

Research Article

Predicting Reading Problems 6 Years Into the Future: Dynamic Assessment Reduces Bias and Increases Classification Accuracy

Douglas B. Petersen,^a Shelbi L. Gragg,^b and Trina D. Spencer^c

Purpose: The purpose of this study was to examine how well a kindergarten dynamic assessment of decoding predicts future reading difficulty at 2nd, 3rd, 4th, and 5th grade and to determine whether the dynamic assessment improves the predictive validity of traditional static kindergarten reading measures.

Method: With a small variation in sample size by grade, approximately 370 Caucasian and Hispanic students were administered a 3-min dynamic assessment of decoding and static measures of letter identification and phonemic awareness at the beginning of kindergarten. Oral reading fluency was then assessed at the end of Grades 2–5. In this prospective, longitudinal study, predictive validity was estimated for the Caucasian and Hispanic students by examining the amount of variance the static and dynamic

assessments explained and by referring to area under the curve and sensitivity and specificity values.

Results: The dynamic assessment accounted for variance in reading ability over and above the static measures, with fair to good area under the curve values and sensitivity and specificity. Classification accuracy worsened when the static measures were included as predictor measures. The results of this study indicate that a very brief dynamic assessment can predict with approximately 75%–80% accuracy, which kindergarten students will have difficulty in learning to decode up to 6 years into the future.

Conclusions: Dynamic assessment of decoding is a promising approach to identifying future reading difficulty of young kindergarten students, mitigating the cultural and linguistic bias found in traditional static early reading measures.

A highly diverse population of students in the United States enter kindergarten each year with varied learning experiences and abilities. Many of these students are unprepared. They have not yet experienced much, if any, formal reading instruction; they are unfamiliar with assessment contexts and procedures; and the expectations of schools are often misaligned with their cultural or linguistic backgrounds (Duncan et al., 2007; Evans, 2004; Isaacs & Magnuson, 2011; Klinger & Edwards, 2006; U.S. Department of Education, Early Learning, 2015; Waldfogel & Washbrook, 2011). Because of this, the majority of these young kindergarten students perform poorly on single time, static reading assessments, yielding considerable

floor effects (Catts, Petschner, Schatschneider, Sittner-Bridges, & Mendoza, 2009; Johnson, Jenkins, Petscher, & Catts, 2009). Consequently, most static reading measures used to assess young students have poor classification accuracy (Badian, 1994; Catts, 1991; Jenkins & O'Connor, 2002; Mantzicopoulos & Morrison, 1994; O'Connor & Jenkins, 1999; Scarborough, 1998; Torgesen, 2002a, 2002b; Uhry, 1993; Wilson & Lonigan, 2009). When used to measure the reading abilities of culturally and linguistically diverse children just entering the school system at the beginning of kindergarten, static measures demonstrate even lower classification accuracy (Donovan & Cross, 2002; Gersten & Dimino, 2006; Klinger & Edwards, 2006).

The Hispanic, Spanish-speaking population in the United States is growing at a rapid rate, accounting for over half of the population increase in the United States between the years 2000 and 2014 (U.S. Census Bureau, 2017). These students are particularly vulnerable to assessment practices that discount their linguistic diversity and their diverse culture, background, and experiences. The use

^aDepartment of Communication Disorders, Brigham Young University, Provo, UT

^bSweetwater County Child Development Center, Rock Springs, WY

^cDepartment of Child and Family Studies, University of South Florida, Tampa

Correspondence to Douglas B. Petersen: dpeter39@byu.edu

Editor-in-Chief: Tiffany Hogan

Received January 29, 2018

Revision received May 3, 2018

Accepted July 9, 2018

https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021

Publisher Note: This article is part of the Clinical Forum: Dyslexia.

Disclosure: Douglas B. Petersen and Trina D. Spencer are entitled to royalties from sales of the Predictive Early Assessment of Reading and Language (PEARL), a dynamic assessment of reading and language similar to that used in this research study. Shelbi L. Gragg has declared that no competing interests existed at the time of publication.

of static measures can lead to costly interventions being provided to far more students than necessary, diffusing and misdirecting finite educational resources.

Because of the floor effects inherent in static assessments, final judgment as to whether a student has a reading disability, such as dyslexia, is often reserved until a profile of limited response to intervention (RTI) is documented over time (Catts, Nielsen, Bridges, Liu, & Bontempo, 2015; Fuchs & Fuchs, 2006; Vaughn & Fuchs, 2003; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). Using this approach, multiple test administrations are needed, with reading instruction provided between assessments to determine who needs intervention and who does not. It stands to reason that the diagnosis of dyslexia, which is characterized by persistent difficulty in learning to automatically recognize written words (or to decode, generally defined; Fletcher & Lyon, 2008; Lyon, Shaywitz, & Shaywitz, 2003; Peterson & Pennington, 2012; Snowling, 2012), should be made using an assessment approach that examines a student's ability to learn to read. Because of the construct alignment (Anastasi & Urbina, 1997) between RTI, which measures learning over time, and dyslexia, which is a learning disability, RTI has been touted as a more valid approach to the identification of learning disabilities (Fuchs & Fuchs, 2006). The RTI process also mitigates the cultural and linguistic bias found in static assessments, because the focus is on the learning process as opposed to current knowledge, which can be confounded by multiple factors, including socioeconomic status, English language learning status, and prior instruction or exposure. Despite the benefits of reducing test bias and decreasing overidentification, RTI takes considerable time and resources (Al Otaiba et al., 2014; Compton et al., 2012; Vaughn, Denton, & Fletcher, 2010). A lengthy RTI process is incompatible with the urgency around identifying struggling students as soon as they enter school so that intensive intervention can prevent early reading difficulties, including dyslexia, from developing into life-long reading problems. (Elbro & Petersen, 2004; Galuschka, Ise, Krick, & Schulte-Körne, 2014; Partanen & Siegel, 2014; Wanzek & Vaughn, 2007).

Dynamic Assessment

Dynamic assessment is an alternative testing process that is focused on how well a student can learn something new, as opposed to what a student currently knows (Feuerstein, Rand, & Hoffman, 1979; Haywood & Tzuriel, 1992; Lidz, 1995; Tzuriel, 2013). Dynamic assessment is similar to RTI in that it yields information from a process that combines assessment and intervention to identify disability, yet it takes much less time (Grigorenko, 2009; Lidz & Peña, 2009). Although there are a variety of approaches to dynamic assessment, a student is usually first assessed in an area that is likely to be outside their current independent functioning and then explicitly taught that deficit skill during a teaching phase with the purpose of identifying how much support the student needs to be successful. The student's responsiveness to this instruction is often

measured using a Likert scale or by examining the number of prompts needed to help a student achieve mastery on the skill being taught (Haywood & Tzuriel, 2002; Peña, Gillam, & Bedore, 2014; Petersen, Allen, & Spencer, 2016; Petersen, Chanthongthip, Ukrainetz, Spencer, & Steeve, 2017; Petersen & Gillam, 2015). Finally, a posttest is often administered. Through this brief cycle of teaching and testing, dynamic assessment, just like RTI, can be a less biased assessment approach (Laing & Kamhi, 2003). It can help educators determine a student's zone of proximal development (Lidz, 1995; Vygotsky, 1978), which yields prognostic information about what the student is capable of with proper support. Dynamic assessments result in multiple variables that can be scored and analyzed, including pretest performance, posttest performance, gain scores from pretest to posttest, duration of the teaching phase, and measures of modifiability. Modifiability refers to how responsive a student is during the teaching phase of the dynamic assessment (Cho & Compton, 2015; Feuerstein et al., 1979; Haywood & Tzuriel, 2002; Peña et al., 2014). Modifiability is most often measured using one or more Likert scales that require the examiner to reflect on their teaching experience with the student. Examiners ultimately determine how difficult it was for the student to learn the task being taught. Measures of modifiability, or responsiveness, are tightly connected to the construct of interest—learning potential (Feuerstein et al., 1979).

Dynamic assessment may be able to identify students with dyslexia more efficiently than the RTI process and may do so more accurately than static measures. Dynamic assessment could accomplish this by predicting how well young prereaders will respond to future reading instruction—effectively identifying students with latent specific word learning difficulty. However, there is limited research on the dynamic assessment of reading, and although dynamic assessment has a long history in psychology, its use in mainstream education is nearly absent (Elliott, 2003). For dynamic assessment of reading to be viable in schools, strong evidence of predictive validity must be established.

Dynamic Assessment of Reading

Caffrey, Fuchs, and Fuchs (2008) reviewed 24 studies published from 1974 to 2000 to examine the predictive validity of dynamic assessments of reading, mathematics, and expressive language. The results of the review indicated that dynamic assessment can explain significant, unique variance over and above traditional static academic and cognitive assessments for students with disabilities. The Caffrey et al. review provided preliminary evidence of the predictive validity of dynamic assessment in an educational context. Since that review, a number of studies have specifically investigated dynamic assessments of reading.

Sittner-Bridges and Catts (2011) and Gellert and Elbro (2017a) found that dynamic assessments of phonological awareness uniquely predicted variance over and above static measures of reading for beginning kindergarten students. Sittner-Bridges and Catts also reported that their dynamic

assessment identified reading difficulty at the end of kindergarten with approximately 86% sensitivity and 63% specificity. *Sensitivity* refers to how well a test correctly identifies children with a disorder, and *specificity* refers to how well a test correctly identifies those without a disorder. For an assessment to classify children well, it must have sufficient (e.g., 70%–80%) sensitivity and specificity (Cicchetti, 1994; Plante & Vance, 1994; Spaulding, Plante, & Farinella, 2006). Thus, this particular dynamic assessment of phonological awareness had less than desirable specificity (63%). Sittner-Bridges and Catts also conducted receiver operating characteristic (ROC) analyses that provided classification accuracy information using area under the curve (AUC) estimates. AUC provides true-positive (sensitivity) and false-positive (1 – specificity) rates for all possible cut points, offering the breadth of predictive accuracy. Values from the AUC analyses range from .50 to 1.00. Models that have an AUC of .50 provide prediction no better than chance, and models with an AUC of 1.00 have perfect prediction. AUC values greater than .70 are considered to be fair, and values greater than .80 are considered good (Swets, 1988). Sittner-Bridges and Catts reported AUCs ranging from .67 to .80 for their dynamic assessment, indicating poor to good prediction. Gellert and Elbro also reported classification accuracy of their dynamic assessment using AUC values. They found that their dynamic assessment of phonological awareness resulted in an AUC of .69, which is poor to fair. In addition, they followed their kindergarten participants until the end of first grade. They found that their dynamic assessment predicted reading performance up to the middle of first grade, but not to the end of first grade. Although not ideal, the classification results of these kindergarten dynamic assessments of phonological awareness reported by Sittner-Bridges and Catts (2011) and Gellert and Elbro (2017a) are superior to those typically found with kindergarten static measures of reading (e.g., Catts et al., 2009).

Instead of focusing on phonological awareness, researchers have also examined the predictive validity of dynamic assessments that focus on decoding and word identification, using nonsense words, hypothesizing that a tighter construct alignment will yield stronger predictive validity. Research on dynamic assessments of decoding by Fuchs et al. (2007) and Cho, Compton, Fuchs, Fuchs, and Bouton (2014) examined the predictive validity of a 20- to 30-min dynamic assessment of decoding that used nonsense words. During the teaching phase of their dynamic assessment, examiners focused on modeling nonsense words and providing instruction on onset–rime sound identification and blending. With primarily first-grade English-speaking students, these researchers found that their dynamic assessment was a significant predictor of response to 11 and 14 weeks of intervention. Their dynamic assessment uniquely accounted for variance in reading performance over and above traditional static measures of reading. Using the same dynamic assessment approach as Fuchs et al. (2007) and Cho et al. (2014), Fuchs, Compton, Fuchs, Bouton, and Caffrey (2011)

followed 318 first-grade students from the beginning to the end of the school year and found that this dynamic assessment accounted for unique variance over and above static predictor measures at the end of first grade. The dynamic assessment of decoding was a significant predictor of response to year-long reading instruction, yet information on sensitivity and specificity was not reported.

Other researchers have investigated dynamic assessments of decoding from the beginning of first grade to the end of first grade using novel orthography (e.g., Chinese, Hebrew, and arbitrary shapes). These studies have focused on evidence of concurrent validity and construct validity (Cho & Compton, 2015; Gellert & Elbro, 2017b), as well as on the extent to which dynamic assessment accounts for variance over and above traditional static measures, including phonological awareness tasks, static rapid automatized naming tasks, and a student's current reading ability (Aravena, Tijms, Snellings, & van der Molen, 2016; Cho & Compton, 2015; Cho et al., 2017; Gellert & Elbro, 2017b). These studies found that dynamic assessment of decoding measured a reading-specific learning construct separate from intelligence and vocabulary and consistently accounted for significant variance over and above traditional static reading measures. In addition, Aravena et al. (2016) reported classification accuracy of dynamic assessments of decoding using indexes of sensitivity and specificity and AUC values. Aravena et al. (2016) found that a 30-min dynamic assessment had 73% sensitivity and 61% specificity in identifying dyslexia in 118 Dutch-speaking children who were 7–11 years old. They reported that the AUC results were .74, which are similar to findings reported by Sittner-Bridges and Catts (2011) in their investigation of a dynamic assessment of phonological awareness. Gellert and Elbro (2017b) also provided AUC estimates and indexes of sensitivity and specificity for their 15-min dynamic assessment of decoding. They reported that the AUC was .85, with the static measures of phoneme synthesis and letter knowledge adding significantly (5%) to the prediction of reading at the end of first grade. For the dynamic assessment alone, sensitivity was 80% and specificity was 71%.

Although studies investigating the predictive validity of dynamic assessments of reading have reported generally positive results, there are a few shortcomings worth noting. First, accounting for unique variance is important to establish, yet more clinically translatable findings, such as classification accuracy, have either not been documented or have been inadequate in most studies. Second, the participants' demographics in most studies lacked linguistic diversity, despite dynamic assessment being particularly helpful for classifying culturally and linguistically diverse children (Peña et al., 2014; Petersen et al., 2016, 2017). Monolingual, English-speaking students do not typically experience the test bias that culturally and linguistically diverse students do with static measures (Laing & Kamhi, 2003). Third, although many studies have incidentally measured modifiability through graduated prompting, none of the dynamic

assessment of reading studies has directly measured modifiability or how easy students respond to effective mediation (i.e., intervention), which is considered vital to the dynamic assessment process (Cho & Compton, 2015; Feuerstein et al., 1979; Haywood & Tzuriel, 2002; Peña et al., 2014). A student's modifiability may be most helpful in predicting later achievement because it most closely reflects the concept of RTI and is independent of past learning experiences (Peña, 2000). In fact, drawing on research studies on dynamic assessments designed to differentiate language differences from language disorder, measures of modifiability have emerged as the most predictive variables derived from the assessment (Peña et al., 2006, 2014; Petersen et al., 2016, 2017; Petersen & Gillam, 2015). A fourth consideration is that most participants have been first graders and have not been followed past the end of first grade. By first grade, most students have already received formal decoding instruction, and the floor effects observed with static reading measures have diminished (Catts et al., 2009); therefore, the predictive validity of dynamic assessment over static measures may not be as pronounced. Also, examining predictive validity over a relatively short time may not be long enough to fully observe a dynamic assessment's predictive power, especially when predicting from the beginning of kindergarten. Finally, many of the dynamic assessments used in previous research have taken considerable time to administer, which reduces their clinical utility when applied as a universal screener for young students across a school district.

Current Study

Petersen et al. (2016) conducted a follow-up study to Petersen (2010) and Petersen and Gillam (2015) to compare the predictive validity of static reading measures (i.e., Dynamic Indicators of Basic Early Literacy Skills [DIBELS] Letter-Naming Fluency [LNF] and First Sound Fluency [FSF]; Good & Kaminiski, 2010) and dynamic measures of decoding administered at the beginning of kindergarten. The participants of the study were diverse; specifically, 300 Caucasian and 300 Hispanic students (with approximately 75% first language [L1] Spanish-speaking, English language learners) attending school in a large district in Utah. At the beginning of kindergarten, participants were administered the two static reading measures and one of two dynamic assessments that took approximately 3 min to administer. During the teaching phase of the dynamic assessments, one group of students was taught an onset-rime decoding strategy, and the other group was taught to use a sound-by-sound decoding strategy. Both dynamic assessments included a measure of modifiability at the end of the teaching phase. At the end of first grade, these same students were administered a norm-referenced assessment of reading.

Because the two dynamic assessments did not differ significantly when compared to each other, their results are summarized together here. The dynamic assessment classified the participants more accurately than a composite of the two static measures, with over 80% sensitivity and

specificity. The composite of the static assessments had 69%–79% sensitivity and 50%–51% specificity. Classification accuracy of the dynamic assessment for Hispanic students was lower, but still adequate, with sensitivity above 85% and specificity at or above 70%. The static measures had very low specificity with the Hispanic students. The results of this study indicated that the dynamic assessment used in Petersen et al. (2016) was an accurate and efficient early identification measure of decoding at the end of first grade for all of the kindergarten participants, including those who were culturally and linguistically diverse.

The evidence supporting predictive validity is growing, and the methodology is evolving. Classification of the dynamic assessment of decoding used in Petersen et al. (2016) most closely approximates the classification accuracy we would like to see for screening measures at the beginning of kindergarten. However, predictive validity that extends past first grade and into later childhood would be even more impactful and convincing. Therefore, the current study examines the extent to which the sensitivity and specificity documented in Petersen et al. (2016) remains adequate when students' decoding skills are tested in later elementary grades.

The primary purpose of this study was to examine and compare the classification accuracy of early static prereading measures and an early dynamic assessment of decoding administered to a large, diverse sample of students whose decoding abilities were examined at the beginning of kindergarten and at the end of every year when the students were in the second through fifth grades (Petersen et al., 2016, report the first-grade data of these students). Oral reading fluency scores obtained by the school district in Grades 2–5 served as the criterion measures for the current study. We asked the following questions for each of the two major subgroups: Caucasian, monolingual English-speaking White students and Hispanic, culturally and linguistically diverse students who could be of any race (White, Black, etc.):

1. (a) How much variance does a kindergarten dynamic assessment of decoding account for second through fifth grade decoding ability? (b) To what extent does a kindergarten dynamic assessment of decoding explain unique variance in second, third, fourth, and fifth grades decoding ability after static measures of letter naming and phonemic awareness are accounted for? (c) Do the combined static and dynamic measures account for variance over the dynamic assessment alone?
2. (a) How well does the dynamic assessment correctly classify students with and without decoding difficulty at each grade? (b) To what extent does the dynamic assessment contribute to classification accuracy when traditional static measures of phonemic awareness and letter identification are first accounted for? (c) Do the combined static and dynamic measures result in better classification accuracy over the dynamic assessment alone?

Method

The 600 kindergarten participants from Petersen et al. (2016) who were still in the same Utah school district at the end of their second, third, fourth, or fifth grades were included in this study. There were 376 original participants remaining at the end of second grade, 367 at the end of third grade, 373 at the end of fourth grade, and 379 at the end of fifth grade. Of those students, approximately 10% were classified by their schools as having a reading disorder at the end of each grade (38, 36, 35, and 36, respectively). Descriptive data collected from the school district for the students still attending school in the district at the end of fifth grade indicated that 45% were receiving English as a second language services, 49.6% were female, 56.1% were Hispanic (of any race), 35.1% were Caucasian (White, monolingual English-speaking students), 5.5% were American Indian/Alaska Native, 1.8% were Black, 0.5% were Asian, and 1% were Other. Of the Hispanic subgroup, 77% were classified as L1 Spanish-speaking, English language learners using parent self-identification and the Utah Language Proficiency Assessment. Analyses in this study primarily focused on the Caucasian students and the Hispanic students.

Procedure

At the beginning of kindergarten, participants were randomly assigned to receive one of two dynamic assessments of decoding, and all were administered the DIBELS Next LNF and FSF static assessments. These dynamic and static kindergarten predictor measures were administered individually and took approximately 3 min each to complete. Dynamic assessments were administered to all of the participants within a period of 3 days in the early fall of their kindergarten school year, and the DIBELS subtests were administered within 3 weeks of the administration of the dynamic assessments. At the end of second through fifth grades, the same participants were administered the Oral Reading Fluency subtest of the DIBELS Next assessment. This assessment served as a measure of decoding ability.

Kindergarten Dynamic Assessment

Both dynamic assessments entailed three distinct phases: (a) the pretest phase, which probed the kindergarten students' ability to decode four nonsense words; (b) the teaching phase, where the examiner explicitly taught each kindergarten student how to decode the nonsense words; and (c) the posttest phase, which once again assessed the students' ability to decode the same nonsense words displayed in a different order. The nonsense words used in the pretest, teaching, and posttest phases were all in consonant–vowel–consonant form, with the short vowel /ae/ and the final consonant /d/ consistent at the end of each word (i.e., *tad*, *zad*, *naad*, *kad*). Sounds used in English (e.g., /ae/) were included because the purpose of the assessment was to identify future difficulty in learning to

read English in U.S. schools, and the items in a dynamic assessment do not need to be sensitive to linguistic diversity because the focus is on modifiability. Both dynamic assessments included identical pretests and posttests, yet the strategy used for teaching word-level reading during the teaching phases differed, with one dynamic assessment using an onset–rime decoding strategy and the other dynamic assessment using a sound-by-sound decoding strategy. In the Petersen et al. (2016) study, results indicated that there were no statistically significant differences between the two dynamic assessments. Therefore, in the current study, the data from both dynamic assessments were analyzed together. Information on fidelity of administration for the dynamic assessments is outlined in Petersen et al. (2016).

Dynamic assessment scoring. Scores derived from the dynamic assessment included the total number of correct sounds and correct words read from the pretest and posttest. Students received 1 point for each correct sound read from the consonant–vowel–consonant pretest and posttest nonsense words, with a total of 12 points possible for sounds at pretest and posttest. Students also received 1 point for each word read (blended) correctly, with a total of 4 points possible for words at pretest and posttest. After the teaching phase of the dynamic assessment, examiners completed a student modifiability form, which used a 0–4 Likert scale that rated students on errors the student made, confidence the student exhibited, disruptions by the child, and rate of acquisition. The last item on the modifiability form asked the examiner to rate a student's overall ability to learn to decode (learning score), which was also scored using a Likert scale from 0 to 4 (where 4 represented *high modifiability* and 0 represented *low modifiability*). Finally, students were awarded points based on the extent to which there was evidence that the decoding strategy taught during the teaching phase was used. Strategy points ranged from 1 to 3 for each posttest word. Detailed dynamic assessment scoring procedures and interrater reliability are outlined in Petersen et al. (2016). The sum of scaled scores of posttest words, posttest sounds, learning, and strategy was used as a composite dynamic assessment score.

Kindergarten Static Assessment

The school district administered the DIBELS Next LNF and FSF subtests (Good & Kaminski, 2010) at the beginning of the academic year to each kindergarten student. For LNF, students were presented a page of uppercase and lowercase letters and were required to name as many letters as possible within 1 min. For FSF, the examiner said a series of words to each student and asked the student to say the first sound of each word. The total score was based on the number of correct responses produced in 1 min. DIBELS Next yields classification categories (i.e., intensive, strategic, and core) for LNF and FSF. The school district used local criteria to assign classifications. The DIBELS Core classification indicates that children are performing as expected. The DIBELS Strategic classification indicates that in order to be successful in classroom

instruction, children likely need additional help or intervention to read at grade level. The DIBELS Intensive classification indicates that children require substantial support to be successful readers (Cummings, Kennedy, Otterstedt, Baker, & Kame'enui, 2011). Students classified as intensive or strategic on the DIBELS subtests were labeled *at risk*, and children classified as core were labeled *not at risk*. These static kindergarten predictor variables were selected based on current practice in the school district, allowing for an authentic analysis of what is commonly administered at the beginning of kindergarten.

Second Through Fifth Grade Word-Level Reading Measures

The school district administered the DIBELS Next (Good & Kaminski, 2010) Oral Reading Fluency (DORF) at the end of second through fifth grades. DORF measures reading fluency by calculating the number of words read in 1 min minus any word errors. Thus, in this study, reading fluency was used as a general measure of decoding ability for second through fifth graders. Test-retest and interrater reliability for the DORF with late elementary students range from .92 to .97 (Tindal, Marston, & Deno, 1983). Correlational evidence of concurrent, criterion-related validity with the Woodcock-Johnson Psycho-Educational Battery assessment (Good, Kaminski, Simmons, & Kame'enui, 2001) is moderate (.65-.71). Students were classified as having decoding difficulty at the end of second through fifth grades if they scored at or below the seventh percentile on the DORF subtest according to local norms (American Psychiatric Association, 2013). Research has clearly established that fluency is highly related to decoding (Hudson, Lane & Pullen, 2005; Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). Word reading accuracy, which is the ability to recognize or decode words correctly, requires clear knowledge of the alphabetic principle and recalling of high-frequency words (Ehri & McCormick, 1998). Reading with automaticity requires the quick identification of those words (Kuhn & Stahl, 2000). Difficulty with reading accuracy and automaticity will negatively impact reading fluency. Children with dyslexia will inevitably have disproportionate difficulty with reading fluency when measured using an index of accuracy and rate. The hypothesized discrete subgroups of children with dyslexia, who are characterized by single deficits in phonological processes or naming speed or who have deficits in both areas (see the double-deficit hypothesis; Lovett, Steinbach, & Frijters, 2000; Manis, Doi, & Bhadha, 2000; Wolf, Bowers, & Biddle, 2000) will have reading fluency difficulty (Wolf & Bowers, 1999, 2000).

Results

Descriptive Statistics and Intercorrelations

Data were analyzed using the Statistical Package for Social Sciences (SPSS Version 24.0; IBM Corp., 2016).

Intercorrelations, means, and standard deviations of kindergarten dynamic and static assessment predictor variables and second through fifth grade reading outcomes are displayed in Table 1. All correlation coefficients were statistically significant ($p < .01$). Floor effects were evident with the kindergarten static predictor measures, with LNF having a positive skew of 1.45 and FSF having a positive skew of 1.25. The dynamic assessment variables and the reading outcome measures had approximately normal distributions according to visual analysis and skew. There were no meaningful outliers. A total of 378 students from the original cohort of 600 kindergarten students remained in the participating school district to the end of fifth grade. The original cohort was composed of 50% Hispanic students ($n = 300$), yet there were 212 (56%) Hispanic students remaining at fifth grade. This indicates that there was slightly greater attrition from the Caucasian students than the Hispanic students. We used a complete case analysis based on those students remaining in fifth grade, making no statistical corrections for missing data (Karahalios, Baglietto, Carlin, English, & Simpson, 2012).

Logistic Regression and ROC Results

The research questions were addressed using hierarchical logistic regression and by referring to the AUC and sensitivity and specificity indexes obtained from ROC analyses. Logistic regression reports the amount of significant variance accounted for by predictor variables, often using Nagelkerke pseudo- R^2 values. Logistic regression uses continuous predictor variables to predict a dichotomized dependent variable. In this study, reading ability was dichotomized with reading disorder characterized by performance at or below the seventh percentile on the DORF subtest. A continuous kindergarten dynamic assessment composite score was calculated by converting raw scores to z scores from correct words read at posttest, correct sounds read at posttest, the learning rating scale, and the total posttest strategy scores, converting those z scores to scaled scores, and then adding those scaled scores together. Continuous raw scores from the LNF and FSF subtests of the DIBELS were used as static kindergarten predictors. These same continuous dynamic and static kindergarten measures were used in the ROC analyses. AUC is referenced from each ROC analysis to provide a single value that represents the percentage of students correctly classified as having or not having a reading disorder by the predictor measures. From the multiple possible combinations of sensitivity and specificity obtained from each ROC analysis, we report sensitivity at or above 70% while maintaining specificity as close to 70% or higher whenever possible (Swets, 1988). Table 2 details the results of the logistic regression and ROC analyses for the Caucasian subsample, and Table 3 displays those results for the Hispanic subsample.

To determine how much variance the kindergarten dynamic assessment accounted for decoding ability (Question 1a), the dynamic assessment of decoding was entered in logistic regression Models 1, 3, 5, and 7 (see Tables 2

Table 1. Descriptive statistics and correlations between measures.

Measure	1	2	3	4	5	6	7	8	9	10
1. DA Post Sounds	—									
2. DA Post Words	.80	—								
3. DA Learning	.58	.53	—							
4. DA Strategy	.94	.93	.58	—						
5. LNF	.50	.60	.45	.58	—					
6. FSF	.49	.56	.43	.55	.60	—				
7. DORF Second Grade	.46	.48	.41	.50	.56	.41	—			
8. DORF Third Grade	.45	.46	.40	.49	.52	.39	.92	—		
9. DORF Fourth Grade	.44	.45	.38	.48	.52	.37	.90	.93	—	
10. DORF Fifth Grade	.42	.42	.38	.44	.52	.37	.89	.91	.94	—
<i>M</i>	7.44	1.46	2.76	6.24	10.00	7.64	88.87	101.45	119.89	129.16
<i>SD</i>	4.01	1.61	1.11	4.04	12.66	9.88	38.97	40.99	40.41	41.01
Minimum	0	0	0	0	0	0	11	5	10	12
Maximum	12	4	4	12	58	40	219	230	246	245
Skewness	-0.44	0.60	-0.41	0.18	1.45	1.25	0.20	0.19	-0.08	-0.24

Note. All coefficients are significant $< .01$. DA = dynamic assessment; LNF = DIBELS Letter-Naming Fluency; FSF = DIBELS First Sound Fluency; DORF = DIBELS Oral Reading Fluency; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

and 3) as the predictor measure, with second, third, fourth, or fifth grade reading (DORF) as the outcome measure. This provided specific information on the amount of variance explained by the dynamic assessment alone. For the Caucasian students, the dynamic assessment accounted for significant variance in second through fifth grade reading ability ranging from .15 to .23, and for the Hispanic students, significant variance ranged from .15 to .24. The

Wald chi-square test (Wald) was consistently $p < .01$ across grades.

Hierarchical logistic regression was conducted to examine the extent to which the kindergarten dynamic assessment explained unique variance in second-, third-, fourth-, and fifth-grade reading (DORF) over and above the kindergarten static measures (Research Question 1b). For the Caucasian subgroup (see Table 2) and the Hispanic

Table 2. Logistic regression and receiver operating characteristic analyses for static and dynamic kindergarten predictors and decoding performance at Grades 2–5 for Caucasian students.

Criterion measure	Model	Step	Predictor	Beta	Exp(B)	B_0	<i>b</i>	R^2	ΔR^2	χ^2	Wald	Sensitivity	Specificity	AUC
2nd Grade DORF	1	1	DA	.16	1.17	-1.80	.54	.23		14.17**	11.15**	.82 ⁺	.80 ⁺	.87
		2	LNF	.05	1.05			.16		9.94**	0.77	.78	.60	.78
	2	1	FSF	.04	1.04			.19	.03	11.75**	0.37	.77	.70	.81
		3	DA	.12	1.12	-1.26	.53	.26	.07	16.35**	4.23*	.87	.80	.88
3rd Grade DORF	3	1	DA	.14	1.15	-1.40	.53	.19		11.86**	9.86**	.81 ⁺⁺	.80 ⁺⁺	.85
		4	LNF	.05	1.05			.15		9.33**	0.69	.66	.70	.77
	2	1	FSF	.07	1.07			.19	.04	12.03**	0.93	.75	.70	.81
3		DA	.09	1.10	-.74	.52	.24	.05	15.03**	2.82	.89	.70	.86	
4th Grade DORF	5	1	DA	.12	1.13	-.84	.53	.15		8.99**	8.89**	.80	.80	.84
		6	LNF	.03	1.03			.12		7.26**	0.37	.78	.60	.76
	2	1	FSF	.09	1.09			.17	.05	10.69**	1.48	.75	.70	.81
3		DA	.07	1.07	-.08	.52	.20	.03	12.41**	1.66	.90	.70	.83	
5th Grade DORF	7	1	DA	.13	1.14	-.83	.53	.15		8.85**	7.84**	.79	.78	.81
		8	LNF	.07	1.08			.16		9.26**	1.22	.90	.56	.79
	2	1	FSF	.04	1.04			.18	.02	10.41**	0.37	.75	.67	.81
3		DA	.07	1.07	-.03	.52	.20	.02	12.04**	1.57	.86	.67	.82	

Note. The dynamic assessment cut points for sensitivity and specificity for all grades was 26 (the sum of dynamic assessment posttest words, posttest sounds, learning, and strategy scaled scores). A discriminant function analysis, for which the dynamic assessment met assumptions, resulted in *90% sensitivity and 75% specificity and **90% sensitivity and 73% specificity. Beta, Wald, and Exp(B) (odds ratio) are from the last step of each model. χ^2 degrees of freedom are equal to the number of predictors in each model. *b* = binomial probability; DORF = DIBELS Oral Reading Fluency; DA = dynamic assessment of decoding; FSF = DIBELS First Sound Fluency (phonemic awareness); LNF = DIBELS Letter-Naming Fluency (letter identification); AUC = area under the curve; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

* $p < .05$. ** $p \leq .01$.

Table 3. Logistic regression and receiver operating characteristic analyses for static and dynamic kindergarten predictors and decoding performance at Grades 2–5 for Hispanic students.

Criterion measure	Model	Step	Predictor	Beta	Exp(B)	B_0	b	R^2	ΔR^2	χ^2	Wald	Sensitivity	Specificity	AUC
2nd Grade DORF	1	1	DA	.15	1.15	-1.41	.54	.16		16.42**	12.97**	.78	.75	.79
		2	LNF	.30	1.35			.17		16.79**	2.74	.68	.65	.73
		3	FSF	.11	1.11			.20	.03	20.53**	1.24	.77	.60	.77
3rd Grade DORF	3	1	DA	.11	1.11	-1.07	.53	.25	.05	26.43**	5.38*	.79	.70	.84
		4	LNF	.16	1.18	-1.56	.54	.18		16.74**	12.73**	.74	.71	.81
		2	FSF	.15	1.16			.17	.04	16.11**	1.26	.78	.47	.75
4th Grade DORF	6	3	DA	.13	1.13	-1.25	.53	.25	.08	23.31**	6.37*	.74	.71	.82
		5	LNF	.14	1.14	-1.24	.53	.15		24.91**	12.64**	.77	.65	.79
		2	FSF	.09	1.09			.08		9.15**	1.40	.67	.50	.65
5th Grade DORF	8	3	DA	.11	1.11	-.915	.53	.18	.06	19.30**	6.52*	.73	.73	.80
		7	LNF	.20	1.22	-2.46	.55	.24		25.30**	17.40**	.79	.80	.85
		2	FSF	.09	1.09			.10		9.68**	1.09	.67	.55	.67
		3	DA	.07	1.07			.13	.03	13.52**	0.69	.78	.45	.71
				.17	1.19	-2.24	.54	.27	.14	28.30**	11.87**	.82	.80	.86

Note. The dynamic assessment cut points for sensitivity and specificity for all grades was 24 (the sum of dynamic assessment posttest words, posttest sounds, learning, and strategy scaled scores). Beta, Wald, and Exp(B) (odds ratio) are from the last step of each model. χ^2 degrees of freedom are equal to the number of predictors in each model. b = binomial probability; DORF = DIBELS Oral Reading Fluency; DA = dynamic assessment of decoding; FSF = DIBELS First Sound Fluency (phonemic awareness); LNF = DIBELS Letter-Naming Fluency (letter identification); AUC = area under the curve; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

* $p < .05$. ** $p \leq .01$.

subgroup (see Table 3), the traditional static kindergarten measures were entered into the model in a forward, step-wise manner (Steps 1 and 2 in Models 2, 4, 6, and 8). After letter identification (LNF) and phonemic awareness (FSF) had been entered into the model, the dynamic assessment of decoding was entered (Step 3). Across all grades (second through fifth) and ethnicities (Caucasian and Hispanic), the dynamic assessment accounted for additional variance after the traditional static measures had been considered. For the Caucasian students, this variance ranged from 2% to 7%, yet for third, fourth, and fifth grades, the Wald statistic for the dynamic assessment was not significant. For the Hispanic students, the dynamic assessment accounted for 5%–14% of significant variance in reading over and above the combined static measures. Across all grades, the Wald test was significant ($p < .05$ or $p < .01$) for the dynamic assessment.

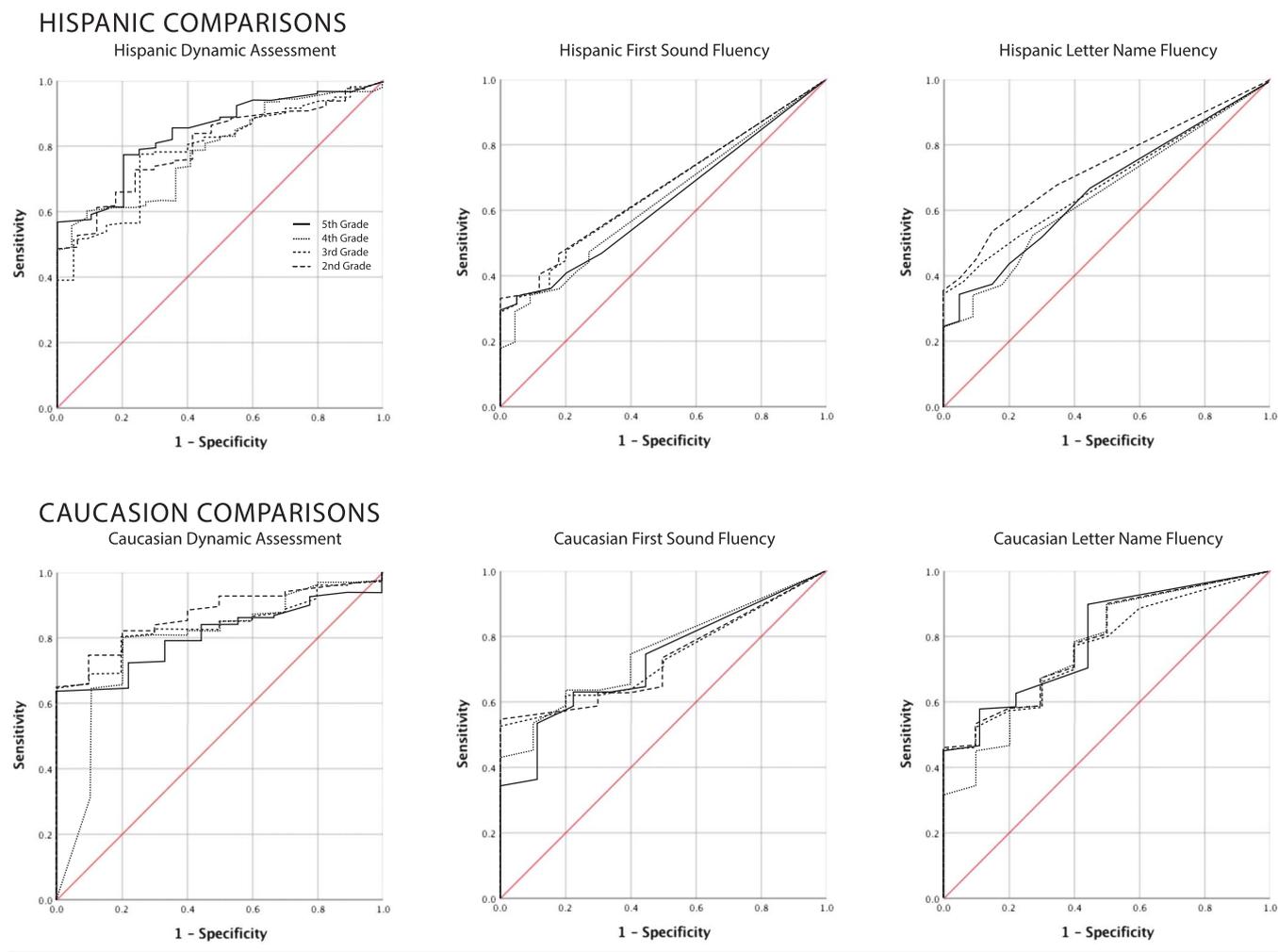
For Question 1c, we simply compared the pseudo- R^2 values from the combined static and dynamic measures at Step 3 to the pseudo- R^2 values from the dynamic assessment alone at Step 1. For the Caucasian students, the static measures explained 3%–5% ($M = 4.5\%$) of the variance in reading in Grades 2–5 over and above the dynamic assessment. For the Hispanic students, the static measures explained 3%–9% ($M = 5.5\%$) of the variance in reading in Grades 2–5 over and above the dynamic assessment.

Question 2a was focused on the classification accuracy of the dynamic assessment across Grades 2–5 for the Caucasian students and Hispanic students with and without reading disorders (± 7 th percentile on DORF). Figure 1 shows the results of the ROC analyses for the Caucasian and

Hispanic subgroups by predictor variable. ROC analyses for the dynamic assessment of decoding revealed AUCs for the Caucasian students ranging from .81 to .87 ($M = .84$). Sensitivity ranged from 79% to 82% ($M = 81\%$), and specificity ranged from 78% to 80% ($M = 80\%$; see Table 2). For the Caucasian students, a cut score of 26 (the sum of scaled scores from the four dynamic assessment variables) resulted in optimal sensitivity and specificity across grades. The dynamic assessment resulted in AUCs for the Hispanic students ranging from .79 to .85 ($M = .81$), with sensitivity ranging from 74% to 79% ($M = 77\%$) and specificity ranging from 65% to 80% ($M = 73\%$; see Table 3). For the Hispanic students, a cut score of 24 resulted in optimal sensitivity and specificity across grades.

ROC analyses resulted in AUC and sensitivity and specificity values for LNF (Step 1 in Models 2, 4, 6, and 8), LNF and FSF combined (Step 2), and LNF, FSF, and dynamic assessment combined (Step 3). These AUC and sensitivity and specificity data were examined to determine whether there was a meaningful improvement in classification accuracy from the dynamic assessment after the static kindergarten measures had been accounted for (Research Question 2b). For the Caucasian students, the AUC for the combined static measures (LNF + FSF, Step 2) was .81 across grades all grades. The addition of the dynamic assessment of decoding in Step 3 increased the AUC values to .88, .86, .83, and .82 in second through fifth grade, respectively. For the Hispanic students, the combined static measures offered AUC values ranging from .71 to .77 ($M = .73$). With the addition of the dynamic assessment to the static measures, AUCs ranged from .80 to .86 ($M = .82$). For the Caucasian students, sensitivity from the combined static

Figure 1. Receiver operating characteristic area under the curve results for Caucasian and Hispanic subgroups by predictor variable.



kindergarten measures in Step 2 ranged from 75% to 77% ($M = 76\%$), with specificity ranging from 67% to 70% ($M = 69\%$; see Table 2). When the dynamic assessment was combined with the static measures in Step 3, sensitivity increased to a mean of 85% and specificity increased to a mean of 72%. For the Hispanic students, sensitivity from the combined static kindergarten measures in Step 2 ranged from 77% to 78% ($M = 78\%$), with specificity ranging from 41% to 60% ($M = 48\%$; see Table 3). For the Hispanic students, sensitivity decreased with the addition of the dynamic assessment from a mean of 78% to a mean of 77%, yet specificity increased from a mean of 48% to a mean of 74%.

To answer Question 2c for the Caucasian and Hispanic students, we compared the sensitivity and specificity from the dynamic assessment alone (Models 1, 3, 5, and 7; Step 1) to the combined static and dynamic assessments (Models 2, 4, 6, and 8; Step 3). This comparison revealed that the addition of the static measures increased sensitivity

but decreased specificity. A similar pattern of results was found for the Hispanic students.

Discussion

The purpose of this study was to examine the predictive validity of a kindergarten dynamic assessment of decoding for second-, third-, fourth-, and fifth-grade reading and to determine whether the dynamic assessment improved predictive validity after traditional kindergarten static measures of reading were taken into account. Results indicated that the kindergarten dynamic assessment was a fair to good predictor of future reading difficulty for Caucasian and Hispanic students 6 years into the future. Several studies have reported the predictive validity of dynamic assessments of decoding, yet this is the first study to report on the extent to which the predictive validity of a dynamic assessment of decoding persists from kindergarten to the end of fifth grade. The addition of traditional

static measures of reading at kindergarten (i.e., letter identification and phonemic awareness) accounted for a small amount of variance over and above the dynamic assessment, yet they did not improve classification accuracy. In most cases, classification accuracy decreased when static measures were included in the analyses.

Variance

To answer our first research question, we examined how much variance a dynamic assessment of decoding accounted for decoding ability at the end of second, third, fourth, and fifth grades for Caucasian and Hispanic students. Our findings indicated that the dynamic assessment accounted for variance in decoding performance not only after several weeks, as has been documented by researchers such as Fuchs et al. (2007) and Cho et al. (2014), or to the end of first grade, as has been documented by researchers such as Cho et al. (2017), Fuchs et al. (2011), Gellert and Elbro (2017a), Petersen (2010), Petersen and Gillam (2015), and Petersen et al. (2016), but from kindergarten to the end of second, third, fourth, and fifth grades. These findings extend the research evidence by documenting the facility of dynamic assessment to account for future reading performance several years into the future.

We also examined the extent to which the dynamic assessment of decoding accounted for unique variance over and above traditional static measures of reading (Research Question 1b). Consistent with findings from previous research studies (e.g., Aravena et al., 2016; Cho & Compton, 2015; Sittner-Bridges & Catts, 2011), we found that the dynamic assessment accounted for variance in reading outcomes that traditional static measures do not explain, especially for the Hispanic subgroup. This additional variance was not particularly large in some cases, yet when stakes are high, accounting for even a small amount of variance can be important (Cho & Compton, 2015). There is clear research evidence that indicates that early identification and prevention are key to reducing severe consequences of reading disorders such as dyslexia (Elbro & Petersen, 2004; Galuschka et al., 2014; Partanen & Siegel, 2014; Torgesen, 2004). We used hierarchical logistic regression to answer this question, and the results of the analyses indicated that none of the static assessment variables were statistically significant according to the Wald test and that, in some cases for the Caucasian subgroup, the dynamic assessment variable was also not statistically significant.

In answer to Question 1c, we found that the static and dynamic measures combined accounted for variance over the dynamic assessment alone. This unique contribution to predicting reading ability from traditional kindergarten reading measures was evident out to the end of fifth grade. This aligns with findings from Gellert and Elbro (2017b), who also found that static kindergarten measures of letter identification and phonemic awareness accounted for unique variance over and above their dynamic assessment. This unique contribution from static measures was not identified in our previous study examining reading

ability at the end of first grade (Petersen et al., 2016). As has been recorded across multiple studies, both letter naming and phonemic awareness account for meaningful variance in reading ability (Hogan, Catts, & Little, 2005; Melby-Lervåg, Lyster, & Hulme, 2012; National Reading Panel, 2000; Stanovich, 1986).

Classification Accuracy

Accounting for variance is only one source of evidence of the predictive validity of an assessment. The correct classification of students with and without reading disorders, which can be evidenced using AUC and sensitivity and specificity values, has a more clinically meaningful interpretation. In fact, when filtered through the purpose of classification accuracy, the traditional static measures used in this study were not helpful.

Our second research question focused on classification accuracy as evidenced by AUC and sensitivity and specificity values. Our findings indicated that the dynamic assessment meaningfully increased the AUC and the sensitivity and specificity of the static measures for the Hispanic and Caucasian students. Although the AUC values from the static measures were generally good ($\geq .80$) for the Caucasian students and fair ($> .70$) for the Hispanic students, the dynamic assessment increased those AUC values, elevating them to at or above $.80$ across ethnicities and grades. This is an important finding given that the culturally and linguistically diverse student population in the United States is often disproportionately misidentified using static assessment procedures (Artiles & Trent, 2000), and the inherent bias found in static measures was mitigated to some degree through the dynamic assessment process.

Without the dynamic assessment, the static measures yielded sensitivity in the fair range (i.e., $< 70\%$, $< 80\%$), but specificity was poor ($< 70\%$). In general, when the static measures were added to the dynamic assessment, sensitivity increased, but specificity decreased. The dynamic assessment alone tended to yield more balanced sensitivity and specificity indexes; this is because the static measures, especially with the Hispanic population, had very poor specificity. This poor specificity is the result of floor effects from the static measures, where too many students were identified as being at risk for future reading difficulty. As has been documented in previous research, the static assessments over identified participants are at risk for decoding difficulty (Catts et al., 2009; Sittner-Bridges & Catts, 2011). Thus, although the combined static and dynamic measures accounted for more variance in reading ability at Grades 2–5, the static measures did not meaningfully improve the classification accuracy of the dynamic assessment. Although previous research has supported dynamic assessment as a supplemental measure to bolster static assessments (Fuchs et al., 2011; Sittner-Bridges & Catts, 2011), our data indicate that a dynamic assessment of decoding might be an appropriate independent assessment option and that static measures may decrease specificity due to their considerable floor effects.

Although 75%–80% classification accuracy from an early kindergarten dynamic assessment indicates fair to good results, these findings are more noteworthy when considering that the dynamic assessment only took 3 min to administer and that this prediction took place before the majority of students had begun formal reading instruction. This predictive, prognostic power has been the topic of recent studies, which have noted that dynamic assessments of decoding appear to measure learning very specific to the decoding process (Cho & Compton, 2015; Fuchs et al., 2011) and that dynamic assessment accounts for almost all variance explained by traditional static measures and then more (Aravena et al., 2016; Cho et al., 2017; Gellert & Elbro, 2017b). Because dynamic assessments appear to measure the ability to learn the very specific construct being assessed, it stands to reason that a dynamic assessment of decoding would best measure the ability to learn to decode. This appears to be the case, because dynamic assessments of phonological awareness, which measure a construct that is peripheral to the actual construct of decoding, tend to have weak predictive validity (Gellert & Elbro, 2017a; Sittner-Bridges & Catts, 2011). The variables derived from the dynamic assessment of decoding used in this study, which were the total correct posttest words, the total correct posttest sounds, the examiner's rating of a student's ability to learn to decode during the teaching phase, and an examination of the strategy the student used at posttest to decode the nonsense words, were all obtained after the students participated in very brief instruction on how to decode nonsense words. Thus, the focus of the assessment was on measuring a student's ability to learn something new, as opposed to measuring their current ability. This measurement of the specific decoding construct—and more specifically the measurement of a student's ability to learn the decoding process (a student's modifiability)—appears to serve as a moderately strong prognostic indicator of how well students will learn the very same decoding process in the coming years. Because dyslexia is a learning disability, an assessment that measures learning, such as a dynamic assessment of decoding, has strong construct validity. No other dynamic assessments of reading have included a direct measure of modifiability similar to those used in language disorder research. The learning rating scale used in the current dynamic assessment contributed significantly to the predictive validity of the dynamic assessment, as has been documented in studies on dynamic assessment of language (e.g., Peña et al., 2006, 2014; Petersen et al., 2017).

Clinical Implications

The results of this study, particularly from the classification findings, have important clinical implications. A dynamic assessment of decoding that takes 3 min to administer could be given to all incoming kindergarten students in a school district. Those students who perform poorly on the dynamic assessment could immediately receive more intensive reading instruction. Evidence is mounting to suggest that when students are identified early in their education

as being at risk for having future reading difficulty and intensive, evidence-based instruction is provided, much of the effects of a latent reading disorder can be mitigated (Ehri et al., 2001). Although certainly not a perfect predictor of future reading ability, dynamic assessment of decoding appears to have considerable promise. Static measures consistently overidentify young students as having reading difficulty, especially when those students are culturally and linguistically diverse (Gersten & Dimino, 2006; Klinger & Edwards, 2006). These floor effects from static measures, which were evidenced through considerable positive skewness and low specificity in the current study, often lead to the misappropriation of limited resources. Although RTI has been proposed as a viable alternative to traditional static approaches to the identification of reading disorders (Fuchs & Fuchs, 2006; Grigorenko, 2009; Vaughn & Fuchs, 2003), dynamic assessment of decoding can potentially identify students with a reading disorder at the beginning of kindergarten more accurately than static measures and more efficiently than the RTI process. At the very least, a dynamic assessment can sort students reasonably well right at the beginning of kindergarten, ensuring students who perform poorly receive immediate, intensive intervention to prevent their reading difficulties from worsening.

Limitations

There are some limitations to this study that should be noted. First, a diagnosis of dyslexia requires more than just evidence of repeated difficulty in learning to read fluently. Although we do not have confirmed diagnosis of dyslexia from the low-performing students in our sample, we do have evidence that the majority of the students that we classified as having a reading disorder consistently performed on a measure of decoding at or below the seventh percentile from Grades 2 through 5. Specifically, 94% of the students reading at or below the seventh percentile on the reading fluency assessment at the end of fifth grade also scored at or below the seventh percentile in second, third, and/or fourth grades. This is further evidence that the students classified as having a reading disorder in this study were indeed persistent limited responders to reading instruction. Although not definitive, this is reasonable evidence to suggest that these students are, at the very least, poor responders to the reading instruction provided in their schools and that, more than likely, these students have a reading disorder that is affecting their ability to decode fluently (Wolf & Bowers, 1999, 2000). Second, the dynamic assessment had more acceptable sensitivity and specificity than the static assessment, yet the accuracy for the Hispanic subsample did not meet the more preferred benchmark of $\geq 80\%$ (Plante & Vance, 1994). Further research is needed to identify what will increase the predictive validity of dynamic assessment of decoding, particularly for a culturally and linguistically diverse population. Findings from this study indicate that it is unlikely that traditional static measures of letter naming and phonemic awareness

are the answers. It could be that the current dynamic assessment of decoding does not contain enough items to fully reflect the decoding–learning construct. If the addition of items does not lengthen the duration of the test considerably, efficiency of dynamic assessment could be retained. Also, although the static predictor measures used in this study are commonly administered to kindergarten students across the United States, it could be that other static measures will more strongly predict future reading ability.

In this study, students were grouped based on ethnicity (Caucasian and Hispanic). Without question, there was heterogeneity within each of those subgroups, particularly within the Hispanic subgroup. Approximately 75% of the Hispanic students were English language learners. Although we did not obtain sufficient data to further disaggregate the subgroups by English language proficiency, this should be explored in future research. It is unclear whether the differences in classification accuracy between the Caucasian and Hispanic groups can be attributed to cultural differences, linguistic differences, or both. Similarly, replication of these research findings with an independent sample of students is needed in order to establish more confidence in the results obtained from this study. Nevertheless, it is very promising that a 3-min dynamic assessment of decoding administered at the beginning of kindergarten can predict future reading difficulty with approximately 75%–80% accuracy for all students, including those who are culturally and linguistically diverse.

Acknowledgments

We wish to thank the school district administration, teachers, and graduate and undergraduate students from the University of Wyoming, Brigham Young University, and Northern Arizona University for their contributions to this project.

References

- Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in tier I or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*, 11–27.
- Al Otaiba, S., Folsom, J. S., Schatschneider, C., Wanzek, J., Greulich, L., Meadows, J., & Connor, C. M. (2011). Predicting first-grade reading performance from kindergarten response to tier I instruction. *Exceptional Children, 77*, 453–470.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorder* (5th ed.). Arlington, VA: Author.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York, NY: Macmillan.
- Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2016). Predicting responsiveness to intervention in dyslexia using dynamic assessment. *Learning and Individual Differences, 49*, 209–215.
- Artiles, A. J., & Trent, S. C. (2000). Representation of culturally/linguistically diverse students. In C. R. Reynolds & E. Fletcher Jantzen (Eds.), *Encyclopedia of special education* (2nd ed., Vol. 1, pp. 513–517). New York, NY: Wiley.
- Badian, N. A. (1994). Preschool prediction: Orthography and phonological skills, and reading. *Annals of Dyslexia, 44*, 3–25.
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*(4), 254–270.
- Catts, H. W. (1991). Early identification of dyslexia: Evidence from a follow-up study of speech-language impaired children. *Annals of Dyslexia, 41*, 163–177.
- Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y., & Bontempo, D. (2015). Early identification of reading disabilities within a RTI framework. *Journal of Learning Disabilities, 48*, 281–297.
- Catts, H. W., Petschner, Y., Schatschneider, C., Sittner-Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Language, Speech, and Hearing Services in Schools, 42*(2), 163–176.
- Cho, E., & Compton, D. L. (2015). Construct and incremental validity of dynamic assessment of decoding within and across domains. *Learning and Individual Differences, 37*, 183–196.
- Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to tier 2 intervention. *Journal of Learning Disabilities, 47*, 409–423.
- Cho, E., Compton, D. L., Gilbert, J. K., Steacy, L. M., Collins, A. A., & Lindström, E. R. (2017). Development of first-graders' word reading skills: For whom can dynamic assessment tell us more? *Journal of Learning Disabilities, 50*(1), 95–112.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.
- Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., & Bouton, B. D. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*, 204–216.
- Cummings, K. D., Kennedy, P. C., Otterstedt, J., Baker, S. K., & Kame'enui, E. J. (2011). *DIBELS data system: 2010–2011 Percentile ranks for DIBELS next benchmark assessments* (Tech. Rep. 1101). Eugene: University of Oregon.
- Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.
- Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 14*(2), 135–164.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National reading panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.
- Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology, 96*, 660–670.
- Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*(1), 15–32.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*(2), 77–92.
- Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential*

- assessment device, theory, instruments, and techniques. Baltimore, MD: University Park Press.
- Fletcher, J. M., & Lyon, G. R. (2008). Dyslexia: Why precise definitions are important and how we have achieved them. *Perspectives on Language and Literacy*, 34, 27–31.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities*, 44, 339–347.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 92–99.
- Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness to intervention: A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children*, 39(5), 58–63.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS One*, 9(2), e89900.
- Gellert, A. S., & Elbro, C. (2017a). Does a dynamic test of phonological awareness predict early reading difficulties? A longitudinal study from kindergarten through Grade 1. *Journal of Learning Disabilities*, 50(3), 227–237.
- Gellert, A. S., & Elbro, C. (2017b). Try a little bit of teaching: A dynamic assessment of decoding as a kindergarten predictor of word reading difficulties at the end of Grade 1. *Scientific Studies of Reading*, 21(4), 277–291.
- Gersten, R., & Dimino, J. A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly*, 41(1), 99–108.
- Good, R. H., & Kaminski, R. A. (Eds.). (2010). *Dynamic Indicators of Basic Early Literacy Skills Next–Seventh Edition*. Eugene, OR: Dynamic Measurement Group.
- Good, R. H., Kaminski, R. A., Simmons, D., & Kame'enui, E. J. (2001). Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model: Steps to reading outcomes. *OSSC Bulletin*, 44(1), n1.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42(2), 111–132.
- Haywood, H. C., & Tzuriel, D. (Eds.). (1992). *Interactive assessment*. New York, NY: Springer-Verlag.
- Haywood, H. C., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education*, 77(2), 40–63.
- Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools*, 36, 285–293.
- Hudson, R., Lane, H., & Pullen, P. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, 58(8), 702–714.
- IBM Corp. (2016). IBM SPSS Statistics for Windows (Version 24.0) [Software]. Armonk, NY: Author.
- Isaacs, J., & Magnuson, K. (2011). *Income and education as predictors of children's school readiness*. Washington, DC: The Brookings Institution.
- Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–149). Mahwah, NJ: Erlbaum.
- Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*, 24(4), 174–185.
- Karahalios, A., Baglietto, L., Carlin, J., English, D. R., & Simpson, J. A. (2012). A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC Medical Research Methodology*, 12(96), 1–10.
- Klinger, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, 41(1), 108–117.
- Kuhn, M. R., & Stahl, S. A. (2000). *Fluency: A review of developmental and remedial practices*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.
- Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools*, 34, 44–55.
- Lidz, C. S. (1995). Dynamic assessment and the legacy of L. S. Vygotsky. *School Psychology International*, 16, 143–153.
- Lidz, C. S., & Peña, E. D. (2009). Response to intervention and dynamic assessment: Do we just appear to be speaking the same language? *Seminars in Speech and Language*, 30(2), 121–133.
- Lovett, M. W., Steinbach, K. A., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double-deficit perspective. *Journal of Learning Disabilities*, 33, 334–358.
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1–14.
- Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities*, 33, 325–333.
- Mantzicopoulos, P., & Morrison, D. (1994). Early prediction of reading achievement: Exploring the relationship of cognitive and noncognitive measures to inaccurate classifications of at risk status. *Remedial and Special Education*, 15(4), 244–251.
- Melby-Lervåg, M., Lyster, S-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from <http://www.nationalreadingpanel.org>
- O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3, 159–197.
- Partanen, M., & Siegel, L. S. (2014). Long-term outcome of the early identification and intervention of reading disabilities. *Reading and Writing*, 27(4), 665–684.
- Peña, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly*, 21(2), 87–97.
- Peña, E. D., Gillam, R. B., & Bedore, L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, 57, 2208–2220.
- Peña, E. D., Gillam, R. B., Malek, M., Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of children from culturally diverse backgrounds: Applications to

- narrative assessment. *Journal of Speech, Language, and Hearing Research*, 49, 1037–1057.
- Petersen, D. B.** (2010). *Using static and dynamic measures to estimate reading difficulty for Hispanic children* (Doctoral dissertation). Retrieved from <http://digitalcommons.usu.edu/etd/540>
- Petersen, D. B., Allen, M. A., & Spencer, T. D.** (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities*, 49(2) 200–215.
- Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W.** (2017). Dynamic assessment of narratives: Efficient and accurate classification of bilingual students with language impairment. *Journal of Speech, Language, and Hearing Research*, 60(4), 983–998.
- Petersen, D. B., & Gillam, R. B.** (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, 48(1), 3–21.
- Peterson, R. L., & Pennington, B. F.** (2012). Developmental dyslexia. *The Lancet*, 379, 1997–2007.
- Plante, E., & Vance, R.** (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25, 15–24.
- Scarborough, H. S.** (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York Press.
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J., & Stahl, S. A.** (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96, 119–129.
- Sittner-Bridges, M., & Catts, H. W.** (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities*, 44(4), 330–338.
- Snowling, M. J.** (2012). Changing concepts of dyslexia: Nature, treatment and comorbidity. *The Journal of Child Psychology and Psychiatry*, 53, e1–e3.
- Spaulding, T. J., Plante, E., & Farinella, K. A.** (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37, 61–72.
- Stanovich, K. E.** (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–406.
- Swets, J. A.** (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.
- Tindal, G., Marston, D., & Deno, S.** (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Torgesen, J. K.** (2002a). Empirical and theoretical support for direct diagnosis of learning disabilities by assessment of intrinsic processing weaknesses. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 565–613). Mahwah, NJ: Erlbaum.
- Torgesen, J. K.** (2002b). The prevention of reading difficulties. *Journal of School Psychology*, 40, 7–26.
- Torgesen, J. K.** (2004). Avoiding the devastating downward spiral: The evidence that early intervention prevents reading failure. *American Educator*, 28(3), 6–19.
- Tzuriel, D.** (2013). Mediated learning experience strategies and cognitive modifiability. *Journal of Cognitive Education and Psychology*, 13, 59–80.
- U.S. Census Bureau.** (2017). *School enrollment of the Hispanic population: Two decades of growth*. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2017/08/school_enrollmentof.html
- U.S. Department of Education, Early Learning.** (2015). *A matter of equity: Preschool in America*. Washington DC: Author. Retrieved from <https://www2.ed.gov/documents/early-learning/matter-equity-preschool-america.pdf>
- Uhry, J. K.** (1993). Predicting low reading from phonological awareness and classroom print. *Educational Assessment*, 1(4), 349–368.
- Vaughn, S., Denton, C. A., & Fletcher, J. M.** (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools*, 47(5), 432–444.
- Vaughn, S., & Fuchs, L. S.** (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research and Practice*, 18, 137–146.
- 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*, 21, 437–480.
- Vygotsky, L. S.** (1978). Interaction between learning and development (M. Lopez-Morillas, Trans.). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (pp. 79–91). Cambridge, MA: Harvard University Press.
- Waldfoegel, J., & Washbrook, E.** (2011). Income-related gaps in school readiness in the U.S. and U.K. In T. M. Smeeding, R. Erikson, & M. Jäntti (Eds.), *Persistence, privilege, and parenting: The comparative study of intergenerational mobility* (pp. 175–208). New York, NY: Russell Sage Foundation.
- Wanzek, J., & Vaughn, S.** (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36, 541–561.
- Wilson, S. B., & Lonigan, C. J.** (2009). Emergent literacy screeners for preschool children: An evaluation of get ready to read! and individual growth and development Indicators. *Annals of Dyslexia*, 59, 115–131.
- Wolf, M., & Bowers, P.** (1999). The “double-deficit hypothesis” for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 1–24.
- Wolf, M., & Bowers, P.** (2000). The question of naming-speed deficits in developmental reading disabilities: An introduction to the double-deficit hypothesis. *Journal of Learning Disabilities*, 33, 322–324.
- Wolf, M., Bowers, P., & Biddle, K.** (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities*, 33, 387–407.