Hammill Institute on Disabilities

Assessment for Effective Intervention
2019, Vol. 44(4) 241–255

DOI: 10.1177/1534508418775805
aei.sagepub.com
SAGE

Predictive Validity of Kindergarten Progress Monitoring Measures Across the School Year: Application of Dominance Analysis

Nathan H. Clemens, PhD[1], Yu-Yu Hsiao, PhD[2], Leslie E. Simmons, MEd[3], Oi-man Kwok, PhD[3], Emily A. Greene, MEd[3], Michelle M. Soohoo, MEd[3], Maria A. Henri, MS[3], Wen Luo, PhD[3], Christopher Prickett, MA[3], Brenna Rivas, PhD[4], and Stephanie Al Otaiba, PhD[4]

Abstract

Although several measures are available for monitoring kindergarten reading progress, little research has directly compared them to determine which are superior in predicting year-end reading skills relative to other measures, and how validity may change across the school year as reading skills develop. A sample of 426 kindergarten students who were considered to be at risk for reading difficulty at the start of kindergarten were monitored across the year with a set of paper-based progress monitoring measures and a computer-adaptive test. Dominance analyses were used to determine the extent to which each measure uniquely predicted year-end reading skills relative to other measures. Although the computer-adaptive test was the most dominant predictor at the start of the year over letter sound fluency, letter naming fluency, and phoneme segmentation fluency, letter sound fluency was most dominant by December. Measures of fluency reading real words administered across the second half of the year were dominant to all other assessments. The implications for measure selection are discussed.

Keywords

early literacy, progress monitoring, curriculum-based measurement, reading

For many students, kindergarten is a time when formal reading instruction begins. In most contemporary kindergarten reading curricula, instruction initially targets the development of basic skills such as phonemic awareness (e.g., segmenting and blending words) and alphabetic knowledge (e.g., identifying letters by name and sound), and gradually transitions across the year to show how those skills are used for decoding words and reading simple texts (e.g., Houghton Mifflin Harcourt, 2014; Pearson, 2011).

Given the variety of early reading skills that children learn in kindergarten, it is important to understand which skills may serve as important indices for monitoring kindergarten reading development. Formative assessment (i.e., progress monitoring) provides teachers with timely feedback on students' learning and responsiveness to instruction, and is particularly important for students who are struggling or at risk for subsequent reading difficulties (Gersten et al., 2009). Although many different forms of assessment can be used for progress monitoring, a popular framework is Curriculum-Based Measurement (CBM), which uses brief skill probes that are indicative of achievement in a broader academic domain (Deno, 1985, 2003). The measurement of oral reading of connected text (i.e., CBM-R) is one of the most widely studied forms of CBM, and research has demonstrated that CBM-R serves as a fairly stable index of overall reading achievement across Grades 2 to 5 (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Reschly, Busch, Betts, Deno, & Long, 2009).

To support intervention efforts with students in earlier grades, such as kindergarten, progress monitoring methodologies have been extended downward to assess foundational and prereading skills. The diversity of skills that represent kindergarten reading achievement is reflected in the variety

[1]The University of Texas at Austin, USA
[2]The University of New Mexico, Albuquerque, USA
[3]Texas A&M University, College Station, USA
[4]Southern Methodist University, Dallas, TX, USA

Corresponding Author:
Nathan Clemens, Department of Special Education, The University of Texas at Austin, 1 University Station, D5300, Austin, TX 78712, USA.
Email: nathan.clemens@austin.utexas.edu

of measures available for monitoring progress. Presently, at least 11 measures exist from different publishers that are designed for monitoring progress in prereading skills including alphabetic knowledge (e.g., letter name or sound fluency, sounds in nonsense words), phonological awareness (e.g., phoneme segmenting, word blending, initial sound identification), decoding (e.g., pseudoword reading), and word identification (for example, word list or sentence reading). The number of available measures increases when considering computer-adaptive tests (CATs) which have been marketed as options for monitoring progress.

The unique and dynamic nature of kindergarten reading instruction, combined with the number of available measures, introduces challenging decisions for kindergarten educators and interventionists to identify the best tools to accurately and efficiently monitor their students' reading growth. Many measures assess similar constructs. Which measures are most predictive of overall, grade-appropriate reading skills in kindergarten? Should measures be changed across the year in accordance with the changing nature of the curriculum, or do particular measures stand out as durable and consistent indicators of overall achievement and better predictors of important year-end reading outcomes?

The current study is part of a larger project that is investigating the technical adequacy of kindergarten progress monitoring measures. To assist educators with decisions on what type of tools to consider for monitoring kindergarten reading progress, this project is investigating several published tools that have been offered by their respective publishers as options for monitoring early reading progress. Using Fuchs's (2004) framework on the three stages of research for establishing the technical adequacy and practical utility of progress monitoring measures, we focused this particular study on "Stage 1," which is concerned with establishing the technical properties of the static score (e.g., the extent to which a score on a progress monitoring measure is associated with achievement in the broader academic domain; Deno, 2003; Fuchs, 2004). Specifically, we investigated the criterion-related predictive validity of the measures by evaluating the relation of scores on the progress monitoring measures administered across the year with reading skills assessed at the end of the year. However, rather than testing measures individually, we directly contrasted the predictive validity of the tools within the same sample of students who were at risk for reading difficulties (i.e., a population who is most often the target of progress monitoring). We used a technique called dominance analysis to compare the relative importance of measures with each other in predicting year-end reading skills.

Dominance Analysis

Dominance analysis (Azen & Budescu, 2003; Budescu, 1993) is a special application of multiple regression that is used to determine the relative importance of individual predictors among models that use all possible combinations of other predictors (i.e., subset models). Via pair-wise comparisons in separate subset regression models, dominance analysis reveals the additional contribution of each predictor relative to other predictors in the same model. Dominance analysis offers advantages to other multiple regression techniques (e.g., hierarchical, stepwise). First, its use of pairwise comparisons, both with and without other predictors in the model, permits the analysis of correlated predictors. Second, unlike other regression techniques that only indicate the statistical significance of a predictor or the proportion of variance it explains, dominance analysis allows for the ranking of variables based on their importance, which allows for determining whether one predictor "dominates" (i.e., has larger additional contribution) over other predictors of the same outcome.

Dominance analysis has been used in several studies that investigated predictors of reading achievement in children and adults. For example, Fuchs, Fuchs, and Compton (2004) observed that students' level of word identification fluency achievement and rate of growth at the beginning of first grade were more important for predicting reading outcomes at year end compared with measures of nonsense word reading fluency (WRF) level and rate of growth. Similarly, Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) utilized dominance analysis to examine the relative importance of phonological awareness, letter sound knowledge, naming speed, and other skills assessed across kindergarten in predicting reading outcomes at the end of first and second grades. Kim, Petscher, Schatschneider, and Foorman (2010) used dominance analysis to determine that growth in oral reading fluency (ORF) across first grade was the most important variable in predicting reading comprehension in later grades, relative to other early literacy and language measures. Dominance analyses have also revealed that skills such as word reading, vocabulary, and auditory working memory are more or less important for adults' ORF depending on an individual's reading level (Mellard, Anthony, & Woods, 2012). Together, the results of these studies illustrate how dominance analysis can help identify the strongest predictors of a complex construct, like reading.

Extending Stage 1 Research on Early Literacy Progress Monitoring Measures

The present study extends the current research base in several ways. First, although data on the predictive validity of early literacy progress monitoring measures are available through evidence submitted by their publishers (National Center on Intensive Intervention, 2018) and peer-reviewed studies (see Goffreda & DiPerna, 2010, for a review; see also Betts, Pickart, & Heistad, 2009; Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009; Elliott, Lee, & Tollefson, 2001; Kamii & Manning, 2005; Ritchey, 2008; Stage, Sheppard, Davidson, & Browning, 2001), additional research using dominance analysis can help extend the research base. For example, Ritchey (2008) compared the strength of correlation coefficients for measures of letter sound fluency (LSF) and nonsense word fluency (NWF), but did not determine the relative importance of one measure over another in predicting year-end word reading skills. The Schatschneider et al. (2004) dominance analysis offered important information on the kindergarten skills that are most predictive subsequent reading achievement. However, the results may not be entirely generalizable to progress monitoring as their assessment tasks (which included untimed assessment of letter names and sounds using flash cards, rapid automatized naming tasks using a set of five letters repeated several times, and subtests from larger standardized achievement tests) differ from published tools available for continuous progress monitoring. Kim et al. (2010) provided insight on measures with a first grade sample; however, kindergarten differs significantly in terms of the nature and content of reading instruction.

Second, studies have commonly used data gathered with a sample of students that represent the full spectrum of achievement levels. This approach is certainly important for revealing the predictive validity of the measures for all students, particularly when used for universal screening purposes. However, frequent formative assessment is important for (and most often used with) students who are at risk for reading difficulties or are receiving supplemental interventions (Gersten et al., 2009). Therefore, more comprehensive information on predictive validity is needed specifically for at risk students so that educators can make more informed decisions regarding the measures that may function best for lower achievers.

Third, many studies used assessment data from the latter half or end of kindergarten (e.g., Elliott et al., 2001; Ritchey & Speece, 2006; Stage et al., 2001) and did not evaluate how predictive validity may change across the school year. The dynamic nature of reading instruction across kindergarten may mean that some measures are more predictive of subsequent reading skills at one time in the school year and less predictive at others. For example, measures that assess skills that are a focus of instruction or best represent overall achievement in the fall of the school year, such as phonemic awareness, may be less predictive later in the year when perhaps more sophisticated or "downstream" skills may take over as the dominant predictors. Catts et al. (2009) investigated the predictive validity of early literacy measures that were administered on a grade-wide basis in the fall, winter, and spring of kindergarten. Although most measures increased their predictive validity across the school year, phoneme segmentation fluency (PSF) did not, and the predictive validity of some measures was quite low for students with lower achievement at earlier administrations. However, Catts et al. did not directly compare the measures to determine which were more important in predicting later reading outcomes.

Fourth, studies of kindergarten progress monitoring measures have rarely included measures of word reading. Word-list reading measures designed for kindergarten have recently become available in progress monitoring tool sets, but few studies have contrasted their predictive validity with other measures of basic early literacy. Clemens and Scholten (2012) found that a measure of fluency reading words in list form administered at the end of kindergarten was more strongly predictive of reading skills 1 year later compared with measures of letter sounds and PSF. In another study, Clemens et al. (2018) found that word-list fluency scores in the middle of kindergarten and slope across the second half of the school year were predictive of year-end reading skills; however, analyses did not determine the amount of unique variance in this prediction relative to other measures.

Finally, the use of CATs has proliferated in educational settings, including use with students in early elementary school. Several CATs are available in the areas of reading and early literacy for kindergarten students (FastBridge Learning, 2018; Istation, 2016; Northwest Evaluation Association, 2013; Renaissance Learning, 2015). In general, adaptive tests were developed to enhance precision while reducing testing time (i.e., estimating a "true score" by administering far fewer items than would be needed on a paper-based test) and have been used in large-scale testing situations such as screening or certification examinations (see van der Linden & Glas, 2010).

Recently, CATs have also been offered as options for monitoring kindergarten reading progress (Istation, 2016; Renaissance Learning, 2010) and, in some cases, have been marketed as being more reliable, valid, and more efficient than paper-based tools (Renaissance Learning, 2009). In addition, the validity of several CATs have been evaluated alongside paper-based progress monitoring tools as options for monitoring kindergarten reading progress (National Center on Intensive Intervention, 2018). As educators have greater evidence for selecting measures, it is important that research evaluate CATs for the roles in which they are marketed and recommended.

Unfortunately, very little independent research has contrasted the predictive validity of computer-adaptive measures with paper-based assessments directly in the same sample. Clemens et al. (2015) observed that although a CAT administered in kindergarten predicted reading skills at the end of kindergarten and first grade, it generally did not improve the prediction of reading outcomes (and in some cases demonstrated weaker concurrent validity) compared with paper-based measures such as word reading fluency. The sample only included students that represented a full range of achievement levels, and only contrasted the CAT with paper-based measures in the spring of the school year. Thus, more work is needed to determine how strongly CATs predict reading outcomes for at risk students, and how the strength of this prediction may change across the kindergarten year.

Study Purpose

The purpose of this study was to contrast the predictive validity of several measures currently available to educators for monitoring kindergarten reading progress on a frequent basis. Because the nature of kindergarten reading instruction and students' reading skills change across the school year, we evaluated the validity of the measures across several time points in predicting kindergarten year-end reading outcomes. We applied dominance analysis to examine the most important predictors of subsequent year-end reading achievement at each time-point. In addition, we focused the analyses on students considered to be at risk for reading difficulty, who are most likely to be the recipients of frequent progress monitoring.

Method

Participants and Settings

This study included a sample of 426 kindergarten students who were participating in a longitudinal investigation of measures for monitoring the reading progress of kindergarten students at risk for reading difficulties. The sample was 45.3% female, 19.7% White, 51.6% Hispanic, and 25.8% Black. Students that were considered to be English learners (31.7%) were included as part of the larger investigation provided they were being taught to read in English and they received at least 50% of reading instruction in English.

Kindergarten students were recruited during two consecutive school years from 10 elementary schools across rural and urban settings in the Southwest United States. Chi-square tests indicated that students in the two cohorts did not differ on a statistically significant basis on English learner status or sex; however, Cohort 2 included fewer Hispanic students. In terms of initial performance, t tests indicated that the cohorts did not differ on the initial administrations of Letter Naming Fluency (LNF), LSF, STAR Early Literacy (STAR), Letter Identification (LID), or Phonemic Awareness. However, Cohort 2 demonstrated stronger initial PSF performance (p = .04). Although the differences between the cohorts were minimal, we controlled for a cohort effect in the multilevel dominance analysis (described below).

Across all schools, the average percentage of students that qualified for free or reduced-price lunch was 77%. To identify participants, teachers first rated alphabetic knowledge, phonological awareness, oral language, and overall reading skills of each student in their classroom using a modified version of the *Reading Rating Form* (Speece et al., 2011). Parent permission forms were sent home with five to eight students who were rated lowest in each classroom. After obtaining parental consent, students qualified for enrollment by scoring at or below the 40th percentile on either the LID or PA subtests of the Woodcock Reading Mastery Test, 3rd edition (WRMT-III; Woodcock, 2011). We selected the 40th percentile as a way to sample the lower end of the achievement spectrum without truncating the range too severely. The 40th percentile is a common cut point on achievement tests (American Institutes for Research, 2007) and has been used to denote risk status in several studies (Brasseur-Hock, Hock, Kieffer, Biancarosa, & Deschler, 2011; Catts et al., 2009; Petscher & Kim, 2011).

Kindergarten Entry Skill Measures

LID. The LID subtest from the WRMT-III was administered to assess letter knowledge at the beginning of kindergarten. Students were asked to identify a series of upper and lower case letters on an untimed basis. The LID subtest demonstrates split-half and alternate-form reliability of .91 and .88, respectively, with kindergarten students.

PA. The PA battery from the WRMT-III includes a series of phonological awareness tasks based on words spoken by the examiner. Tasks are untimed and include first-sound matching, last sound matching, rhyme production, phoneme blending, and phoneme deletion. Scores across the tasks are summarized in an overall PA score. PA demonstrates split-half and alternate-form reliability of .92 and .78, respectively, with kindergarten students.

Progress Monitoring Measures

LNF. LNF assesses fluency in correctly identifying names of randomly ordered upper and lower case letters. LNF probes from the EasyCBM system were used, which demonstrated average 2-week alternate form reliability of .88 with the current sample. Correlations between LNF at the time points included in the present analyses and the year-end reading variables ranged from .37 to .59.

LSF. LSF is an assessment of fluency in correctly identifying sounds of randomly ordered lower case letters. LSF probes from the AIMSweb system were used, which demonstrated average 2-week alternate form reliability of .86 in our sample. Correlations between LSF at the time points included in the present analyses and the year-end reading variables ranged from .40 to .67.

PSF. PSF is an assessment of fluency in identifying phonemes and other sound segments in words. Students orally segment a series of words spoken by the examiner, and receive one point for each separate and unique sound segment produced within 1 min, which can include individual phonemes, syllables, or other word parts (no points are awarded when the student repeats a whole word). PSF from the EasyCBM system was used in this study, which demonstrated average 2-week alternate form reliability of .82. Correlations between PSF at the time points included in the present analyses and the year-end reading variables ranged from .26 to .53.

WRF. WRF from the EasyCBM system is an assessment of fluency in reading high-frequency words in list form. Word lists contain both phonetically regular (i.e., "decodable" words, such as "it") and phonetically irregular words (e.g., "me"). The total score consists of the number of words read correctly in 1 min, and any words read incorrectly or omitted, or hesitations of longer than 3 s were scored as errors. In our sample, WRF demonstrated average 2-week alternate form reliability of .91. Correlations between WRF at the time points included in the present analyses and the year-end reading variables ranged from .65 to .93.

Decodable Word Reading (DWR). DWR from the FastBridge system (FastBridge Learning, 2016) is an assessment of fluency reading phonetically regular real words in list form. All words follow a consonant–vowel–consonant (CVC) pattern. DWR is scored in terms of the number of words read correctly in 1 min. DWR demonstrated average 2-week alternate form reliability of .88 in our sample. Correlations between DWR at the time points included in the present analyses and the year-end reading variables ranged from .67 to .87.

NWF. NWF from the Dynamic Indicators of Basic Literacy Skills Next system (Dynamic Measurement Group, 2011) is an assessment of fluency in reading decodable pseudowords in list form. All pseudowords follow a VC or CVC pattern (e.g., ip, lut). Students are instructed to read the words in the list as best they can, and may say the sounds of the letters in the words if they cannot read the whole word. Two scores were derived from each NWF administration. The NWF correct letter sounds (NWF-CLS) score was based on the number of sounds students correctly produced in isolation (e.g., "b-i-m" = 3 points), as part of a word segment (e.g., "b-im" = 3 points), or whole word ("bim" = 3 points). A maximum of three points was awarded for each word. The NWF-Words score consisted of the number of words the students read correctly as a whole unit without segmenting or "sounding out" the word first. In our sample, the NWF-CLS and NWF-Words scores demonstrated average 2-week alternate form reliability of .82 and .83, respectively. Correlations between NWF-CLS at the time points included in the present analyses and the year-end reading variables ranged from .59 and .77. Correlations between NWF-Words at the time points included in the present analyses and the year-end reading variables ranged from .51 to .79.

STAR. STAR (Renaissance Learning, 2015) is a CAT designed for students in pre-kindergarten through Grade 2. Each administration consists of 27 multiple-choice items (three answer choices per item) that include graphic display and audio dictation. Examinees make answer choices using the computer mouse or keyboard. Each test administration begins with an exercise to verify that the student understands the use of the mouse or keyboard followed by a brief practice exercise in which the student must pass prior to starting the test. Kindergarten students complete STAR in approximately 15 min. STAR was selected for inclusion in the larger study given its suggested use as a progress monitoring measure (National Center on Intensive Intervention, 2018; Renaissance Learning, 2010) and its wide use in U.S. schools (Renaissance Learning, 2015).

STAR assesses skills in 10 subdomains: the alphabetic principle (i.e., identify letters and letter sounds), concept of word (i.e., discriminate words from letters, identify the number of words in a set), visual discrimination (i.e., differentiate upper and lower case letters, identify words that are the same or different), phonemic awareness, phonics (which includes decoding CVC words), vocabulary (i.e., matching words to pictures, reading high-frequency words, identifying word meanings), sentence-level comprehension, paragraph-level comprehension (i.e., identifying the main topic of a text), structural analysis (i.e., reading words with affixes, compound words), and early numeracy (Renaissance Learning, 2015). An adaptive branching process is used to select individual test items based initially on the examinee's age or grade, and subsequently on his or her accuracy of responding to test items. Typically, easier items follow incorrect responses, and more difficult items follow correct responses until the software has determined the student's scale score or 27 items have been administered. Some students may see more items from some skill domains depending on response accuracy. During test administration, STAR estimates a student's ability level using a proprietary procedure, and selects items based on a Rasch 1-parameter logistic response model. A Rasch scale is then used to express students' overall reading abilities, which are transformed as scaled scores ranging from 300 to 900.[1] The STAR technical manual reports a kindergarten split-half reliability of .75, 1-week test–retest reliability of .66, and "generic" reliability (i.e., an upper bound estimate of overall reliability estimated by calculating the ratio of error variance and scaled score variance and subtracting from 1) of .77 (Renaissance Learning, 2015). Correlations between STAR administrations and the year-end reading variables ranged from .36 to .46.

Year-End Criterion Measures

Word Identification (WID). The WID subtest from the WRMT-III is an assessment of reading isolated words that increase in difficulty. The measure is untimed, and administration ends after the student incorrectly reads four consecutive words. The year-end administration of WID demonstrated Kuder–Richardson Formula 20 (KR-20) internal consistency of .84.

Word Attack (WA). The WA subtest from the WRMT-III was used to assess decoding accuracy. Students read from a list of phonetically regular pseudowords that increase in difficulty. The subtest is untimed and is scored according to the number of pseudowords read correctly as a whole unit (segmented or partially blended words did not receive credit). In our sample, the KR-20 consistency coefficient was .79.

Sight Word Efficiency (SWE). On the SWE subtest from the Test of Word Reading Efficiency–Second Edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012), students are asked to read a list of real words that increase in difficulty. The measure is scored in terms of the number of words read correctly in 45 s. The test authors report an average 2-week test–retest reliability of .93 for students aged 6 to 7 (Torgesen et al., 2012). SWE demonstrated internal consistency (KR-20) of .87 with our sample.

Phonemic Decoding Efficiency (PDE). The PDE subtest of the TOWRE-2 assesses fluency in reading a list of phonetically regular pseudowords that increase in difficulty. The measure is scored in terms of the number of pseudowords read correctly in 45 s. The test authors report average 2-week test–retest reliability of .91 for students aged 6 to 7 (Torgesen et al., 2012). In our sample, PDE demonstrated internal consistency (KR-20) of .87.

ORF. Students' fluency reading connected text was assessed using the passage "Mac Gets Well" (Makar, 1995), which consists of high frequency and phonetically decodable words. This measure has been used in previous studies with low-achieving kindergarten students (Coyne et al., 2013; Vadasy, Sanders, & Peyton, 2006). Students were awarded one point for each word they read as a complete unit within 1 min (partially blended words did not receive credit). Vadasy et al. reported Cronbach's alpha coefficient of .93 with kindergarten students.

Procedures

LID and PA were administered in October prior to the first administration of the progress monitoring measures. The fall progress monitoring battery (October through December) included LNF, LSF, and PSF. The spring progress monitoring battery (January through April) included the measures administered across the fall, as well as WRF, DWR, and NWF, to assess students' emerging word reading skills across the second half of kindergarten. As part of the larger project, students were administered progress monitoring measures once every 2 weeks (5 times across the fall and 8 times across the spring). STAR, which is not designed for administration more frequently than once per month (Renaissance Learning, 2015), was administered in October, December, February, and April. For the present study, in the interest of parsimony in the analyses and interpretation, we used a subset of progress monitoring assessment points that were representative of the school year (October, December, January, March, April). The criterion battery of year-end reading skills was administered in late April/early May following the final progress monitoring data point. All measures were administered under standardized conditions using the administration procedures specified by the publishers of each measure. Testing locations included empty classrooms and other quiet locations in the schools.

Examiner training and assessment fidelity. Data collectors included senior research staff, graduate students, and advanced undergraduate students. A series of training sessions included an overview of best practices in assessment with systematic explanations and modeling of the specific protocol requirements for each measure. Following individualized practice sessions, trainee reliability was established through mock test administrations, which included feedback and modeled administration by senior project staff. Finally, trainees administered assessments to study participants under supervision by senior project staff, and were required to demonstrate 100% fidelity to assessment procedures and at least 95% inter-scorer agreement with the senior staff member before they were allowed to administer assessments independently. Follow-up fidelity checks were conducted by research staff 6 to 8 times per year, and interrater reliability (calculated as agreements divided by agreements plus disagreements) met or exceeded 95% for all measures at each time point. All assessments were double-scored at the item level at data entry (with the exception of STAR, which students take individually on a computer).

Data Analyses

Year-end reading skills latent variable. A latent variable was formed using WID, WA, SWE, PDE, and ORF to summarize overall reading achievement at the end of kindergarten. This model demonstrated good fit to the data, $\chi^2 = 7.84$ ($df = 3$, $p = .05$), comparative fit index = .997, root mean square error of approximation = .066 (90% confidence interval [CI] = [.003, .124]), Tucker–Lewis index = .989; standardized root mean square residual = .007, and factor loadings ranged from .77 to .94. Each student's factor score on this variable was saved and used as the dependent variable in the analyses.

Missing data. Across the measures and time points used in this study, the mean percentage of cases that were missing data on a measure for a given time point was 15.93% (range = 5.39%–22.30%). Data were missing primarily due to student absences or unanticipated changes in school or classroom schedules on days of testing that could not be made up at a later date. The mechanism of missing data was determined to be missing completely at random (MCAR) according to the results of Little's MCAR test (Little, 1988). Missing data were handled via a multilevel multiple imputation approach using the Pan package in the R software (Schafer & Yucel, 2002; Schafer & Zhao, 2013), which is the most commonly used approach to impute values in data that have a multilevel structure (Lüdtke, Robitzsch, & Grund, 2017). Twenty imputed datasets were created and all analyses results were pooled across the 20 imputations.

Dominance analysis. Dominance analyses were conducted using a SAS Macro developed by Luo and Azen (2013) to determine the relative importance of individual predictors among all possible models of other predictors (i.e., subset models) while accounting for the multilevel structure of the data (i.e., students clustered within classrooms). The SAS Macro was used to conduct multilevel dominance analyses at each of the five time points using the predictors available at that time, while controlling for early literacy skills at kindergarten entry (LID and PA), as well as demographic variables which included English learner status, cohort membership (Year 1 or Year 2 of data collection), and urban vs. rural school region. The baseline model containing only the control variables was specified as

$$Y_{ij} = \gamma_{00} + \beta_1\text{Cohort} + \beta_2\text{EL} + \beta_3\text{Region} + \beta_4\text{LID} + \beta_5\text{PA} + U_{0j} + e_{ij},$$

where i indexes students within school j, γ_{00} represents the fixed effect (intercept), βs are the regression weights, and U_{0j} and e_{ij} represent the classroom random effects and Level 1 residuals, respectively.

Due to the multilevel structure of the data, the additional contribution of a student-level predictor can be quantified by various computations of R^2 statistics (see Luo & Azen, 2013, for a discussion of R^2 computations). Because we were interested in predicting students' end-of-year outcome within classrooms (i.e., controlling for classroom random effects), we used a Level 1 approach proposed by Raudenbush and Bryk (2002; which we refer to as R&B R^2) as the criterion measure, which quantifies the contribution of a specific predictor in explaining within-classroom variation:

$$\text{Level 1 } R\,\&\,B\,R^2 = \frac{\left[\sigma_e^2 \middle| M_1 - \sigma_e^2 \middle| M_2\right]}{\sigma_e^2|M_1},$$

where σ_e^2 is the Level 1 residual variance, M_1 is the hierarchical linear model without the predictor of interest (i.e., baseline model), and M_2 is the model containing the variable of interest.

Via pair-wise comparisons, dominance analysis reveals the additional contribution of each predictor relative to other predictors when added to all possible subset multilevel models. The contribution of each predictor variable can be further interpreted in terms of its level of dominance, which refers to how strongly and clearly one predictor dominates another (Azen & Budescu, 2003). *Complete dominance* occurs when the additional contribution of a predictor is always greater than another predictor in *every* possible subset model. *Conditional dominance* occurs when the average additional contribution of one predictor within models of the same size is always greater than that of the other predictor. *General dominance* is observed when the average additional contribution of one predictor across all the possible subset models is greater than that of another predictor.

Results

Descriptive data are reported in Table 1. Unique contributions of each predictors to the subset models across time points are reported in Table 2, which can be interpreted in terms of the proportion of variance (i.e., R^2) accounted for by each predictor while controlling for other predictors in each model. Results of the dominance analyses from each assessment point are reported in Table 3. The results reported in Tables 2 and 3 are discussed together, as follows.

Of the measures administered in October, STAR made the largest average contribution across subset models, and demonstrated complete dominance over LNF, LSF, and PSF in predicting kindergarten year-end outcomes. LNF was dominant over LSF and PSF, and LSF demonstrated complete dominance over PSF.

A shift was observed in the dominance of the measures when considering the December administration. As reported in Table 2, LSF made the largest average contribution across subset models, and demonstrated complete dominance over LNF, PSF, and STAR (see Table 3). PSF, previously not dominant over any predictors in October, now completely dominated LNF and STAR. LNF was completely dominant over STAR. STAR did not dominate any of the measures at the December administration.

In January, the WRF, DWR, and NWF measures were added to the progress monitoring battery. As reported in Table 3, WRF and DWR were the most dominant predictors of the measures administered at this time, but it is notable that LSF demonstrated either conditional or general dominance over LNF, PSF, and the two NWF scoring methods. Of the other predictors, PSF did not demonstrate dominance over any other measures (STAR was not administered).

Of the measures administered in February, DWR and WRF were the most dominant predictors of year-end reading skills, respectively, followed by the NWF scoring methods. LSF demonstrated conditional or general dominance over LNF, PSF, and STAR. STAR demonstrated general

Table 1. Descriptive Statistics.

Month and Measure	Letter Naming Fluency	Letter Sound Fluency	Phoneme Segmentation Fluency	Word Reading Fluency	Decodable Words Fluency	NWF-Correct Letter Sequences	NWF-Words	STAR Early Literacy
October								
M	20.17	8.45	10.84					488.07
SD	13.86	8.47	11.97					87.43
Range	0–60	0–43	0–49					319–815
December								
M	28.62	17.42	19.26					534.80
SD	16.26	13.67	16.30					93.14
Range	0–75	0–60	0–60					318–782
January								
M	33.16	19.90	24.90	3.69	2.67	15.40	1.15	
SD	17.54	14.68	17.98	4.90	5.72	16.00	4.35	
Range	0–76	0–64	0–71	0–47	0–46	0–114	0–40	
February								
M	35.48	25.12	25.85	4.15	3.90	20.14	1.53	562.33
SD	18.17	16.51	15.94	5.59	6.31	18.77	4.56	97.21
Range	0–84	0–73	0–62	0–59	0–45	0–113	0–30	339–828
April								
M	40.27	29.32	32.52	8.27	6.34	28.20	3.04	587.59
SD	19.06	17.61	16.40	8.98	9.06	23.22	6.80	105.53
Range	0–88	0–79	0–77	0–64	0–56	0–161	0–43	322–842
	WID	WA	SWE	PDE	ORF			
May								
M	4.46	1.91	10.14	5.03	13.11			
SD	4.35	2.81	9.04	5.38	14.35			
Range	0–26	0–16	0–61	0–34	0–113			

Note. Scale scores reported for STAR Early Literacy. Raw scores are reported for all other variables. NWF = Nonsense Word Fluency; WID = Word Identification; WA = Word Attack; SWE = Sight Word Efficiency; PDE = Phonemic Decoding Efficiency; ORF = Oral Reading Fluency.

dominance over PSF, and PSF did not dominate any of the measures.

In April, WRF and DWR were the most dominant predictors, respectively, of year-end reading skills. LSF demonstrated dominance over LNF, PSF, and STAR.

Discussion

A number of measures are available for monitoring reading development with struggling students in kindergarten, but several questions remain regarding which are most preferable. Although several factors are relevant for selecting a progress monitoring measure, one important aspect is the degree to which scores on a measure are associated with important reading outcomes (Deno, 1985; Fuchs, 2004), and understanding the strength of the prediction relative to other available measures may assist educators in decisions regarding measure selection. Toward this end, the present study contrasted the predictive validity of several measures for monitoring kindergarten reading progress that were administered on a repeated basis, and determined the degree to which they were predictive over and above measures that assessed different but related skills. This study added uniquely to the research in this area by utilizing a large and relevant sample, data collection across the kindergarten year, and dominance analysis to determine which measures were the most important predictors of year-end reading achievement.

To summarize the results, dominance analyses indicated that in October of kindergarten, STAR was the most dominant predictor of year-end reading outcomes. By December, however, the picture had changed dramatically; LSF had become the most dominant predictor of year-end reading skills, and STAR was the least dominant. In addition, whereas PSF was not dominant over any other predictor in October, by December, PSF was dominant over LNF and STAR. Measures of word reading and decoding (WRF, DWR, NWF) were added to the progress monitoring battery during the second half of kindergarten. Of the measures administered in January, WRF and DWR were the strongest

Table 2. Unique Predictor Contributions to Year-End Reading Skills Across Subset Models and Time Points.

Time	Predictor	1 IV	2 IVs	3 IVs	4 IVs	5 IVs	6 IVs	7 IVs	8 IVs	Average
October	STAR	0.129	0.096	0.076	0.064	—	—	—	—	0.091
	LNF	0.078	0.051	0.036	0.027	—	—	—	—	0.048
	LSF	0.057	0.028	0.011	0.003	—	—	—	—	0.025
	PSF	0.020	0.008	0.003	0.001	—	—	—	—	0.008
December	LSF	0.219	0.157	0.118	0.088	—	—	—	—	0.145
	PSF	0.091	0.053	0.036	0.025	—	—	—	—	0.051
	LNF	0.083	0.038	0.014	0.000	—	—	—	—	0.034
	STAR	0.000	0.000	0.000	0.000	—	—	—	—	0.000
January	WRF	0.510	0.267	0.156	0.100	0.069	0.052	0.043	—	0.171
	DWR	0.439	0.205	0.102	0.051	0.024	0.010	0.003	—	0.119
	LSF	0.302	0.134	0.075	0.050	0.035	0.024	0.016	—	0.091
	NWF-WWR	0.314	0.136	0.063	0.029	0.012	0.005	0.003	—	0.080
	NWF-CLS	0.330	0.118	0.038	0.009	0.000	0.000	0.000	—	0.071
	LNF	0.149	0.057	0.031	0.020	0.013	0.009	0.006	—	0.041
	PSF	0.113	0.041	0.022	0.015	0.010	0.007	0.005	—	0.031
February	DWR	0.592	0.327	0.189	0.110	0.064	0.037	0.022	0.013	0.169
	WRF	0.556	0.306	0.178	0.106	0.065	0.042	0.030	0.023	0.163
	NWF-WWR	0.453	0.239	0.130	0.069	0.035	0.016	0.008	0.005	0.119
	NWF-CLS	0.457	0.216	0.102	0.045	0.018	0.007	0.004	0.002	0.106
	LSF	0.240	0.096	0.044	0.023	0.014	0.009	0.006	0.003	0.054
	LNF	0.155	0.056	0.025	0.014	0.009	0.006	0.004	0.003	0.034
	STAR	0.072	0.029	0.019	0.015	0.013	0.011	0.010	0.009	0.022
	PSF	0.106	0.033	0.013	0.006	0.003	0.002	0.001	0.001	0.021
April	WRF	0.807	0.508	0.346	0.252	0.196	0.164	0.145	0.132	0.319
	DWR	0.642	0.349	0.193	0.104	0.054	0.026	0.011	0.001	0.173
	NWF-WWR	0.559	0.294	0.155	0.078	0.036	0.015	0.006	0.003	0.143
	NWF-CLS	0.524	0.257	0.124	0.056	0.023	0.010	0.006	0.004	0.126
	LSF	0.205	0.070	0.025	0.011	0.006	0.004	0.003	0.002	0.041
	LNF	0.143	0.037	0.008	0.001	0.000	0.000	0.000	0.000	0.024
	PSF	0.084	0.028	0.014	0.009	0.006	0.005	0.004	0.003	0.019
	STAR	0.041	0.012	0.007	0.005	0.003	0.002	0.001	0.000	0.009

Note. For each time point, predictor variables are ordered according to their average unique contribution to end of ear reading outcome, from highest to lowest. Values reflect the proportion of unique variance accounted for by each predictor while controlling for other predictors in each subset model. STAR = STAR Early Literacy (computer adaptive measure); LNF = Letter Naming Fluency; LSF = Letter Sound Fluency; PSF = Phoneme Segmentation Fluency; WRF = Word Reading Fluency; DWR = Decodable Words Fluency; NWF-WWR = Nonsense Word Fluency whole words read score; NWF-CLS = Nonsense Word Fluency correct letter sounds score.

predictors of year-end reading skills, but it is notable that LSF remained a strong predictor and demonstrated dominance in some form over the two NWF scoring methods, LNF, PSF, and STAR. WRF and DWR were the strongest predictors across February and April (with negligible differences between the two), followed by the NWF scoring methods. LSF continued to dominate LNF, PSF, and STAR across February and April.

Overall, the results revealed several important findings that have implications for educators, school leaders, and researchers. First, among the measures that assessed basic skills in alphabetic knowledge and phonological awareness, LSF emerged as the strongest predictor across the school year compared with LNF and PSF. With the exception of the October administration, in which STAR was most predictive of year-end outcomes relative to the other measures administered at that time, LSF was the dominant predictor during the later portion of the fall semester and continued to be strongest in relation to LNF, PSF, and STAR across the spring. Knowledge of letter-sound correspondence provides the essential raw material for decoding and word recognition and is thus a critical skill for reading acquisition (Ehri, 1998, 2005; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012), and the findings in the present study are consistent with this view. It is also helpful to consider that measures like LSF that show low scores at the outset but greater scores over time, especially when those scores are predictive of subsequent reading outcomes, may be useful for progress monitoring. Although only a limited number of studies have directly compared the predictive validity of LSF with other measures in kindergarten, existing studies have included only small number of measures for comparison, have limited

Table 3. Dominance Analysis Results Predicting Year-End Reading Outcomes From Each Time Point.

Time	Predictor	Complete Dominance Over	Conditional Dominance Over	General Dominance Over
October	STAR	LNF, LSF, PSF	—	—
	LNF	LSF, PSF	—	—
	LSF	PSF	—	—
	PSF	—	—	—
December	LSF	PSF, LNF, STAR	—	—
	PSF	LNF, STAR	—	—
	LNF	STAR	—	—
	STAR	—	—	—
January	WRF	LSF, NWF-WWR, NWF-CLS	DWR, LNF, PSF	—
	DWR	—	NWF-WWR, NWF-CLS	LSF, LNF, PSF
	LSF	—	LNF, PSF	NWF-WWR, NWF-CLS
	NWF-WWR	—	—	NWF-CLS, LNF, PSF
	NWF-CLS	—	—	LNF, PSF
	LNF	—	PSF	—
	PSF	—	—	—
February	DWR	LSF, LNF	NWF-WWR, NWF-CLS, STAR, PSF	WRF
	WRF	NWF-CLS, LSF, LNF, STAR, PSF	NWF-WWR	—
	NWF-WWR	—	LSF, LNF, PSF	NWF-CLS, STAR
	NWF-CLS	—	PSF	LSF, LNF, STAR
	LSF	—	LNF, PSF	STAR
	LNF	PSF	—	STAR
	STAR	—	—	PSF
	PSF	—	—	—
April	WRF	DWR, NWF-WWR, NWF-CLS, LSF, LNG, PSF, STAR	—	—
	DWR	STAR	—	NWF-WWR, NWF-CLS, LSF, LNG, PSF
	NWF-WWR	—	LSF, PSF, STAR	NWF-CLS, PSF
	NWF-CLS	LNF	LSF, PSF, STAR	—
	LSF	LNF	STAR	PSF
	LNF	—	—	PSF, STAR
	PSF	—	STAR	—
	STAR	—	—	—

Note. STAR = STAR Early Literacy (computer adaptive measure); LNF = Letter Naming Fluency; LSF = Letter Sound Fluency; PSF = Phoneme Segmentation Fluency; WRF = Word Reading Fluency; DWR = Decodable Words Fluency; NWF-WWR = Nonsense Word Fluency whole words read score; NWF-CLS = Nonsense Word Fluency correct letter sounds score.

assessment to the second half of kindergarten or end of the year, and did not investigate dominance (Elliott et al., 2001; Ritchey, 2008; Ritchey & Speece, 2006; Stage et al., 2001). Thus, the present results expand on this research base and demonstrate the educational implications that LSF is a strong predictor of subsequent reading skills relative to other "sublexical" skills assessed in the fall of kindergarten. It is also notable that LSF remains a good predictor during the spring, where it dominated LNF, PSF, and STAR.

LNF is offered as a progress monitoring measure by several assessment suites such as AIMSweb, EasyCBM, and Fastbridge. Although it was dominated by STAR in October, LNF was a stronger predictor than LSF and PSF earliest in the school year, which is consistent with prior work that has demonstrated that letter name knowledge is an important predictor of future reading achievement in preschool and kindergarten entry (Badian, 1995; Piasta, Petscher, & Justice, 2012). Greater strength on the part of LNF at kindergarten entry compared with measures of letter sounds or phoneme segmenting may be due in part to a greater range of scores at that time of year. It is also possible that letter name knowledge at kindergarten entry may serve as an index of home literacy environment and exposure to literacy activities and instruction prior to kindergarten, which affords numerous benefits to promoting subsequent reading development (Foulin, 2005). After October, as students began to develop their knowledge of letter sounds, LNF was dominated by LSF, suggesting that LSF may be a more important alphabetic fluency measure as time goes on.

Word reading measures have received less attention for use in monitoring the reading progress of kindergarten students compared with other measures such as LSF, PSF, and NWF; however, our results revealed that from January and beyond, they were the most important predictors of year-end reading skills. Even among our population that was considered to be at risk for reading difficulties, the WRF and DWR measures from January onward were the most dominant predictors of year-end outcomes relative to all others.

These results suggest that measures of word reading may serve as key indicators of progress toward year-end reading outcomes for at risk kindergarten students. As noted earlier, word reading becomes an area of emphasis in kindergarten reading curricula primarily during the second half of the school year, and national and state standards for kindergarten reading achievement typically include expectations that students should be able to decode simple words, read high-frequency words, and read grade appropriate texts by the end of kindergarten (.g., National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; Texas Education Agency, 2016). The emphasis of contemporary kindergarten reading curricula on decoding and word reading skills, as well as expectations in state and national standards, suggest that measures of word reading may be viable indicators of reading acquisition for at-risk kindergarten students, at least across the second half of the year.

The findings observed for PSF are worthy of additional discussion. PSF measures are commonly available across most commercially available progress monitoring toolsets for kindergarten students. However, PSF demonstrated very little predictive validity relative to the other measures, consistent with prior work (Catts et al., 2009; Clemens, Shapiro, & Thoemmes, 2011; Goffreda & DiPerna, 2010). Although there is no debate regarding the importance of phonological awareness in supporting reading development (Wagner & Torgesen, 1987), two factors may influence the inability of PSF in particular to more strongly predict later reading skills. First, students respond to verbal stimuli rather than print, and therefore, it may be less indicative of reading skills compared with other print-based measures. Second, students can obtain scores on PSF in a number of different ways, including segmenting by phonemes, syllables, an onset-rhyme pattern, or stating initial sounds. Thus, it is possible for two students may earn the same score, although one student may respond in a way that reflects greater sophistication in phonological awareness (e.g., segmenting phonemes) compared with a student who responds in a less refined or sophisticated manner (e.g., segmenting syllables).

With regard to the NWF scoring methods, both were subordinate to LSF as predictors when assessed in January. Although they demonstrated dominance over LSF, LNF, PSF, and STAR across February and April administrations, NWF scores were always dominated to some extent by measures of reading real words (i.e., WRF and DWR). NWF has featured prominently in the DIBELS and AIMSweb systems as an option for monitoring kindergarten reading progress. Intended as a bridge between letter sound awareness and word reading, common administration formats for NWF have provided students with an option to identify the sounds in the letters of each word, or by reading the word as whole units (Good & Kaminski, 2002; Pearson, 2012). When using the NWF-CLS score, students receive credit for sounds read in isolation, as part of word segments, or within whole words. Prior work has suggested the possibility that, like with PSF, two different students may arrive at the same score in ways that reflect different levels of sophistication in early literacy skills, thus creating ambiguity in interpreting the results regarding students' reading skill needs, and possibly leading to problems in the predictive validity of the measure (Clemens, Shapiro, Wu, Taylor, & Caskie, 2014; Fuchs et al., 2004). It should also be noted that the NWF-Words score used in this study, which only provided credit for words that were read correctly as whole units, typically dominated the traditional NWF-CLS score but never improved upon WRF or DWR. Therefore, our results suggest that the prediction of year-end reading skills may be more clear when using a measure of reading real words, rather than NWF.

STAR, the computer-adaptive measure, is marketed as an option for progress monitoring (Renaissance Learning, 2009, 2010, 2015) and has been reviewed together with other paper-based progress monitoring tools (National Center on Intensive Intervention, 2018). STAR was the most dominant predictor of year-end outcomes at the October time point. Additional inspection of the multi-level data revealed that within each classroom, there was a group of students that scored very low on LNF, LSF, and PSF. Had DWR, WRF, and NWF been administered in October, it is likely nearly all students would have scored low on these measures as well. With its ability to assess a wider range of language and beginning literacy skills, STAR likely has a lower "floor" and may be more sensitive to individual skill differences early in kindergarten. This may make it a better screening measure at kindergarten entry in contrast to other measures of specific skills. However, after October, STAR was the least predictive of year-end reading skills and was dominated by nearly all of the paper-based measures for the remainder of the year.

There may be several reasons why STAR was dominated by the other measures across the remainder of the year. First, our year-end assessment battery was focused on word reading, decoding, and text reading. STAR evaluates multiple skill areas that are related to reading outcomes; however, the computer-adaptive nature of the assessment may have resulted in scores that were less associated with word reading skills more specifically. In any given STAR administration, students see 27 items. The adaptive nature of the assessment means that some students may not see any items in skill areas that correspond more closely to decoding and word recognition, which may limit the relation between scores and word reading outcomes. It is also possible that the inclusion of language skills in STAR may make it a stronger predictor of literacy outcomes in subsequent grades (reading comprehension in particular), and our future work will investigate this possibility.

Second, in contrast to several of the measures in the year-end battery as well as all of the paper-based progress monitoring measures, STAR scale scores are not rate-based. Thus, the fluency-based nature of the paper progress monitoring assessments may have resulted in a stronger relation to year-end outcomes compared with STAR.

Third, students complete STAR independently on the computer by responding to multiple choice questions. Thus, in contrast to most of the paper-based predictors which required students to respond orally to printed letters or words, STAR shared less method variance with the year-end outcome assessments. In addition, we do not know how often students guessed at the multiple-choice questions in the assessment, but it is possible that impulsivity and guessing may have played a role in reducing the predictive validity of STAR. An earlier study of a computer-based intervention with kindergarten children revealed a high percentage of random and meaningless mouse activity among lower performing students (Kegel, van der Kooy-Hofland, & Bus, 2009); therefore, future research might investigate guessing on computer-based assessments that are designed for students to complete independently.

Limitations

There were several limitations to this study. We used a limited set of data points from our available set of progress monitoring data to make the analyses and results more manageable. Although the data points were representative of the school year, and it is doubtful that the relative differences observed among the measures would have varied a great deal from the present results, it is possible that we may have missed other times in which measures (such as PSF) were more important predictors.

The generalizability of the findings of this study must be considered carefully. First, our conclusions are limited to the measures included in this project. Although the progress monitoring measures used in this study are representative of those available for use with kindergarten populations (e.g., similar PSF measures are available from DIBELS Next, AIMSweb, and EasyCBM), the results of the dominance analyses should be applied cautiously when considering measures from other publishers. Second, no attempts should be made at generalizing the results to typically or high-achieving students, as our sample was considered "at-risk" for reading difficulties. Third, caution should also be exercised when generalizing these findings to high-socioeconomic status (SES) populations, as the students in this study attended schools in which the majority of students came from economically disadvantaged households. Fourth, English learners who were learning to read in English and received at least 50% of their instruction in English were included in the analyses, and although we controlled for English learner status in the analyses, readers should use caution when extending the results to English learners.

Furthermore, the purpose of the study must be understood when considering its conclusions and implications. This was not an investigation of classification accuracy and our results should not be interpreted as evidence for the accuracy of the measures as universal screening tools. Although we investigated the predictive validity of several measures that have been used for universal screening purposes (see Jenkins, Hudson, & Johnson, 2007), students in our sample were already considered to be at-risk for reading difficulty (both by their teachers and through verification with our preliminary assessment battery), and the majority were receiving supplemental intervention supports when available in their schools.

Implications and Conclusion

The number of measures that are available for monitoring kindergarten reading progress and the complex set of skills that develop and converge across the kindergarten year presents a challenging decision regarding what measures to use to monitor kindergarten reading progress. Indeed, there are many reasons why an educator may select a particular measure for progress monitoring. This study investigated one aspect that may play into that decision, that being the extent to which scores on a measure are predictive of year-end reading outcomes relative to other measures of early literacy skills. Of the measures evaluated in this study, measures of WRF were the strongest predictors of year-end reading skills (at least across the second half of the school year) compared with measures of pseudoword reading, letter name and sound fluency, PSF, and a CAT. Of measures that evaluated more basic alphabetic or phonemic awareness skills, with the exception of the first time point, LSF generally outperformed measures of fluency in naming letters, segmenting phonemes, and a CAT. Overall, across the measures included in this study, the most dominant measures at different points in time across the year were those that were more likely to be more reflective of the predominant skills targeted in instruction for that period of time. As noted earlier, contemporary kindergarten reading curricula tend to emphasize alphabetic knowledge (letter-sound correspondence in particular) and phonological awareness across the first half of kindergarten, with gradual introduction of decoding and word recognition, and a stronger emphasis on decoding and word recognition skills by the end of the year.

With regard to instructional decision making, the present results suggest that although initial scores may not be highly predictive of year-end outcomes, a measure such as LSF that assesses skills in the acquisition of letter-sound correspondence may be a good option for monitoring students' acquisition of the alphabetic principle across the fall and winter of the school year. Subsequently, across the second half of the school year as instruction tends to place increasing emphasis on decoding and word recognition skills, a measure of word reading is suggested as an index students' acquisition of skills target through instruction. Our subsequent work will continue to evaluate kindergarten progress monitoring measures with at risk learners, specifically in terms of their sensitivity to growth and the extent to which slope of improvement is associated with reading outcomes.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by a grant from the U.S. Department of Education, Institute for Education Sciences, Grant R324A130214. The opinions expressed in this article are those of the authors and do not represent the views of the Institute of Education Sciences or the U.S. Department of Education.

Note

1. STAR provides subscale scores for the domains listed earlier; however, we only used total scaled scores in our analyses. Due to STAR's adaptive technology, students may only see a limited number of items in some domains based on their response accuracy; therefore, scaled scores are considered the strongest estimate of a student's overall reading skills at a particular time (Renaissance Learning, 2015).

ORCID iD

Nathan H. Clemens https://orcid.org/0000-0003-3635-4241

References

American Institutes for Research. (2007). *Reading first state APR data*. Washington, DC: Author.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129–148.

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.

Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of readability statistics and equating for alternate forms. *Journal of School Psychology*, *47*, 1–17.

Brasseur-Hock, I. F., Hock, M. F., Kieffer, M. J., Biancarosa, G., & Deshler, D. D. (2011). Adolescent struggling readers in urban schools: Results of a latent class analysis. *Learning and Individual Differences*, *21*, 438–452.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542–551.

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

Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C. A., Blakely, A., Frosch, J., . . . Jones, M. (2015). The predictive validity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review*, *44*, 76–97.

Clemens, N. H., & Scholten, T. L. (2012, February). *The implications of growth in word identification fluency across first grade*. Paper presented at the 2012 Annual Convention of the National Association of School Psychologists, Philadelphia, PA.

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

Clemens, N. H., Shapiro, E. S., Wu, J. Y., Taylor, A., & Caskie, G. L. (2014). Monitoring early first grade reading progress: A comparison of two measures. *Journal of Learning Disabilities*, *47*, 254–270.

Clemens, N. H., Soohoo, M., Wiley, C. P., Hsiao, Y., Estrella, I., Allee-Smith, P.J., & Yoon, M. (2018). Advancing stage 2 research on measures for monitoring kindergarten reading progress. *Journal of Learning Disabilities*, *51*, 85–104.

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O. M., Kim, M., . . . Rawlinson, D. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*, *80*, 25–44.

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

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

Dynamic Measurement Group. (2011). *DIBELS next*. Eugene, OR: Author. Available from https://dibels.org/dibelsnext.html

Ehri, L. C. (1998). Research on learning to read and spell: A personal-historical perspective. *Scientific Studies of Reading*, *2*, 97–114.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*, 167–188.

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills—Modified. *School Psychology Review*, *30*, 33–44.

FastBridge Learning. (2016). *Decodable word reading*. Minneapolis, MN: Author. Available from http://www.fastbridge.org/

FastBridge Learning. (2018). aReading (Adaptive Reading). Minneapolis, MN: Author. Retrieved from http://www.fastbridge.org/assessments/reading/areading/

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

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–193.

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

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239–256.

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 for reading 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.

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

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available from http://dibels.uoregon.edu/

Houghton Mifflin Harcourt. (2014). *Houghton Mifflin reading* (Instructional curriculum). Retrieved from http://www.hmhco.com/shop/education-curriculum/reading/core-reading-programs/houghton-mifflin-reading

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read combining intervention studies with mediation analyses. *Psychological Science*, *23*, 572–577.

Istation. (2016). *Istation early reading assessment*. Dallas, TX: Author. Retrieved from http://www.istation.com/Product/EarlyReading

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*, 582–600.

Kamii, C., & Manning, M. (2005). Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A tool for evaluating student learning? *Journal of Research in Childhood Education*, *20*, 75–90.

Kegel, C. A., van der Kooy-Hofland, V. A., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences*, *19*, 549–554.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*, 141–165.

Luo, W., & Azen, R. (2013). Determining predictor importance in hierarchical linear models using dominance analysis. *Journal of Educational and Behavioral Statistics*, *38*, 3–31.

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

Mellard, D. F., Anthony, J. L., & Woods, K. L. (2012). Understanding oral reading fluency among adults with low literacy: Dominance analysis of contributing component skills. *Reading and Writing*, *25*, 1345–1364.

National Center on Intensive Intervention. (2018). *Academic progress monitoring* (Tools chart). Washington, DC: Author. Retrieved from http://www.intensiveintervention.org/chart/progress-monitoring

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*. Washington, DC: Authors.

Northwest Evaluation Association. (2013). *Measures of academic progress*. Available from www.nwea.org

Pearson. (2011). *Scott Foresman reading street* [Instructional curriculum]. Retrieved from http://www.pearsonschool.com/index.cfm?locator=PS1dH9

Pearson, N. C. S. (2012). *AIMSweb technical manual*. Bloomington, MN: Author.

Petscher, Y., & Kim, Y. S. (2011). Efficiency of predicting risk in word reading using fewer, easier letters. *Assessment for Effective Intervention*, *37*, 17–25.

Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology*, *104*, 945–958.

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: SAGE.

Renaissance Learning. (2009). *With the right information, you can help your students shine like stars* (Product brochure). Wisconsin Rapids, WI: Author.

Renaissance Learning. (2010). *Getting the most of STAR early literacy: Using data to inform instruction and intervention*. Wisconsin Rapids, WI: Author.

Renaissance Learning (2015). *STAR Early Literacy technical manual*. Author. Available: http://doc.renlearn.com/kmnet/r004384710gj119f.pdf

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

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

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

Schafer, J. L., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, *11*, 437–457.

Schafer, J. L., & Zhao, J. H. (2013). Pan: Multiple imputation for multivariate panel or clustered data (R package version 1.4) [Computer program]. Retrieved from http://CRAN.R-project.org/package=pan

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–279.

Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal*, *111*, 585–598.

Stage, S. A., Sheppard, J., Davidson, M. M., & Browning, M. M. (2001). Prediction of first-graders' growth in oral reading fluency using kindergarten letter fluency. *Journal of School Psychology*, *39*, 225–237.

Texas Education Agency. (2016). *Texas essential knowledge and skills for kindergarten*. Retrieved from http://www.tea.state.tx.us/index2.aspx?id=6148

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency* (2nd ed.). Austin, TX: Pro-Ed.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*, 508–516.

van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. New York, NY: Springer.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–203.

Woodcock, R. (2011). *Woodcock Reading Mastery Tests—Third Edition (WRMT–III)*. Circle Pines, MN: American Guidance Service.