

A Comparison of Reading Measures as Indicators of Risk across Grade Levels

Chuang Wang

Department of Educational Leadership, University of North Carolina at Charlotte
9201 University City Blvd, Charlotte, NC 28223, USA
Tel: 1-704-687-8859 E-mail: cwang15@uncc.edu

Bob Algozzine (Corresponding author)

University of North Carolina at Charlotte
9201 University City Blvd, Charlotte, NC 28223, USA
Tel: 1-704-687-8859 E-mail: rfalgozz@uncc.edu

Received: October 12, 2011

Accepted: November 7, 2011

Published: December 30, 2011

doi:10.5539/jedp.v1n1p85

URL: <http://dx.doi.org/10.5539/jedp.v1n1p85>

Support for this research was provided in part by Grant No. H238X00001 from the U.S. Department of Education, Office of Special Education Programs, awarded to the University of North Carolina at Charlotte. The opinions expressed are those of the authors and do not reflect a position, policy, or endorsement from the Department of Education.

Abstract

The purpose of this study was to assess relationships between early literacy skills of at-risk students assessed with curriculum-based and norm-referenced measures. Participants were students in first, second, and third grade. Trained examiners administered five curriculum-based reading passages and three norm-referenced reading subtests to each student. Strong relationships were evident between performance on the curriculum-based oral reading fluency assessments and the norm-referenced reading subtests of word attack, word identification and passage comprehension skills. Findings are discussed with regard to previous research and recommendations are provided for helping teachers and other professionals engaged in efforts to use assessment information in response to intervention as well as in targeting instructional support needed in reading.

Keywords: Curriculum-based measurement, Fluency assessment, At-risk students

1. Introduction

More than 25 years ago, researchers collaborating at the University of Minnesota simultaneously demonstrated that assessment practices for identifying students with learning disabilities were seriously flawed and that the use of curriculum-based rather than norm-referenced assessments held promise in efforts to improve their educational lives (Deno, 1985; Deno, Mirkin, & Chiang, 1982; Shinn, 1989; 1998; Ysseldyke, 2005; Ysseldyke & Algozzine, 1979; Ysseldyke, Thurlow, Graden, Wesson, Algozzine, & Deno, 1983). Position papers, scholarly reviews, and reports of research on curriculum-based measurement have consistently proposed its practical value and documented its technical adequacy in efforts to monitor progress in core academic skills (Christ, 2006; Christ & Silbergliitt, 2007; Marston, 1989; Reschly, Busch, Betts, Deno, & Long, 2009; Wayman, Wallace, Wiley, Tichá, & Espin, 2007; Woodward & Brown, 2006). Recently, a continuing "...identity crisis or, perhaps more accurately, an identification crisis..." has driven researchers, practitioners, and policy-makers to "response to intervention and progress monitoring" in their efforts to improve assessment and identification practices related to meeting the needs of the very large number of students at risk of failure in America's schools, especially those likely to be classified with learning disabilities (Case, Speece, & Mollow, 2003, p. 557).

According to Vaughn and Fuchs (2003), efforts to improve identification of students with learning disabilities grew from a National Research Council study (Heller, Holtzman, & Messick, 1982) that focused on improving

placement practices in special education. Response to intervention (RTI) initiatives have emerged as the promising practice being advocated and embraced as an improved approach for identifying children with learning disabilities (Fuchs & Fuchs, 2006). In this context, regular monitoring of progress provides evidence of the extent to which instructional environments are promoting learning and students are not learning as hallmarks of the approach (Vaughn & Fuchs, 2003). For a problem-solving approach like RTI to be effective, especially for at-risk students, practitioners need continuing evidence that supports the accuracy of predictive relationships between progress monitoring measures and outcome performance in key academic areas (Catts, Petscher, Schatschneider, Sittner Bridges, & Mendoza, 2009; Jenkins, Hudson, & Johnson, 2007; Johnson, Jenkins, & Petscher, 2010; Johnson, Jenkins, Petscher, & Catts, 2009; Reschly et al., 2009; Speece, Ritchey, Silverman, Schatschneider, Walker, & Andrusik, 2010).

Comprehensive analyses and reviews (cf. Marston, 1989; Reschly et al., 2009; Shinn, 1989; 1998; Wayman et al., 2007) have supported the technical adequacy and practicality of curriculum-based measures (CBM). For example, Hosp and Fuchs (2005) investigated the extent to which relationships between CBM and traditional reading assessment scores (e.g., Woodcock Reading Mastery Test-Revised) differed as a function of grade level; they also explored how well cut scores served as benchmarks for mastery on summative measures. Their study included students ($n = 310$) in first through fourth grade in four schools. Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Wanzek, et al. (2010) evaluated the validity and usefulness of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (ORF) in predicting performance on the summative end-of-grade assessments; and, the strongest correlations were for third grade. Wayman et al. (2007) reviewed 160 documents addressing reading, writing, and math CBM in kindergarten (K) to Grade 12 and concluded that the data supported the value of the CBM reading-aloud measure for use by classroom teachers as an indicator of the performance and progress in reading for elementary school students, Grades 2 to 5.

Accepting that replication is a hallmark of science, we reasoned that continued study of relationships between CBM and academic skills was warranted especially in light of shifts in use from one of monitoring progress of students receiving special education to one more closely linked to decision-making with different educational consequences for students placed at risk (i.e., placement in special education). We followed the lead of others in documenting relationships at different grades and reporting the extent to which curriculum-based cut-off scores corresponded with benchmark performance on norm-referenced and other measures (cf. Hintze, Owen, Shapiro, & Daly, 2000; Hintze & Silberglitt, 2005; Hosp and Fuchs, 2005; McGlinchey & Hixson, 2004; Reschly et al., 2010; Schilling, Carlisle, Scott, & Zeng, 2007; Silberglitt, Burns, Madyun, & Lail, 2006; Silberglitt & Hintze, 2007; Wanzek, Roberts, Linan-Thompson, Vaughn, Woodruff, & Murray, 2010; Wayman et al., 2007).

We focused on descriptive and predictive analyses of use in high-stakes decisions that carry important social consequences for young students from diverse, at-risk backgrounds. Specifically, we addressed three research questions:

- 1) To what extent are scores for early literacy skills of at-risk students assessed with CBM statistically similar to those for norm-referenced measures across grades (descriptive validity)?
- 2) To what extent are scores for early literacy skills of at-risk students assessed with CBM statistically related to those for norm-referenced measures across grades (concurrent validity)?
- 3) To what extent are scores for early literacy skills of at-risk students on CBM statistically predictive of performance on norm-referenced measures across grades (predictive validity)?

2. Method

In this study, we investigated the state and usefulness of CBM indicators as benchmarks of performance on norm-referenced subtests of critical reading skills. We reasoned that such an investigation contributes to the existing literature in two important ways. First, most prior research comparing CBM with assessments of specific reading skills focused mostly on predictive relationships with comprehension. Our study expands previous work by exploring the value of CBM as predictors of specific reading skills, including, but not limited to, comprehension. Second, adding to what is known about curriculum-based and norm-referenced measurement provides a clear replication path in formative efforts to use assessment to improve instruction, such as in those defining the promises of RTI.

2.1 Participants

The participants were 475 first grade, 360 second grade, and 181 third grade students from seven public schools in an urban, integrated school system enrolling more than 120,000 students each year in the southeastern region of the United States. The ethnic backgrounds of students in the district were rich and diverse including

African-American (43%), American Indian/Multi Racial (3%), Asian (4%), Caucasian (40%), and Hispanic (10%) groups. The children in this study were in seven elementary schools representative of those serving large numbers of students at risk for failure in the participating district, other states, and the nation. The percent receiving free or reduced lunch was also generally high for students participating in our study. Additional demographic information regarding these students' gender, ethnicity, social economic status represented by school lunch program, and special education status is reported in Table 1. None of these characteristics were statistically significantly different across the grade levels as indicated by chi-square tests. As a result, demographic variation across these variables was not considered in our subsequent analyses.

2.2 Procedures

One CBM set was collected three times during the school year as part of district-mandated formative and summative assessment practices. For research purposes, another CBM and norm-referenced subtests were administered at the end of the year. The specific assessments were identified in collaboration with district-level administrators interested in supporting the overall research effort while minimizing the time that participating students were being tested. All students were tested with the same battery of tests.

2.2.1 Curriculum-based Measures

Teachers trained by the school district administered and used the *Oral Reading Fluency* (ORF) subtest of *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) for progress monitoring benchmark assessments of oral reading proficiency (cf. Good & Kaminski, 2003). DIBELS ORF was used for K-2 benchmark, program monitoring, and end-of-year assessments within the district. It is a standardized, individually administered test of accuracy and fluency with connected text. The passages and procedures are designed to (a) identify children who may need additional instructional support, and (b) monitor progress toward instructional goals. The passages are calibrated for the goal level of reading for each grade level. Student performance is measured by having students read a passage aloud for 1 minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The number of correct words per minute from the passage is the oral reading fluency rate. In general, the DIBELS assessments have excellent technical adequacy (cf. Elliott, Lee, & Tollefson, 2001; Fuchs, Fuchs, & Compton, 2004; Good & Kaminski, 2002, 2003; Good, Kaminski, Simmons, & Kame'enui, 2001; Hintze, Ryan, & Stoner, 2003; Speece, Mills, Ritchey, & Hillman, 2003; Vadasy, Sanders, & Peyton, 2005).

At the end of the school year, teachers also administered two additional curriculum-based reading passages developed by Fuchs and Fuchs (1992). Each student was graded according to the average number of words read correctly in 1 minute across the two passages. Words read correctly were those pronounced accurately while reading the passage. Repetitions and self-corrections within 3 seconds were counted as correct. Errors were substitutions, omissions, and hesitations of more than three seconds. Previous studies have shown that this measure is a reliable and valid indicator of oral reading fluency in young children (Deno et al., 1982; Fuchs, Fuchs, & Maxwell, 1988; Hosp & Fuchs, 2005).

2.2.2 Norm-referenced Measures

Trained research assistants supervised by a school psychologist administered subtests of The Woodcock Reading Mastery Test-Revised (WRMT-R: Woodcock, 1998) at the end of the school year. The WRMT-R is an individually-administered norm-referenced measure of reading achievement and we used three of its six subtests in our study. The Word Attack subtest requires the student to read nonsense words (i.e., letter combinations that are not actual words) or low-frequency words from English. Each form of this subtest includes 45 items arranged in order of difficulty to measure students' ability to use phonic and structural analysis skills and knowledge. The Word Identification subtest requires students to identify isolated words in a text. Each form of this subtest has 108 items. The Passage Comprehension subtest measures students' ability to read a short passage (approximately 2 to 3 sentences) and identify the key word missing from the passage. To produce the correct response, the student needs to understand not only the sentence in which the word is missing, but also the entire passage. Each form of this subtest includes 68 items arranged in order of difficulty. We used decoding, word reading, and comprehension as labels for scores on these measures in our research. Grade-based standardized scores with a mean of 100 and standard deviation of 15 were used for each of these subtests which all have strong reliability and validity (Hosp & Fuchs, 2005; Woodcock, 1998).

2.3 Design and Data Analysis

Our study was a descriptive comparison of reading skills of children at risk for failure. First, we inspected the distribution of each measure but did not notice significantly skewed performances. As a result, we used the

original test scores without corrective transformations. The average of the two CBM scores, the average of the three ORF scores, and the grade-based standardized scores for each of the three subtests of WRMT-R were used in data analyses. We reasoned that the mean was the best overall marker of performance on assessments provided on measures available on multiple administrations (i.e., ORF). Second, significant differences in the correlations between CBM and the three subtests of WRMT-R were tested with z -tests by transforming r values into z values across the grades and t -tests within each grade (Blalock, 1960). Finally, predictive discriminant analysis (PDA) was used to measure classification accuracy or how well CBM predicted the students' performance on WRMT-R subtests. These procedures have been used in previous research to identify the CBM cut-off scores to determine the benchmark associated with mastery versus non-mastery (cf. Hosp & Fuchs, 2005).

3. Results

We were interested in relationships between early literacy skills of at-risk students assessed with curriculum-based and norm-referenced measures across grades. We used descriptive and predictive statistics to address each of our research questions.

3.1 Descriptive Comparison of Curriculum-Based and Norm-Referenced Measures

Means and standard deviations for CBM and ORF are in Table 2. Average oral reading fluency scores for the participating students were at levels placing them at risk for academic failure (First Grade < 40, Second Grade < 90, Third Grade < 110) based on benchmarks recommended by Good and Kaminski (2003). They were also below levels for students needing intervention (i.e., 10 or more words below the 50th percentile) based on norms and recommendations published by Hasbrouck and Tindal (2006) in an article in *The Reading Teacher* and in a technical report available online. Means and standard deviations for norm-referenced performance on WRMT-R word reading, decoding, and comprehension subtests (see Table 2) reflected below average and average performance which was similar (e.g., 96-102, 101-103, 94-95, respectively) across the three grade levels.

3.2 Relationships between Curriculum-Based and Norm-Referenced Measures

The correlation coefficients between CBM and ORF, word reading, decoding, and comprehension for each grade are also in Table 2. Because the relationship between CBM and ORF was high across all three grade levels, we used CBM as the only measure to correlate with the three subtests of WRMT-R. Results indicated that CBM was strongly related to all three subtests of WRMT-R and across all three grade levels. Statistically significant ($p < .01$) differences were noted with z -tests for the relationships between CBM and word reading as well as the relationships between CBM and comprehension across grade levels, but the relationship between CBM and decoding was not statistically significantly different between first and second grade students ($z = 1.35, p > .01$). Except for the relationship between CBM and decoding for the first and second grade students, the relationship between CBM and all three subtests of WRMT-R was lower as students moved up in grade levels.

We also found that the relationship between CBM and word reading within each grade level was significantly different from the relationship between CBM and decoding. Similar results were evident for the relationship between CBM and decoding and the relationship between CBM and comprehension. The relationship between CBM and word reading, however, was not statistically significantly different from the relationship between CBM and comprehension within each grade level. The differences indicated that the relationship between CBM and word reading was stronger than that between CBM and decoding. The relationship between CBM and comprehension was also stronger than that between CBM and decoding. Word reading and comprehension were comparable with respect to their relationships to CBM and these relationships are both stronger than the relationship between CBM and decoding.

3.3 Predictive Analysis for Curriculum-Based and Norm-Referenced Measures

Hit rate, sensitivity, and specificity of the PDA were displayed in Table 3. Overall, CBM separated students who mastered the skills from their counterparts who did not master the skills measured by WRMT-R at each grade level (i.e., greater numbers of true negatives and true positives). The numbers for false negatives were small, indicating that CBM predicted students who did not master the skills measured by WRMT-R very well because very few of them were predicted as having mastered the skills. In contrast, the numbers in false positives were relatively large, suggesting that CBM predicted less accurately students who mastered the skills measured by WRMT-R because quite a few of these students were predicted as not having mastered the skills. This result that CBM predicted better for students who did not master skills measured by WRMT-R was also indicated by larger sensitivity indices in comparison with specificity indices. Although not identical, these scores were similar to those reported by Hosp and Fuchs (2005) for each grade and each reading skill. Since diagnostic accuracy statistics are always influenced by the cut-scores used to define levels of risk, continued study will help to clarify

the usefulness of CBM in predicting norm-referenced test performance.

4. Discussion

We followed similar procedures to those used in previous research and we obtained similar results although the populations represented by the samples were different. For example, Silbergliitt et al. (2006) included children in upper grades and the Hosp and Fuchs' (2005) distributions of gender and ethnicity (Caucasian and African American) were somewhat balanced. The majority of our lower grades sample was males by gender and African American by ethnicity with Hispanic students representing the second largest group and Caucasian students as the third largest group in schools enrolling large numbers of students at risk for school failure. Moreover, our sample was predominantly children eligible for free or reduced price lunch program. While Hosp and Fuchs (2005) did not report the socioeconomic status of children participating in their study across grade levels, the percentage of students receiving free or reduced lunch in all but one of their schools (82%, 42%, 41%, and 34%) was lower than in our sample (79-86%). Students in Silbergliitt et al. (2006) were predominantly "94.3% White, not of Hispanic Origin" and the "percentage of students meeting the federal definition of poverty...districts varied from 5.07 to 18.63" (p. 529). These demographic differences should be considered when interpreting relationships of CBM and other measures of achievement.

Although our samples differed, we obtained similar results to Hosp and Fuchs (2005) with regard to statistically significant relationships between CBM and (a) word reading, (b) decoding, and (c) comprehension at each grade level. These outcomes suggest that CBM is a good indicator of students' performance on standardized tests in reading. This conclusion is supported by the results from PDA. Both studies reached high hit rate, sensitivity, and specificity in using CBM to predict students who mastered or did not master the reading skills measured by WRMT-R. Similar to Hosp and Fuchs, we found statistically significant differences between the relationship between CBM and word reading and the relationship between CBM and decoding for the first and third grade students. The finding that the relationship between CBM and word reading and the relationship between CBM and comprehension were not statistically significantly different in Grades 2 and 3 (cf. Hosp & Fuchs) was also supported in our study.

We also obtained some different results. For example, in Hosp and Fuchs' (2005) study, the relationship between CBM and decoding was higher in Grades 2 and 3 than in Grade 1, but we found it to be higher in Grades 1 and 2 than in Grade 3. The relationship between CBM and word reading was not significantly different from each other in Grades 1-3 in Hosp and Fuchs' study, but in our study it was highest in Grade 1, the second highest in Grade 2, and the lowest in Grade 3. A similarly different outcome was also noted for the relationship between CBM and comprehension. For each grade level, the relationship between CBM and decoding and the relationship between CBM and comprehension was not statistically significantly different in Hosp and Fuchs' study but was statistically significantly different in our study.

4.1 Implications for Future Research and Practice

Data are keys to successful efforts to use response to intervention (RTI) to improve the lives of children. One of the critical elements of RTI is the use of a technically adequate system of progress monitoring to inform decision-making (Deno et al., 2009; Fuchs, Mock, Morgan, & Young, 2003). More importantly, "[a] distinct feature of progress-monitoring methods is that educators evaluate student performance on material that represents or is closely associated with either the skills or general outcome that students should achieve by the end of the year" (Stecker, Lembke, & Foegen, 2008, p. 51). In our study, we obtained similar results to those reported in earlier research and extended them to students not previously studied. Clearly, continued research is warranted to support efforts to improve early literacy outcomes especially for students placed at risk by diverse ethnic, cultural, and language backgrounds.

In this context, then, learning more about how well widely-used assessments, such as DIBELS ORF and other CBM, predict performance on other indicators of achievement, included norm-referenced measures often used to identify students in need of special education and end-of-grade tests used for high stakes decision making, is clearly important. Additionally, because many schools using curriculum-based progress-monitoring assessments serve a large number of students who are traditionally considered at risk for continued school failure, it is essential to study their predictive usefulness with different student groups. While there is evidence that oral reading fluency measures are unbiased predictors of reading comprehension for African American and White students (Hintze, Callahan, Matthews, Williams, & Tobin, 2002; Hixson & McGlinchey, 2004; Wanzek, et al., 2010), continued study of the usefulness of these measures with students with diverse ethnic, cultural, and language backgrounds is clearly warranted (Klein & Jimerson, 2005; Roehrig et al., 2010).

Tiered continuous-improvement frameworks, such as RTI, direct schools and teachers to draw strong

connections between student performance data and instructional decision making (Cummings, Atkins, Allison, & Cole, 2009; Mercier Smith, Fien, Basaraba, & Travers, 2009; Stuart & Rinaldi, 2009). Typically, in reading this means that all students participate in a comprehensive core program (Tier I) and receive strategically-focused (Tier II) or intensively-individualized (Tier III) instruction when evidence suggests inadequate progress is being made. The take-away message for schools and teachers from this study is that simple progress monitoring measures (e.g., one-minute oral reading fluency assessment) are excellent predictors of performance on individually- and group-administered standardized tests that provide a strong foundation for collaborative planning and instruction frameworks that guide RTI and other models focused on improving education. Working together in teams, teachers and other professionals should use progress monitoring measures to decide if individuals or groups of students are performing at levels likely to prevent progress. Then they should document the specific nature and level of the problem and the intervention to be directed to improving it, provide an intervention to correct the problem, peruse (i.e., examine or consider with focused attention and in detail) data to determine if the problem has improved, stayed the same, or worsened. The “decide-document-provide-peruse” decision making should occur regularly to support and ensure success for all students.

5. Conclusion

Reading is a critical tool skill for children to learn early in school. Reading has received and continues to receive attention at the national level as the focus of major initiatives designed to correct extant problems and ensure renewed progress. Efforts to enable school psychologists and other professionals to efficiently and effectively monitor children’s early reading progress are essential to the success of practice-altering promises such as RTI. Our data support the validity of oral reading fluency measures as indicators of the progress and performance of elementary school students in first through third grade. Our outcomes are, in general, consistent with those reported by Hosp and Fuchs (2005) and also documented and discussed by others using different outcome measures (see Roehrig et al., 2007; Schilling et al., 2007; Wayman et al., 2007; Wanzek et al., 2010; Wiley & Deno, 2005). Different from Hosp and Fuchs’s (2005) findings, our study suggested that the relationship between CBM and norm-references tests was lower as students moved up in grade levels. This difference warrants further investigation and future studies with students of higher grade levels. Teachers and school administrators should be cautious when applying this result, especially with older students. Moreover, we have added to the extant knowledge base by including younger students and those from diverse cultural and ethnic backgrounds that place them at risk for early and continued failure in school. In the end, our results reiterate that CBM is appropriate for formative and summative monitoring of early reading skill progress and outcome. Additionally, our research further illustrates that cut scores are useful in identifying students who require additional instruction in reading—a finding with broad value in efforts to improve educational outcomes for all with enhanced instruction grounded in improved progress monitoring and RTI.

References

- Blalock, H. M., Jr. (1960). *Social statistics* (2nd ed.). New York: McGraw-Hill.
- Case, L. P., Speece, D. L., & Mollow, D. E. (2003). The validity of a response-to-intervention paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. *School Psychology Review*, 32, 557-582.
- Catts, H.W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, 42, 163-176. <http://dx.doi.org/10.1177/0022219408326219>
- Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review*, 35, 128-133.
- Christ, T. J., & Silberglitt, B. (2006). Estimate of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, 36, 130-146.
- Cummings, K., Atkins, T., Allison, R., & Cole, C. (2008). Response to Intervention: Investigating the new role of special educators. *TEACHING Exceptional Children*, 40(4), 24-31.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., & Windram, H., et al. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools*, 46, 44-55. <http://dx.doi.org/10.1002/pits.20353>

- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.
- 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, 33-49.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41, 93-99. <http://dx.doi.org/10.1598/RRQ.41.1.4>
- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, 18, 157-171. <http://dx.doi.org/10.1111/1540-5826.00072>
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification versus nonsense word fluency. *Exceptional Children*, 71, 7-21.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 20-28. <http://dx.doi.org/10.1177/074193258800900206>
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21, 45-58.
- Good, R. H., & Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency Passages for First through Third Grades* (Technical Report No. 10). Eugene, OR: University of Oregon.
- Good, R. H., & Kaminski, R. A. (2003). *Administration and scoring guide: Dynamic Indicators of Basic Early Literacy Skills* (DIBELS). Longmont, CO: Sopris West.
- Good, R. H., III, 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*. Unpublished manuscript, University of Oregon at Eugene.
- Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59, 636-644. <http://dx.doi.org/10.1598/RT.59.7.3>
- Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.
- Hintze, J. M., Callahan, J. E., III, Matthews, W. J., Williams, S. A. S., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, 31, 540-553.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32, 541-556.
- Hintze, J.M., & Silbergliitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, 34, 372-386.
- Hintze, J.M., Owen, S.V., Shapiro, E.S., & Daly, E.J., III. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly*, 15, 52-68. <http://dx.doi.org/10.1037/h0088778>
- Hixson, M. D., & McGlinchey, M. T. (2004). The relationship between race, income, and oral reading fluency and performance on two reading comprehension measures. *Journal of Psychoeducational Assessment*, 22, 351-364. <http://dx.doi.org/10.1177/073428290402200405>
- Hosp, M., K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, 34, 9-26.
- 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.
- Johnson, E.S., Jenkins, J.R., & Petscher, Y. (2010). Improving the accuracy of a direct route screening process. *Assessment for Effective Intervention*, 35, 131-140. <http://dx.doi.org/10.1177/1534508409348375>
- Johnson, E.S., Jenkins, J.R., Petscher, Y., & Catts, H.W. (2009). Can we improve the accuracy of screening instruments? *Learning Disabilities Research and Practice*, 24, 174-185. <http://dx.doi.org/10.1111/j.1540-5826.2009.00291.x>

- Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly*, 20, 23-50. <http://dx.doi.org/10.1521/scpq.20.1.23.64196>
- Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp.18-78). New York: Guilford Press.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33, 193-203.
- Mercier Smith, J., Fien, H., Basaraba, D., & Travers, P. (2009). Planning, evaluating and improving tiers of support in beginning reading. *Teaching Exceptional Children*, 41(5), 16-22.
- 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. <http://dx.doi.org/10.1016/j.jsp.2009.07.001>
- Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS Oral Reading Fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology*, 46, 343-366. <http://dx.doi.org/10.1016/j.jsp.2007.06.006>
- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal*, 107, 429-448. <http://dx.doi.org/10.1086/518622>
- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: The Guilford Press.
- Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford Press.
- Silbergliitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-ZCBM growth rates by level of performance. *Exceptional Children*, 74, 71-84.
- Silbergliitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*, 43, 527-535. <http://dx.doi.org/10.1002/pits.20175>
- Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2008). *Digest of Education Statistics 2007* (NCES 2008-022). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education*, 36, 223-233. <http://dx.doi.org/10.1177/002246690303600403>
- Speece, D.L., Ritchey, K.D., Silverman, R., Schatschneider, C., Walker, C.Y., & Andrusik, K.N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*, 39, 258-276.
- Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure*, 52, 48-58. <http://dx.doi.org/10.3200/PSFL.52.2.48-58>
- Stuart, S., & Rinaldi, C. (2009). A collaborative planning framework for teachers implementing tiered instruction. *Teaching Exceptional Children*, 42(2), 52-57.
- Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities*, 38, 364-380. <http://dx.doi.org/10.1177/00222194050380041401>
- Vaughn, S., & Fuchs, L. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, 18, 137-146. <http://dx.doi.org/10.1111/1540-5826.00070>
- Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruff, A. L., & Murray, C. S. (2010). Differences in the relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessment for Effective Intervention*, 35, 67-77. <http://dx.doi.org/10.1177/1534508409339917>

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, 41, 85-120. <http://dx.doi.org/10.1177/00224669070410020401>

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*, 26, 207-214. <http://dx.doi.org/10.1177/07419325050260040301>

Woodcock, R. M. (1998). *Woodcock Reading Mastery Test – Revised*. Circle Pines, MN: American Guidance Service.

Woodward, J., & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in middle grade mathematics. *The Journal of Special Education*, 40, 151-159. <http://dx.doi.org/10.1177/00224669060400030301>

Ysseldyke, J. (2005). Assessment and decision making for students with learning disabilities: What if this is as good as it gets? *Learning Disability Quarterly*, 28, 125-128. <http://dx.doi.org/10.2307/1593610>

Ysseldyke, J. E., & Algozzine, B. (1979). Perspectives on assessment of learning disabled students. *Learning Disability Quarterly*, 2(4), 3-13. <http://dx.doi.org/10.2307/1510820>

Ysseldyke, J., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalizations from five years of research on assessment and decision making: The University of Minnesota Institute. *Exceptional Education Quarterly*, 4(1), 75-93.

Table 1. Demographic information for students by grade level

	Grade			df	χ^2
	1	2	3		
	(n = 475) [74]	(n = 360) [81]	(n = 181) [76]		
Gender				2	2.45
Male	266 (56%) [57%]	220 (61%) [56%]	109 (60%) [53%]		
Female	209 (44%) [43%]	140 (39%) [44%]	72 (40%) [47%]		
Ethnicity				6	9.36
African American	252 (53%) [40%]	191 (53%) [37%]	92 (51%) [56%]		
Caucasian	48 (10%) [49%]	54 (14%) [56%]	22 (12%) [39%]		
Hispanic	152 (32%) [n/a]	90 (25%) [n/a]	56 (31%) [n/a]		
Other	23 (5%) [11%]	25 (7%) [7%]	11 (6%) [5%]		
Lunch Status				4	7.58
Free	352 (74%) [n/a]	248 (69%) [n/a]	144 (79%) [n/a]		
Reduced	43 (9%) [n/a]	36 (10%) [n/a]	12 (7%) [n/a]		
Regular	80 (17%) [n/a]	76 (21%) [n/a]	25 (14%) [n/a]		
Special Education Status				2	0.25
Yes	42 (9%) [5%]	30 (8%) [10%]	12 (7%) [11%]		

Note. Numbers in parentheses are percentages of each group of students for each grade level; numbers in brackets are those reported by Hosp and Fuchs (2005), Table 1, p. 13. None of the reported χ^2 values reported above were statistically significant.

Table 2. Means, standard deviations, and correlations for CBM and other measures

	Grade								
	1			2			3		
	M	SD	r	M	SD	r	M	SD	r
CBM	36.71	25.25	--	72.04	30.59	--	108.55	33.56	--
ORF	37.06	25.74	.96	66.78	32.41	.92	81.82	27.04	.88
Word Reading	101.66	14.74	.82	96.31	10.28	.73	95.72	9.86	.60
Decoding	103.16	15.54	.68	101.49	14.45	.63	100.94	12.30	.46
Comprehension	94.46	13.15	.83	95.36	11.07	.72	95.04	7.90	.59

Note. All correlations significant at $p < .01$.

Table 3. Hit rates, sensitivity, and specificity indices for CBM predicting mastery/non-mastery of word reading, decoding, and comprehension on the WRMT-R

Reading Skill	TP	FN	TN	FP	Hit Rate (%)	Sensitivity (%)	Specificity (%)
Grade 1 (Based on 85 Standard Score)							
Word Reading	46	1	324	104	78	98	76
Decoding	50	5	293	127	72	91	70
Comprehension	90	5	285	94	79	95	75
Grade 2 (Based on 85 Standard Score)							
Word Reading	31	0	245	84	77	100	74
Decoding	27	4	216	113	68	87	66
Comprehension	52	5	215	88	74	91	71
Grade 3 (Based on 85 Standard Score)							
Word Reading	9	1	123	48	73	90	72
Decoding	11	2	110	58	67	85	65
Comprehension	14	2	120	45	74	88	73

Note. TP = true positive; FN= false negative; TN = true negative; FP = false positive.

Hit rate = $(TP + TN)/N$; sensitivity = $TP/(TP + FN)$; specificity = $TN/(TN + FP)$