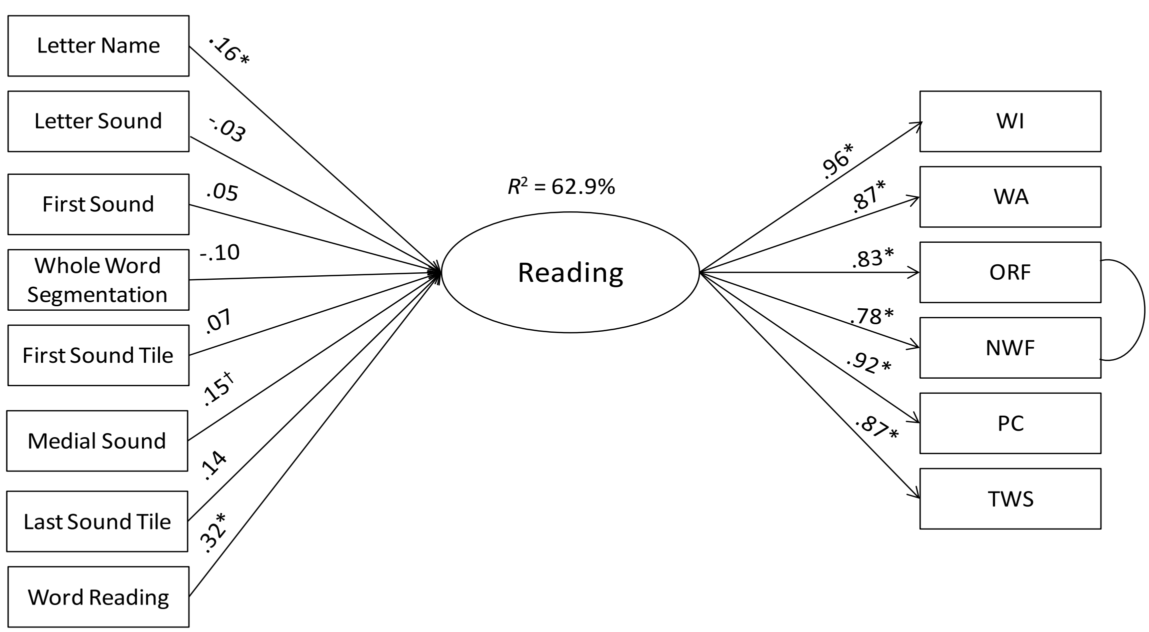

Figure 9. Structural equation model for curriculum-embedded measure 3 on end-of-first outcomes. * $p < 0.05$. $^{\dagger}p < 0.10$.

End of Second Grade

An EFA for end-of-second grade reading outcomes was conducted to examine the factor 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 using a CFA analysis. The model fit for the measurement model was good (RMSEA = 0.03, CFI = 1.00, SRMR = 0.04) with a non-significant chi-square value of 11.80, $p = 0.299$. 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 7-9 are the same for the end-of-second grade model. The difference in the endogenous variables can be seen in the end-of-second grade measurement model in Figure 10.

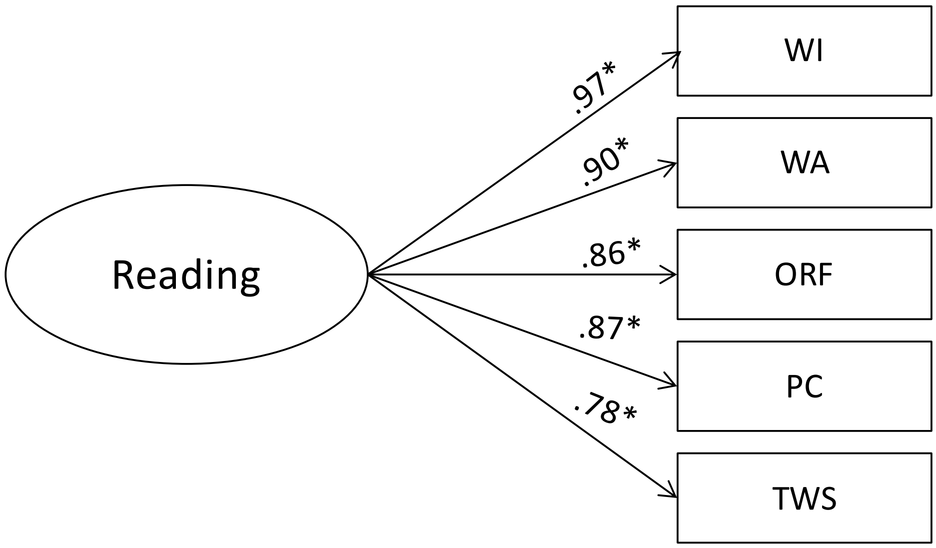

Figure 10. End-of-second grade measurement model. * $p < 0.05$.

Results from the first CEM subtests predicting end-of-second grade outcomes revealed a statistically significant chi-square value of χ^2 (53) = 96.53, $p < 0.001$ with fit indices indicating adequate model fit (RMSEA = 0.05, CFI = 0.94, SRMR = 0.03). While none of the predictors were statistically significant, the First Sounds subtest ($\gamma = 0.15$, $p = 0.060$) was nearly significant 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 of χ^2 (57) = 95.04, $p = 0.001$. Fit indices indicated good model fit with RMSEA = 0.05, CFI = 0.95 and SRMR = 0.03. 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 of χ^2 (61) = 111.07,

$p < 0.001$. The Letter Sounds subtest was a statistically significant predictor ($\gamma = .21$, $p = 0.045$) 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

This study evaluated CEM subtests administered throughout the kindergarten year and examined their utility for predicting reading outcomes at the end of first and second grades. The author was interested in which individual skills were statistically significant predictors of future reading outcomes, when were they able to predict, and whether their predictive power changed over time following reading development. Previous studies and Ehri and McCormick's (1998) model suggested the role of predictors would change over time according to reading development, with easier skills being predictive. The author used structural equation modeling to evaluate the effectiveness and parsimony of CEM subtests in predicting latent reading outcomes for the end of first and second grades.

Predicting First-Grade Outcomes

Findings indicated that end-of-first grade outcomes for children at risk for reading problems can be predicted by January of the kindergarten year. Specifically, the ability to produce the last sound of a word presented orally (phonemic task) predicted the reading latent outcome variable. Prior studies identified phonological processing tasks taken 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).

In those studies, the phonological processing predictor was composed of at least six individual tasks, whereas the present study isolated one phonological processing predictor. A task that integrated alphabetic and phonemic knowledge (i.e., correctly providing the tile representing the first sound in a word presented orally) was also a statistically significant predictor of first-grade reading outcomes. Finally, knowledge of letter names in kindergarten approached statistical significance ($p = 0.063$); letter naming is supported 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 mid-year kindergarten CEMs can provide substantial amounts of information vital to making informed instructional decisions.

The CEM administered in March 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 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 mid-March CEMs was whole word segmentation, which was the most difficult task measured. 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 is predictive following reading development and easier tasks (e.g., letter naming or letter identification).

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 (1993) model of reading. In addition, 62.9% of the variance was explained by the April CEM, which is higher than previous studies have found.

The findings for the end-of-first outcomes support the first hypothesis of this study that predictors would change over time, which aligns with earlier findings (Morris, Bloodgood, Lomax, & Perney, 2003). In addition, the predictors that achieved 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

There was only one statistically significant predictor of end-of-second grade reading outcomes – knowledge of letter sounds. Catts et al. (1999) and Hogan et al. (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 ($p = 0.06$), which also aligns with these prior studies.

For the April CEM administration, the integrated Medial Sounds subtest approached statistical significance ($p = 0.08$). However, findings are insufficient to support the hypothesis that predictors 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 supports the second hypothesis.

Limitations and Future Directions

Findings from this study must be considered in light of several limitations. First is sample attrition 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. That limitation noted, the maximum likelihood estimation method used in Mplus handled missing data very well, so having the full sample data at the end of second may have not changed the results.

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 as well as other sources of curriculum-derived measures (e.g., teacher-made tests). Finally, the author has 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, dosage) differed across children and grade levels. This introduces error variance that may have reduced the predictors' power as the time difference between collecting the predictors and the outcomes they were predicting grew.

Implications of the Study

This study examined the predictive validity and changing roles of measures embedded in an early reading curriculum gathered three times in the kindergarten year. For end-of-first grade outcomes, predictors (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) emerged at each of the three measurements that explained a majority of the variance on a broad latent reading outcome. Additionally, the predictors changed over time in alignment with the Ehri and McCormick (1998) model of reading development. The end-of-second grade findings are less clear; only one predictor reached statistical significance. However, when near-significant results are considered, the general trend for end-of-second grade outcomes also aligned with their model, with more difficult predictors becoming more salient throughout the kindergarten year.

School resources are often limited and having viable, parsimonious tools that are reliable and have predictive validity could improve the cost-to-benefit ratio. The ability to predict longitudinal outcomes is important, especially for children who enter kindergarten at risk for developing reading difficulties. Knowing whether a student is likely to continue to be at risk in first grade could enable teachers to make instructional decisions early in kindergarten, which might in turn lead to improved outcomes that could mitigate problems that become more intractable over time. Additionally, it is important for teachers to understand the changing role of predictors, as this provides information regarding which and when indicators serve as optimal predictors of future reading performance.

CONCLUSION

Within the context of RTI, measures that can be used to inform educational decisions for children in Tier 2 intervention are vitally important. The ability to predict reading outcomes allows teachers to make data-driven adjustments to instruction that may alter the course of students at-risk for reading problems. Curriculum-embedded measures have the potential to predict outcomes and serve as a basis for intervention modifications that better match student needs.

Gersten et al. (2009) recommended using CEMs to monitor student performance and as the basis for modifying instruction; however, little research exists to support this recommendation. Other studies have also highlighted the lack of valid early measures for predicting early-reading outcomes (McCardle et al., 2001; Speece et al., 2003; Torgesen, 1998). The purpose of the research discussed in this dissertation was to evaluate the predictive validity of CEMs for later reading outcomes and examine whether the role of CEMs changed throughout the kindergarten year using a structural equation modeling framework. To achieve the goal, two studies were conducted using CEMs gathered at multiple times in the kindergarten year to predict later reading outcomes for students receiving Tier 2 intervention.

Summary of Study 1 Findings

Study 1 focused on the ability of CEMs administered four times during the kindergarten year to predict end-of-kindergarten outcomes. The CEMs' subtests probed multiple pre-reading skills that changed throughout the year to reflect reading development. The end-of-kindergarten reading outcomes were two latent factors (i.e., phonemic and decoding) composed of multiple early reading skills that reflected a broad view of reading. The study evaluated whether a parsimonious set of early reading indicators could predict reading outcomes and whether the predictive validity changed throughout the year in alignment with a prominent model of reading development (Ehri & McCormick, 1998). The following research questions were addressed:

1. What skill-specific progress-monitoring measures embedded in the curriculum and administered during the kindergarten year are most predictive of broad end-of-year outcomes for children receiving Tier 2 supplementary instruction?
2. At what time points can skill-specific CEMs be predictive of end-of-year outcomes and does the predictive validity of skill-specific CEMs change following a developmental progression?

At the first measurement point in October, a majority (62.0%) of the variance in phonemic outcomes was explained, and 36.3% of the variance in decoding outcomes was explained. This finding suggests that CEMs administered during the first half of the kindergarten year can be viable predictors of end-of-year outcomes. Additionally, a parsimonious set of predictors was supported in that the ability to name letters predicted both outcomes and knowledge of the first sounds in words predicted the phonemic outcome. These findings indicate that early in the kindergarten year, CEMs provide valuable information that can be used to inform educational decisions.

At the second measurement point in December, the CEMs explained a majority of variance on both latent outcomes (79.0% of phonemic outcomes and 55.0% of decoding outcomes). As in October, the ability to identify the first sounds in words was predictive of end-of-kindergarten phonemic outcomes. Knowledge of the last sounds in words also predicted phonemic outcomes. A student's ability to place the appropriate letter tile to form the initial sound of a word predicted both phonemic and decoding outcomes. This task was the most difficult of the second CEM subtests; that fact that that it reached statistical significance at the second measurement time point provides evidence that predictors of reading outcomes change as reading develops.

For the CEM administered in January, 79.6% of the variance on the latent phonemic outcome variable was explained and 57.7% of the decoding outcome was explained. Letter-sound knowledge predicted both phonemic and decoding outcomes, and identifying the first sounds in words predicted the phonemic outcome. Although the most difficult CEM subtest administered at this measurement point (i.e., whole word segmentation) was not statistically significant, it had the second-highest beta weight. The wide range in student performance, which resulted in a large amount of variance, likely prevented this finding from being statistically significant.

The final CEM administered in February explained 86.8% and 65.2% on the phonemic and decoding outcomes, respectively. Whole word segmentation was a statistically significant predictor for both latent outcomes while knowledge of first sounds in word and identifying first sounds of words with tiles predicted phonemic outcomes. Letter name identification and letter sound knowledge predicted decoding outcomes.

Results from the first study provided evidence that:

1) a majority of variance could be explained on outcome measures as early as October,
2) the amount of explained variance increased over time,
3) the status of statistically significant predictors changed throughout the year according to reading development, and
4) a parsimonious subset of predictors emerged at each measurement occasion.

These findings have important implications for the field because they demonstrate the utility of CEMs for predicting early reading outcomes and their usefulness in an RTI framework, which focuses on monitoring student progress and using those data to make decisions that improve student outcomes. The statistically significant findings and relatively large amounts of variance explained may give teachers more confidence in decision-making for students in Tier 2 intervention. Additionally, requiring fewer measures, as indicated by the parsimonious subset of indicators, can help reduce costs while improving data on which educational decisions are based.

Summary of Study 2 Findings

The second study focused on the predictive validity of CEMs gathered in kindergarten for children in Tier 2 intervention. CEMs were administered in January, March, and April of kindergarten, and longitudinal outcomes were gathered at the end of first and end of second grades. A single latent variable was regressed on by the CEM subtests. The research question posed for the second study was:

1. Which specific early reading skills measured by CEMs multiple times throughout the kindergarten year are predictive of comprehensive longitudinal reading outcomes and does their predictive validity change throughout the year following the development of reading?

Findings at the end of first grade indicated there was a parsimonious set of statistically significant predictors for each CEM administration in kindergarten. For the CEM gathered in January, knowledge of the last sounds in words and the first sounds in words using tiles predicted a majority (54.3%) of the variance on the latent reading outcome. For the second CEM administered in March, letter-sound knowledge and whole word segmentation predicted a majority of variance (55.2%) on the reading outcome variable. The final CEM administered in April explained 62.9% of the variance on reading outcomes and had two statistically significant predictors – letter-name knowledge and word reading. The pattern of statistically significant predictors on end-of-first grade outcomes follows Ehri and McCormick's (1998) model; as the year progressed, the strongest predictors transitioned from easier tasks (i.e., knowledge of last sounds in words in January) to more difficult tasks (i.e., whole word segmentation in March and word reading in April).

Although there were several near-statistically significant predictors for end-of-second grade outcomes, only one predictor (i.e., letter-sound knowledge in April) reached statistical significance. The ability to identify the first sounds in words from the January CEM approached statistical significance, as did medial-sound knowledge assessed in April. Although there were few statistically significant findings at the end of second grade, the pattern of near significant results follows reading development. Additionally, the amount of variance explained on the end-of-second grade reading outcome increased from 34.1% for measures from January of kindergarten to 40.6% for kindergarten measures collected in April.

Limitations and Future Directions

There are several limitations to be acknowledged in these studies. First, the relatively small sample sizes limit the power to detect statistically significant findings. Particularly, the sample size for Study 2 was small for the end-of-second grade outcomes. Given the pattern of near-significant results, a larger sample size may have produced more statistically significant findings. Second, the CEMs used in both studies were from one intervention; findings cannot be generalized to CEMs from other interventions. Future research should examine the predictive utility of CEMs from other interventions and teacher-designed CEMs. Building a research base that supports CEMs from a variety of interventions may lead to increased use and more informed decision-making. Another limitation of the second study is that little is known about the type and amount of reading intervention that children received during first and second grades. It is likely that some students continued to receive supplementary support in the first and second grades, while others may not have. These intervention differences may have contributed to greater variance in the sample as time progressed, which in turn would have reduced the likelihood of statistically significant findings.

Implications

Both studies in this research examined the predictive validity of CEMs gathered during the kindergarten year for predicting future reading outcomes. For end-of-kindergarten and end-of-first grade outcomes, subsets of statistically significant predictors emerged for each CEM measurement time point. This finding indicates that not only can kindergarten CEM subtests be valid predictors of future reading outcomes, but that a parsimonious set of measures is viable. The amount of variance explained increased across CEM measurement time points, which could provide teachers with better data on which to base decisions as the year progresses, as it provides more information about future reading performance. The roles of kindergarten CEM predictors changed over time for predicting end-of-first grade outcomes. This finding has importation implications for teachers' decision-making, as it allows them to focus on the most appropriate indicators at a given point in time. These findings suggest that CEMs can provide schools with better data for decision-making at a critical time for a vulnerable population and potentially at a reduced burden.

Findings for end-of-second grade outcomes indicated only one statistically significant CEM predictor, although several approached statistical significance. The overall trend of results is similar to the findings for end-of-kindergarten and end-of-first grade; predictors of increasing difficulty explained higher amounts of variance on reading outcomes as time progressed. The evidence, although weaker than it is for end-of-first grade, supports that data from CEMs can help teachers focus on the most salient indicators for decision-making throughout the kindergarten year.

Conclusion

Failure to learn to read early can lead to difficulties throughout a child's education and beyond. Students at risk for reading failure who do not receive sufficient early intervention rarely catch up to their typically performing peers. Informed decision-making is a critical component of RTI, and much of the necessary data are provided by progress-monitoring measures. To ensure that teachers have the most accurate data to make the best educational decisions, valid measures of student progress are required.

The CEMSs examined in this dissertation illustrate the potential of curriculum-embedded progress-monitoring measures. Findings from this research support that the CEMs investigated had good predictive validity, were parsimonious, and were dynamic. Educators often need tools that are accurate and affordable, and CEMs warrant further examination as possible resource to improve student outcomes.

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: 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.

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

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*, 192-208.

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.

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.

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.

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

Fuchs, L. S. (n.d.). *Validated forms of progress monitoring in reading and mathematics.* Retrieved April 28, 2012, from http://www.rtinetwork.org/essential/assessment/progress/validated-forms-progressmonitoring

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18*, 157-171.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188-192.

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

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 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/publications/practiceguides/.

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.

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. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research*

Institute on measuring growth and development. (Tech. Rep. No. 7). Eugene, OR: 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 & Hearing Services in Schools, 36*, 285-293.

Jenkins, J. R., Graff, J. L., & Miglioretti, D. L. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children, 75*, 151-163.

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—A cross-classified multilevel model approach. *Journal of Educational Psychology, 102*, 313-326.

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

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.

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

Marston, D., Pickart, M., Reschly, A., Heistad, D., Muyskens, P., & Tindal, G. (2007). Early literacy measures for improving student reading achievement: Translating research into practice. *Exceptionality, 15*(2), 97-117.

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

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24*, 186-195.

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.

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

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.

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

Ritchey, K. D. (2004). From letter names to word reading: The development of reading in kindergarten. *Reading Research Quarterly, 39*, 374-376.

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 & Developmental Disabilities Research Reviews, 2*, 54-63.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In Bruce K. Shapiro, Pasquale J. Accardo, Arnold J. Capute (Eds.), *Specific reading disability: A review of the spectrum* (pp. 75-119). Timonium, MD: York Press.

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.

Seethhaler, P., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*, 37-59.

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.

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.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement.* (No. 9). Minneapolis, MN: 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.

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.

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.

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

Woodcock, R. W. (1987). *Woodcock reading mastery tests -Revised*. Circle Pines, MN: American Guidance Services.

Woodcock, R. W. (1998). *Woodcock reading mastery test - Revised/Normative update*. Bloomington, MN.: Pearson Assessments.

Woodcock, R. W., & Johnson, M. (1989). *Woodcock-Johnson psychoeducational battery-Revised*. Allen, TX: DLM.

VITA

Eric Lars Oslund received his Bachelor of Science degree in Business Administration from California Polytechnic State University in San Luis Obispo, California, in 2000. He entered the school psychology program at West Texas A&M in Canyon, Texas, in 2004 and earned his Master of Science degree in May 2008. He then worked as a Licensed Specialist in School Psychology and began his doctoral program in educational psychology with an emphasis in special education, graduating from Texas A&M University in College Station, Texas, in August 2012. His research interests include assessment, methodology, and applied statistics.

Dr. Oslund may be reached at Department of Educational Psychology, 704 Harrington Tower, 4225 TAMU, College Station, TX 77843-4225. His email is ericoslund@yahoo.com.