5-2016

The Relationship of Curriculum-Based Measurement Oral Reading as an Indicator of Reading Achievement

Hyunchang Moon
Saint Cloud State University, mohy1401@stcloudstate.edu

The Relationship of Curriculum-Based Measurement Oral Reading as an Indicator of Reading Achievement

by

Hyunchang Moon

A Starred Paper

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

For the Degree

Master of Science in

St. Cloud, Minnesota

May, 2016

Starred Paper Committee:
Janet Salk, Chairperson
Mary Beth Noll
Yun Claire Park

Table of Contents

Chapter Page

List of Tables

Chapter I: Introduction

Assessment is the process of gathering and discussing information from multiple and diverse sources to identify student's learning needs and plan instruction (Huba & Freed, 2000). The reauthorized Individuals with Disabilities Education Improvement Act of 2004 (IDEA) emphasized the important role of assessment by requiring all public schools to provide screening and benchmark testing of students at least three times per year to determine instructional needs (Yell, Shriner, & Katsiyannis 2006). Both IDEA and the No Child Left Behind Act of 2001 (NCLB) contain regulations that require educators to use scientifically based assessment practices to assess student performance (Christine, Kristen, Susan, & Miya Miura, 2012; Shinn, 2007; Tindal, 2013).

Progress monitoring is a type of assessment that is considered to be a scientifically based practice (Deno, 2003; Fuchs & Fuchs, 2006). It enables teachers to determine what their students learned and what still needs to be taught. In contrast to standardized tests that simply compare a student's performance with other children or state standards, progress monitoring provides meaningful data that enables the teacher to assess each student's academic performance and evaluate instructional effectiveness (McMaster & Wagner, 2007; Stecker & Fuchs, 2000). Traditionally, educational assessment has been used for identifying students with special needs through the data of standardized measures (Deno, 1997). However, the use of norm-referenced tests has been revealed to be unreliable for tools in monitoring students' progress (Deno, 1992; Fuchs, 2004). Assessments focused on progress monitoring need to be reliable and valid so that they can be used to sensitively monitor student progress as an index performance and change to more effective interventions over time (Deno, 1997; Stecker & Fuchs, 2000; Vaughn & Fuchs, 2003).

The most common progress monitoring approach is curriculum-based measurement (CBM), which is used to monitor individual or class performance on a regular basis—weekly, biweekly, or at least once monthly (Stecker & Fuchs, 2000). Progress monitoring is used most frequently in the area of reading because reading is fundamental to learning in other subject areas. Learners who do not perform at grade level in reading are far more likely to experience academic failure (Jenkins, Hudson, & Lee, 2007).

The purpose of this paper is to examine how curriculum-based measurement of oral reading correlates with other standardized measures of reading achievement for students in grades K-6. Chapter I provides background information on data-based decision making and assessment, progress monitoring and response to intervention, and problem solving shift and CBM. It also presents the research question, focus of the review, importance of the study, and definitions of terms.

Data-Based Decision Making and Assessment

Data-based decision making can create a significant difference for both students and teachers. (Lai & Hsiao, 2014; Mandinach, 2012). There are two ways to collect decision-making data: formal and informal assessment. Formal assessments are standardized measures that compare a student's performance with that of peers who are similar in age or grade level. Formal assessments produce data that are mathematically computed and reported in percentiles, stanines, or other standard scores. Informal assessments are considered to be more authentic and are typically curriculum- or performance-based measures used to describe student functioning and inform instruction effectiveness. Informal assessments provide educators with more information about content and performance (Marchand & Furrer 2014; Spinelli, 2011).

Whether formal or informal, assessments should (a) reflect the subject content that is most essential for students to learn, (b) improve learning through a link with instruction, (c) provide consistently reliable data of student performance, and (d) produce valid inferences about student learning (Spinelli, 2011). Progress monitoring systems meet these four criteria, quantify a student rate of responsiveness to instruction, and can be implemented with individual students or an entire class (Stecker, Fuchs, & Fuchs; 2005). Research from the past 2 decades indicates when teachers use progress monitoring for instructional decision making, the benefits include increased student achievement, improved teacher decision making, and enhanced student awareness of their school performance (Batsche et al., 2005; Fuchs & Fuchs, 2011; Lai & Hsiao, 2014).

Progress Monitoring and Response to Intervention

Progress monitoring refers to individualized decision making about academic skill development (Fuchs & Fuchs, 1992). Typically, it is conducted frequently to estimate rates of improvement and to identify students who are not demonstrating sufficient progress (Fuchs & Fuchs, 2011). Students can also track their own progress, which involves them more directly in meeting their educational goals. A wide range of progress monitoring is used in classrooms to monitor student performance across core subject areas: reading, mathematics, writing, and spelling (Stecker, Lembke, & Foegen, 2008).

At one time, schools were required to use the IQ-achievement discrepancy model for identifying students with a learning disability (LD) for special education services (Fuchs & Fuchs, 2006). After the implementation of IDEA 2004, schools can use response to intervention (RTI) as an alternative model for identifying students for LD. By using the RTI model, students who are falling behind academically but do not qualify for special education services are able to obtain and access earlier intervention (Fuchs & Fuchs, 2006; Justice, 2006; Reschly & Bergstrom, 2009). RTI uses a three-tiered framework to provide preventive services prior to the onset of serious deficits and before a student fails (Fuchs, Fuchs, & Hollenbeck, 2007). Different instructional interventions are used at each tier, based upon each student's demonstrated need. In the first tier, universal screening measures are used to identify students who are at risk for academic failure. When students do not respond to the instructional program and do not make adequate yearly progress, Tier 2 individual or small-group instructional modifications are implemented. Tier 3 interventions are implemented for the 1-5% students who are at the highest risk and require intensive, individualized interventions such as special education services (Burns & Gibbons, 2008; Greenwood, Carta, Baggett, Buzhardt, Walker, & Terry, 2008; Justice, 2006; Prasse, Breunlin, Giroux, Hunt, Morrison, & Thier, 2012; Reschly & Bergstrom, 2009).

Progress monitoring is now an essential Tier 2 and 3 component in RTI models that have been adopted in schools throughout the nation. Progress monitoring in RTI is an effective and resourceful tool for identifying children who are not making expected rates of short-term progress and who may need more targeted interventions (Fuchs & Fuchs, 2007). Students' growth during Tier 2 and Tier 3 interventions are monitored to determine students' response to interventions. Progress is monitored frequently, and interventions of increasing intensity designed to match a learner's demonstrated response to intervention (Greenwood et al., 2008; Greenwood, Kratchowill, & Clements, 2008).

Problem Solving Shift and Curriculum-Based Measurement

The origins of CBM are in mastery measurement. Mastery measurement is the traditional progress monitoring approach in which teachers provide instruction based on each unit and objective sequence and then tests students' criterion-referenced mastery (Tindal, 2013). When the students reach criterion, they move to the next instructional phase in the hierarchy (Fuchs, 2004). However, mastery measurement does not typically track progress across the academic year. Norm-referenced tests were unreliable for tools in monitoring students' progress (Deno, 1985, 1992). To address this limitation, Deno (1978) developed a different approach to progress monitoring named curriculum-based measurement (CBM), which is also referred to as general outcome measurement (GOM). Minnesota Institute for Research on Learning Disabilities (IRLD) set some criteria to create a scientifically based progress monitoring tool which should be research-based, curriculum-based, time-efficient, multiple-formed, low-cost, and easy-to-apply (Jenkins, Deno, & Mirkin, 1979). Over the past 30 years, studies have investigated CBM in both general and special education settings and have found it to be a reliable and valid instrument (Christine, Kristen, Susan, & Miya Miura, 2012; Hosp, Hosp, & Howell, 2012; Tindal, 2013).

Curriculum-based measurement is a standardized process used to assess students' academic progress on a regular and frequent basis which guides in making decisions about students (Deno, 2003). CBM is a broad type of curriculum-based assessment (CBA) and has three major criteria: (a) materials are associated with the school's curriculum, (b) measurement regularly occurs, and (c) assessment data is used for instructional decision making (Tucker, 1987). Curriculum-based measurement consists of six steps: (a) select appropriate test probes, (b) administer and score the probes, (c) graph the scores, (d) set goals, and (e) communicate progress (Hosp, Hosp, & Howell, 2012).

In addition to frequent monitoring of learner progress and planning to make instructional decisions, CBM is easy to use and is sensitive to student progress and growth over short instructional periods (Burns & Gibbons, 2008). Standardized probes at regular intervals provide progress data that teachers can use to gauge students' growth and establish long-term goals that will lead to proficiency (Fuchs & Fuchs, 2011). Although probes may differ in that they are based on the curriculum, they measure reading fluency consistently with regard to accuracy and automaticity (Roehrig, Nettles, S. M., Hudson, & Torgesen, 2008). The results are graphed so that teachers can easily determine if students meet their reading goals or their educational programs are effective (Fuchs, Fuchs, & Compton, 2004).

Curriculum-based measurement allows comparison of an individual's performance on other similar tasks and to classroom or grade-level peers (Deno, 1985). It can be used as a more individualized approach for making decisions regarding special education eligibility, placement, instruction, and accountability (Mercer, Mercer, & Pullen, 2009). In addition, students are more aware that they are progressing toward a long-term goal, which helps them pay attention to individual learning (McGlinchey & Hixson, 2004).

Research Question

One research question guides this comprehensive literature review: What is the relationship between curriculum-based measurement oral reading measures and other standardized measures of reading achievement?

Focus of the Review

In this literature review, I reviewed empirical studies that provided correlational evidences of the relationship between curriculum-based measurement of oral reading (CBM-R) and other standardized measures of reading achievement. To find literature relevant to my topic, research participants in Chapter II studies must include general education or special education students in grades K-6. The Chapter II literature review was limited to studies published in the United States from 2005 to 2015.

I reviewed journal articles pertaining to curriculum-based measurement of oral reading and other standardized tests of reading achievement by using Academic Search Premier, EBSCOhost, ERIC, and PsycINFO search engines and searching *Journal of Psychoeducational Assessment, Journal of Exceptional Children, Journal of School Psychology, and Journal of Psychology in the Schools*. I used a variety of different keywords and combinations to locate relevant information: *assessment, progress monitoring, Individuals with Disabilities Education Act, data-based making decision, reading measure, reading achievement, general outcome measure, oral reading fluency, curriculum-based measurement, CBM-R, standardized reading test, response to intervention, correlation,* and *special education*.

Importance of the Topic

The five essential components of elementary reading instruction should target phonics, phonemic awareness, fluency, comprehension, and vocabulary (National Institute of Child Health and Human Development, 2000). At this level, reading fluency is necessary in improving reading comprehension because reading fluency development is critical for early reading success (Armbruster, Lehr & Osborn, 2001). In addition to the importance of reading fluency for reading comprehension, developing reading fluency is also important for educators and parents to improve special education accountability and effectiveness (Silberglitt, Burns, Madyun, & Lail, 2006). Public law mandates that progress toward IEP goals and objectives must be reported at the same frequency as progress is reported to parents of student without disabilities (IDEA, 2004). The impact of individual progress monitoring in reading fluency should be investigated and shared with the individual education plans (IEP) team so that they can implement scientific- and evidence-based instructional practices (Stecker, Lembke, & Foegen, 2008). The use of student achievement data is critical to successful intervention and data-based instructional decision making (Lai & Hsiao, 2014; Reschly, Busch, Betts, Deno, & Long, 2009). How do teachers know whether their students are improving satisfactorily in reading achievement? The most common means of monitoring progress is to carefully observe students' performance during reading instruction. However, it is more informative to actually measure reading performance. The point is finding a suitable reading achievement tool that can be given repeatedly to assess student progress.

To monitor students' reading progress, CBM-R is used to gather and chart data and then evaluate it to make decisions regarding students' instructional needs. It can involve the students in monitoring progress of their academic goals and also is currently used as an assessment tool by special educators. The CBM-R tool can give information that can help students track their own progress, help teachers to design more effective plans, and make better decisions about the type of instruction that will work best with their students. Student progress monitoring also helps teachers evaluate how effective their instruction is for each student who receives special education services or for the entire class and improve education accountability and effectiveness.

Definitions of Terms

Academic Improvement Measurement System based on the web (AIMSweb) is built on general outcome measurement, a form of CBM, and the K-12 assessment system that provides academic assessments in reading, math, and language arts for universal screening and progress monitoring (Shinn, 2012). It measures 1-minute standardized reading probes to assess reading fluency.

Assessment refers to the process of collecting data related to a goal and objectives to identify students in need of additional education or monitor their progress, such as tests, observations, or interviews (Overton, 2006).

Correlation coefficient is a quantity that measures the direction and the degree to which two variable's movements are associated, meaning statistical relationships between two or more variables or observed data values (Gay, Mills, & Airasian, 2006).

Curriculum-based measurement (CBM) is the representative example of GOM and is a standardized assessment approach for gathering student performance data across core subject areas: reading, mathematics, writing, and spelling that empirically supports efficient educational decision making such as screening, progress monitoring, and instructional diagnosis. (Deno, 1985; Fuchs, 2004)

Curriculum-based measurement of oral reading (CBM-R) typically consists of three standardized reading passages at grade-appropriate level. The student reads each passage aloud for 1 min, and the examiner calculates the median number of words read correctly across the three passages (Silberglitt, Burns, Madyun, & Lail, 2006).

Data-based decision making is the ongoing process of analyzing and evaluating student data to determine the efficacy of instruction and intervention (Mandinach, 2012).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of standardized measures to assess the acquisition of students' early literacy skills (Good & Kaminski, 2002). It has been developed based on measurement procedures for CBM and is designed for seven measures to monitor the development of reading skills aligned with the essential literacy domains of National Reading Panel and National Research Council (Hintz, Ryan, & Stoner, 2003). The AIMSweb and DIBELS systems are commonly used to assess oral reading fluency and other literacy skills.

Evaluation is procedures to determine whether a student meets a certain criteria, such as qualifying for special education services. It uses assessment to make a determination of qualification (Overton, 2006).

General outcome measure (GOM) is designed to individually approach continual measurement of a student's progress toward long-term goals instead of assessment of component target skills. It is the critical indicator of growth and comparison between the achievements of a student and a group (Overton, 2006).

Maze CBM consists of one standardized grade-level reading passage. The first sentence of each passage is left intact, and then every subsequent seventh word is replaced with three words provided inside parentheses, one of which was from the original passage and the other two being near and far distractors that are words with the incorrect meaning. (Silberglitt et al., 2006).

Measurement refers to the set of procedures and the principles for how to use the procedures in educational tests and assessments (e.g., raw scores, percentile ranks, standard scores) (Overton, 2006).

Multi-tiered system of supports (MTSS) is a multi-tiered interventions delivery system with the intensity of intervention increasingly based on frequent student performance monitoring (Batsch et al., 2005). Most multi-tiered systems use three tiers: (a) Tier 1 is a universal instruction for all students, (b) Tier 2 is targeted interventions for some students, and (c) Tier 3 is intensive intervention for few students. MTSS includes systems such as RTI and positive behavioral intervention and supports (PBIS) (Sugai & Horner, 2009; Sanetti & Collier-Meek, 2015).

Progress monitoring is defined as a scientifically based practice that is used to assess students' academic performance and evaluate the effectiveness of instruction and is based on principles of simple repeated measurement of student performance toward a long-range instructional goal (Deno, 1992, 2003).

Oral Reading Fluency (ORF) is the ability to read a text with speed, accuracy, and prosody (Hudson, Pullen, Lane, & Torgesen, 2009). The National Reading Panel (NRP) identified five skills that contribute to reading development: phonemic awareness, phonics, fluency, vocabulary, and comprehension. CBM-R is based on oral reading fluency because fluency provides a close link between fluency and reading comprehension (Armbruster, Lehr & Osborn, 2001; National Institute of Child Health and Human Development, 2000).

Chapter II: Review of the Literature

In this chapter, I review the literature that investigates the relationship between curriculum-based measurement of oral reading (e.g., AIMSweb and DIBELS) and other standardized measures of reading achievement (e.g., 6 nationwide- and 7 statewide-standardized tests) for K-6 students to evaluate if CBM-R is a reliable tool as an alternative indicator of students' overall reading achievement. In addition, I discuss recommendations and limitations of each study. Ten studies are organized in chronological order from oldest to most recent.

Colon and Kranzler (2006)

Colon and Kranzler (2006) investigated the effect of three different administrations of CBM-R (i.g., baseline, fast, and best condition) and the relationship between CBM-R and standardized reading tests. The study participants included 50 fifth graders in North Central Florida. The demographic distribution was 58% Caucasian, 22% African American, 6% Asian, 4% Hispanic, and 10% Other.

Participants were randomly assigned to 1 of 2 groups. Two baseline CBM-R probes were administered by having the participant read the passage aloud. During the counterbalance intervention phase, Group A first received instruction to read as "fast as they can" without making mistakes (fast condition) and then a different instruction to do their best reading (best condition). As the participants read aloud the reading probe in the 1-min period, the examiner counted the number of mispronunciations, omissions, and word reversals. Following the CBM-R probes, they were also administrated the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests of the *Woodcock-Johnson Tests of Achievement-Third Edition* (WJ-III; Woodcock, Mather, & McGrew, 2001). All testing was conducted during the fall of the academic year.

According to *t*-test results between Group A and B, no significant differences were found between groups for the mean number of words read correctly or the mean number of errors (all *p*s > .03). The data for Groups A and B depended on the different instruction. As the results of *t* tests to examine mean differences in the words read correctly across CBM-R condition, statistically significant differences were found between each of the conditions. The mean of the fast condition was significantly greater than the mean of both the baseline ($t = -5.65$, $p < .01$) and best conditions ($t = -5.65$, $p < .01$). The mean of the baseline condition was greater than the mean of the best condition ($t = 2.56$, $p < .01$). Correlations between the CBM-R and WJ-III reading scaled scores exceeded .80, which indicated that the relationship between CBM-R and reading achievement was not affected by the use of different directions. Differences between the correlations for each CBM-R condition and WJ-III subtests were not statistically significant (*p*s > .05).

This study indicated that instructions were related to CBM-R performance. When asked to read as fast as participants could, they read significantly more words correctly per minute. The effect size between the fast and best conditions of approximately one-half standard deviation is not insubstantial. Each of the reading instruction conditions correlated greater than .80 with WJ- III reading achievement, which indicated that the relationship between CBM-R and reading achievement was not affected by the use of different directions. Nonetheless, the importance of using standardized instructions on CBM-R results both within and across settings were underscored.

Shapiro, Keller, Lutz, Santoro, and Hintze (2006)

Shapiro, Keller, Lutz, Santoro, and Hintze (2006) examined the relationships between CBMs (i.e., reading and math) and standardized reading tests including both one statewide and two nationwide norm-referenced standardized achievement tests in two districts in eastern Pennsylvania. The 1,048 students who participated in this research were from six elementary schools in one urban and one suburban school district. The percentage of participants' overall free-and-reduced lunch level was 32%, and no students had Individualized Education Plans (IEPs). Two types of measures were used: (a) CBMs for reading, math computation, and math concepts/applications and (b) standardized assessments (*Pennsylvania System of School Assessment* (PSSA; Pennsylvania Department of Education, 2002), *Stanford Achievement Test-Ninth Edition* (SAT-9; Harcourt Brace Educational Measurement, 1996), *Metropolitan Achievement Test-Eighth Edition* (MAT-8; Harcourt Brace Educational Measurement, 2002), and *Stanford Diagnostic Reading Test* (SDRT; Karlsen & Gardner, 1995).

AIMSweb CBM data were collected from the two local norming projects for reading and math computation for Grades 1 through 5 and for math concepts/applications only in District 1 at fall, winter, and spring in the school year 2002-2003. The PSSA was administered for third and fifth grades for both districts at fall, winter, and spring assessments. In District 1, MAT-8 data in reading and math were collected in Grade 4, and SDRT reading data were collected in Grade 5. In District 2, SAT-9 data were collected in reading and math in Grades 2 and 4. Data analyses are presented for reading scores only in this chapter.

All Pearson correlations between the CBM-R scores and the PSSA scores obtained from subtests of each of standardized achievement test at fall, winter, and spring across districts were statistically significant at the $p < .001$ level, and correlation coefficients ranged between .62 and .69, except for the fall assessment for District 2. The MAT-8 showed moderate to strong correlations with CBM-R scores across all subtests for fourth graders in District 1. On the MAT-8 subtests of Total Reading, Sounds and Print, Vocabulary, and Comprehension, correlation coefficients ranged from .519 to .724, with most correlations .633 or higher. The fall, winter, and spring SDRT and CBM correlations were .524, .518, and .551, respectively. Moderate to strong relationships were reported for the correlation between the SAT-9 and CBM-R for second and fourth graders in District 2, ranged from .438 to .744. In addition, the correlations between CBM-Math outcomes and the SAT-9 math subtests were significant and ranged from .45 to .72.

Results showed that CBMs had moderate to strong correlations with the high stakes reading and mathematics assessments across two school districts. With these findings, the authors suggested that CBM-R have potential for identifying students who are likely to pass or fail the statewide assessments. Although not the focus of this study, math CBM and state assessment correlations were not as strong as those found in reading.

Silberglitt, Burns, Madyun, and Lail (2006)

Silberglitt, Burns, Madyun, and Lail (2006) examined the relationship between CBM-R and state-standardized test scores, potential grade differences in relationship magnitude, and differences in relationship magnitude between two CBM reading assessments (e.g., CBM-R and Maze-CBM) by comparing two state reading tests. The study participants included 5,472 students in grades 3, 5, 7, and 8 from five rural and suburban districts in Minnesota, equally representing both female and male. Over 94% of the students were White, and the percentage of students in poverty ranged from 5.07% to 18.63%.

All students in Minnesota in grades 3, 5, 7, and 8 take one or more of the state accountability tests: the *Minnesota Comprehensive Assessments-Reading* (MCA-R) in Grades 3, 5, and 7 and *Basic Standards Test-Reading* (BST-R) in Grade 8. These are criterion-referenced, standardized achievement tests. In this study, trained school personnel also administered CBM-Rs and Maze-CBMs and collected CBM data in the same grades. The means and standard deviations for each cohort across seven years were used to convert raw scores to *z* standard scores.

All coefficients met or exceeded .50 and were significant ($p < .001$). Correlation coefficients ranged from .51 (eighth graders) to .71 (third graders) for CBM-R and .49 (eighth graders) to .54 (seventh graders) for Maze-CBM. The magnitude of the coefficients between CBM-R and state test scores for third and fifth graders were significantly larger than those for eighth graders. No significant differences in correlation coefficient's magnitude were reported between the CBM-R and Maze-CBM when compared to MCA-R state test scores. Coefficients for seventh and eighth graders were .54 and .48, respectively.

Silberglitt et al. (2006) noted that although CBM-R continued to account for a substantial amount of the variance in student performance in the later grades, the overall value of this as a predictor diminished in later grades. Thus, this decreased relationship should be investigated further. Due to diminished sample size due to only 1 year of state test data, conclusions from data for seventh and eighth graders could not be evaluated as confidently as conclusions from earlier grades. Future research could be reflected on cross-grade analyses with larger sample size.

Uribe-Zarain (2006)

Uribe-Zarain (2006) conducted a 2-year study to determine that the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measure of Oral Reading Fluency (ORF) is a reliable predictor of reading performance on the *Delaware Student Testing Program* (DSTP) in third grade. The participants included 652 third-grade students from nine schools participating in *Reading First* program in the state of Delaware during 2004 - 2005 school year. Fifty percent of the students were female, 15% were classified as special education students, less than 3% were considered limited English proficient, and 59% received free or reduced lunch. They were divided into six groups of equitable racial composition, socioeconomic status, and special education placement according to DIBELS-ORF outcome levels of *at risk, some risk,* and *low risk*.

The data considered in this study were obtained from both the DIBELS-ORF subtest and the reading portion of DSTP. Although the ORF was administered three times a year, the winter ORF score was selected because the reading DSTP was taken in March. The ORF outcomes were classified according to the three risk levels of *at risk, some risk*, and *low risk*; reading DSTP scale scores were classified into five performance levels: *well below* the standard, *below* the standard, *meeting* the standard, *above* the standard, and *distinguished*. A series of logistic regression analyses were conducted to test if the optimal risk-level cutoffs for ORF predicted reading performance on DSTP equally well for students with selected independent variables (e.g., free or reduced-priced lunch status, English language learner status, or race/ethnicity).

The results revealed a significant correlation between ORF scores and reading DSTP scores ($r = .61$, $p < .01$). Students in the *some risk* category who scored between 67 and 91 correct words per minute were more likely to meet the reading standard. Students classified as *low risk* with ORF score of 92 or more were very likely to meet the standards. For racial composition and the ORF classification, White students had higher concentration in the groups in which students met or exceeded the state reading standards. White students in the *low risk* group met the DSTP standard more than White students in the *at risk* group. Socioeconomic status was characterized by whether the students received free or reduced lunch. Likewise, low-income students had a higher concentration in the groups in which students were classified as *at risk* by ORF data. In special education sample, the highest concentration of students with special education needs was in the groups where ORF classified them as *at risk*, whereas students with no special education were classified as *low risk.*

Results indicated that although all the correlations among six groups were significant, the correlation coefficients varied widely. For example, the relationship between DIBELS-ORF scores and reading DSTP scores of the group classification was stronger (Hispanic $r = .723$, White $r = .624$, and African American $r = .536$, $p < 0.01$). In addition, the relation between these two scores was stronger at higher socioeconomic status levels ($r = .643$, $p <0 .01$) and the group with no special education placement ($r = .579$, $p < 0.01$).

Uribe-Zarain discussed that even though approximately 72% of the total sample of *Reading First* third graders during the 2004-2005 school year met or exceeded the reading DSTP standard, the other 28% needed to be addressed to comply with district and state guidelines. The DIBELS-ORF screened 16% of the total sample as achievers below reading standard and 12% as achievers above reading standard. The study empasized that although they seemed to be a small number, educators needed to concentrate on why the DIBELS-ORF scores of certain student groups more correlated than others.

McIntosh, Graves, and Gersten (2007)

For 2 consecutive years, McIntosh, Graves, and Gersten (2007) studied the effectiveness of response-to-intervention (RTI) in four first-grade classrooms of English learners (ELs) from 11 native languages in three schools in an urban school district in California. The study participants included 4 teachers and 111 ELs who came from homes where nine foreign languages were spoken. The Year 1 study included 51 ELs, and Year 2 included 60 ELs. All of them received free or reduced-cost lunch at schools.

The four different first-grade classrooms were observed during a 2.5 hour reading/ language arts between five and seven times from Year 1 to Year 2 using the *English Language Learners Classroom Observation Instrument* (ELCOI; Haager et al. 2003). The ELCOI uses a 4-point Likert scale to assess 30 literacy practices, including the degree to which Tier 1 alone or Tier 1 plus Tier 2-type instruction was implemented. After observations, each teacher was interviewed for 30 mins and oral reading assessments were conducted. At the end of third grade, the DIBELS was administered to measure ORF, and the Passage Comprehension subtest of the *Woodcock Reading Mastery Test-Revised* (WRMT-R, Woodcock, 1987) was administered as a measure of reading comprehension. Participants read passages at the beginning and then again at the end of the year to determine the number of words read in 1 min. In Year 1, pretests were conducted in November and posttests in June. In Year 2, pretests were conducted in September and posttests in June. At the end of third grade, three 1-min timings were given, and words read correctly per minute were recorded. After the three timings, the WRMT-R Passage Comprehension subtest was administered to students.

The correlation between classroom ratings on the ELCOI and DIBELS-ORF gains from pre- to protest in first grade on ORF was moderately strong in both Year 1 (r = .61) and Year 2 (r = .57). The correlation between teacher ratings and ORF gains was strong in both Year 1 (r = .75) and Year 2 (r = .70). Results indicated a strong correlation (r = -.81) between the number of students below DIBELS benchmark thresholds at the end of first grade and the teacher rating data on the amount of instruction provided for low-performing participants. Follow-up data at the end of third grade in ORF and reading comprehension demonstrated moderate correlations to first-grade scores (r = .51 - .73).

The authors pinpointed that 8 of 9 students labeled with learning disabilities were reading below 20 words per minute (WPM) at the end of first grade, and 1 was reading 23 WPM. The results indicated a moderately strong relationship between teacher rating on the ELCOI measure and ORF outcomes at the end of first grade. ORF was strongly correlated with reading comprehension in the .6 to .7 range by the end of third grade. These data suggested that the ORF score of 23 or less was a serious cause for concern in English learners, which implied that those who were identified with learning disabilities required special education services that general education could not provide.

Riedel (2007)

Riedel (2007) investigated the relation between DIBELS and two standardized reading comprehension tests at the end of first grade and second grade in a sample of 1,518 first-grade students from a large urban school district in Memphis, Tennessee. Students in the sample completed the DIBELS tests at the beginning, middle, and end of their first-grade year and were assessed with a measure of reading comprehension at the end of first grade, *Group Reading Assessment and Diagnostic Evaluation* (GRA+DE; Williams, 2001) and second grade, *TerraNova Reading* (CTB/McGraw-Hill, 2003). The DIBELS and GRA+DE were administered during the 2003-2004 school year at 26 district schools, and TerraNova was administered during the 2004-2005 school year at the same schools to determine second-grade reading test outcomes.

Students were predominantly African American (*n* = 1,395, 92%), with a nearly equal representation of females (*n* = 760) and males (*n* = 758). The poverty rate in the sample was high, with 85% of the students qualifying for free or reduced-cost lunch. The population of ELs within the participating schools was small (*n* = 59), and their results were analyzed separately. Students receiving special education services were not included in the study

Receiver Operating Characteristic (ROC) analyses (Swets, Dawers, & Monahan, 2000) were used to examine the relation between DIBELS subtests and standardized reading comprehension test. Through ROC analyses, each DIBELS measure available at each time period (beginning, middle, and end of the year) was examined as a predictor of reading comprehension status at the end of first grade. Pearson correlations were calculated between DIBELS subtests and reading comprehension measures. Analysis of variance (ANOVA), chi-square, and logistic regression analyses were used to examine students for whom DIBELS was a poor predictor of reading comprehension.

Results from the ROC analyses showed that the ORF subtest of the DIBELS was a better predictor (r = .67) of comprehension than the remaining subtests (e.g. Letter Naming Fluency [LNF], Nonsense Word Fluency [NWF], phoneme segmentation fluency [PSF], and retell fluency [RF]). The use of other subtests in combination with ORF did not substantially improve predictive power beyond that provided by ORF alone. However, Vocabulary was a significant factor in the relation between ORF scores and reading comprehension. Participants with satisfactory ORF scores but poor comprehension (M = 34.2) had lower vocabulary scores than students with satisfactory ORF scores and satisfactory comprehension (M = 57.5). The ORF subtest was most strongly related to comprehension (r = .54 to .80), and PSF had a weak relation with comprehension (r = .14 to .23). The relationship between ORF and reading comprehension was stronger for ELL students (r = .72 to .80) than for non-ELL students (r = .49 to .67).

Riedel (2007) suggested that the relatively strong relation between DIBELS-ORF and reading comprehension support the use of DIBELS-ORF as a screening and outcome measure. However, the value of DIBELS-ORF as a diagnostic assessment was less clear. Although DIBELS ORF predicted current and future comprehension difficulties, Riedel contended that it may not provide any details regarding the student's reading difficulties or the interventions needed to remedy them. In addition, the current results did not support intervention instruction in phoneme segmentation or decoding for those who score poorly on the PSF or NWF.

Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, and Beck (2008)

Baker et al. (2008) investigated the relation between ORF and comprehensive reading tests in the context of the *Oregon Reading First* initiative conducted in low-performing and high-poverty schools. The participants included four cohorts of students in grades 1-3, with each cohort representing approximately 2,400 students. Students from 34 *Oregon Reading First* schools participated in this study representing 16 independent school districts that were located in large urban areas (17 schools), midsize cities (8 schools), and rural areas (9 schools) in Oregon. Approximately 69% of students across all *Reading First* schools qualified for free or reduced-cost lunch rates, and 27% of third graders did not pass minimum proficiency standards on the *Oregon Statewide Reading Assessment*. Ten percent of the students received special education services, and 32% of the participants were English learners.

Quantitative data were collected during the first 2 years of *Oregon Reading First* implementation. In 2003-2005, all students in kindergarten through third grade participated in four assessments per year. In the fall, winter, and spring, the DIBELS-ORF was administered as part of benchmark testing during the year; median scores was used as the representative performance scores. Two standardized tests were administered in the spring: *Stanford Achievement Test* (SAT-10, Madden, Gardner, Rudman, Karlsen, & Merwin, 1973) and *Oregon Statewide Reading Assessment* (OSRA). Growth curve analyses were used to test how well ORF predicted performance on SAT-10 or OSRA administered at the end of Year 2. The means and standard deviations for each cohort across 7 years were used to convert raw scores to *z* standard scores.

The correlations between DIBELS-ORF, SAT-10, and OSRA were consistently associated and ranged from .58 to .82 ($p < .0001$), with most correlations between .60 and .80. Grade 1 ORF correlated .72 in the winter and .82 in the spring with the Grade 1 SAT-10. For the second grade SAT-10, correlations with the five ORF assessments from winter of grade 1 through spring of grade 2 were .63, .72, .72, .79, and .80. Six ORF assessments from fall of Grades 2 through spring of Grade 3 correlated with the OSRA at .58, .63, .63, .65, .68, and .67. In addition, ORF intercept and slope predicted a statistically significant portion of performance on the Grade 2 SAT-10 ($p < .0001$) in first and second grade. ORF level and slope explained 70% of the variance on the SAT-10 high-stakes reading test at the end of Grade 2. In second and third grade, ORF intercept and slope also predicted a significant portion of performance on the third-grade OSRA ($p < .0001$). The ORF intercept and slope accounted for 52% of the variance on the OSRA. Although ORF slope accounted for a statistically significant amount of the variance in predicting the high-stakes measure, the contribution of slope in first and second grade was greater than in second and third grade.

Baker et al. (2008) recommended that DIBEL-ORF could be part of comprehensive assessment systems in which schools made decisions to screen students, monitor progress, and adjust instruction to meet students' overall reading needs. They discussed two reasons why ORF might provide a stronger index of overall reading proficiency in Grade 2 than Grade 3: the nature of reading development may be different in the two grades, and the ability of reading fluency to provide an overall index of reading proficiency may diminish over this period of time. These changes were typically more apparent when the grade difference was larger than 1 year, and

ORF should be administered once or twice a month for progress monitoring. The limitation of the study was that it did not focus on a subset of student populations, such as students in special education or in grades other than 1-3.

Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008)

Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008) evaluated the validity of the DIBELS-ORF measure for predicting performance on the *Florida Comprehensive Assessment Test* (FCAT-SSS) and *Stanford Achievement Test* (SAT-10, Madden, Gardner, Rudman, Karlsen, & Merwin, 1973), which are group-administered, criterion-referenced, standardized measures of reading comprehension. The participants included 35,207 third-grade students enrolled in *Florida Reading First Schools* during the 2004-2005 school year who were divided into two samples: n = 16,539 (S1) and n = 16,908 (S2). The groups were determined to be equitable on all demographic variables. Thirty-six percent of participants were White, 36% African American, and 23% Latino. Seventy-five percent were eligible for free or reduced-price lunch, and 17% had individual education plans. Eighty percent of students were identified as having English proficiency.

The data used in this study were obtained from the *Progress Monitoring and Reporting Network* (PMRN) as part of its role in providing support for statewide *Reading First* programs. The SAT-10 was administered at the end of each grade, whereas the FCAT-SSS and ORF were given only to third graders. Students in both samples were administered DIBELS-ORF four times per year by district-based assessment teams and also completed the FACT-SSS and SAT-10 during the same school year. A series of logistic regression analyses were conducted to determine if the ORF scores predicted performance on the FCAT-SSS outcome equally well for students with selected independent variables: free or reduced-priced lunch status, English language learner status, and race/ethnicity.

The results showed that moderate to strong correlations for both the S1 and S2 samples between third graders ORF, FCAT-SSS, and SAT-10 scores. The relationship between ORF and FCAT-SSS in both groups increased in magnitude over time, peaking at the ORF Winter 2 Assessment (rS1 = .71; rS2 = .70, $p < .01$). Similarly, the correlations between ORF scores and SAT-10 peaked at the Winter 2 Assessment, with comparable magnitudes to the relationship between ORF and FCAT-SSS (rS1 = .71; rS2 = .70, $p < .01$). Thus, ORF predicted reading comprehension performance on FCAT-SSS equally as well as with the SAT-10. Results of an ANOVA determined that the means of the two groups did not statistically differ on the selected outcomes. The authors concluded the DIBELS-ORF was significantly related to standardized measures of reading comprehension and identified students at risk for poor performance on the FCAT-SSS and SAT-10.

The authors discussed the finding that there was no evidence of predictive bias across several demographic groups. However, the data did not provide information on which home languages were spoken at home, so all students were identified as English language learners. More research is needed in the area of evaluating the implications of using the DIBELS-ORF as cut scores to identify students at risk for poor reading performance in school contexts.

Merino and Beckman (2010)

Merino and Beckman (2010) examined the possibility of a predictive relationship between CBM-R, Maze-CBM, and the *Measure of Academic Progress* (MAP) standardized test with a diverse population in Nebraska. The study participants included 376 elementary students in grades 2 through 5 from a Nebraska public school. Seventy-four percent of participants were

Hispanic, and 20% were Caucasian. Thirty-nine percent of students were ELL, and 15% were students with disabilities.

All participants were administered CBM-R, Maze-CBM, and MAP reading composite tests in the spring and fall of 2009. The MAP reading composite test measured word recognition, decoding, and comprehension via a computer-based and multiple-choice assessment. Multiple regression analyses were used to determine if any correlations existed among the scores of these three tests existed.

Both CBM-R alone and the combination of CBM-R and Maze-CBM were significantly predictive of performance on the MAP's reading test in all grades 2 through 5. However, Maze-CBM alone did not predict MAP's reading scores in any of the grades in the spring and fall of 2009 except fourth grade in the fall of 2009. Maze-CBM predicted performance on the MAP test in the fall of 2009 for fourth grade only. The results from this study provide an evidence that CBM-R is a significantly better predictor of MAP reading scores in both semesters. The correlations were also significant that CBM-R, Maze-CBM, and MAP scores in the spring of 2009 predicted their scores in the fall of 2009 (r = .624 - .720, $p < .05$).

The ORF alone and the combination of ORF and Maze scores significantly predicted MAP reading scores. The authors discussed that although the evidence obtained in this study supported the idea of using CBM-R to monitor students' progress and to target students who are at-risk of failing high-stakes reading tests, Maze-CBM was not a better predictor of performance on a Nebraska's standardized assessment than the CBM-R tool. This research strengthened that the CBM-R procedures were suggested for screening, progress monitoring, and instructional planning. The authors also recommended that future studies be conducted using larger sample sizes with different standardized assessments and demographics.

Marchand and Furrer (2014)

Marchand and Furrer (2014) conducted a study to evaluate the relationship between CBM-R and state-standardized reading scores. Data were collected from six schools located in high-risk neighborhoods in a large, urban school district in the Southwest area of Nevada. The final sample of participants consisted of 563 students. The 215 third graders, 240 fourth graders, and 134 fifth graders were approximately equally represented by gender. Fifty-seven percent of participants were Hispanic, the remaining participants were African American and Caucasian. Thirty-two percent of students were English language learners (ELLs), and nine percent had an individual education plans (IEPs).

Two subtests from AIMSweb were used to measure reading competence: CBM-R for oral reading fluency and Maze-CBM for reading comprehension measure. The *Nevada State Criterion Referenced Tests* (CRT) reading scores were obtained as a standardized assessment. Trained testers collected individual CBM-R data in October, and these testers also collected Maze-CBM data with group-administered procedures. District personnel administered CRT reading tests, and scores were obtained from district records. An ANCOVA was conducted to analyze the data by means, standard deviations, and correlations between the key variables.

The results showed that low reading performance was generally associated with ELLs, students with IEPs, boys, and minority students. The strongest negative associations with reading performance were ELL and IEP status in demographic variables (r = .29). The fall CBM reading score was significantly positively correlated with CRT standardized test in the spring (r = .71). All correlations among reading fluency, reading comprehension, and reading performance were significant at the $p < .001$ level for third, fourth, and fifth graders and accounted for approximately 29% of the variance over and above the demographic variables. The CBM-R data collected in this study predicted state-standardized reading performance scores.

Although not the purpose of review of literature, the study revealed a significant correlation between students' classroom engagement and higher reading scores, particularly for those with higher reading competence. Marchand and Furrer (2014) recommended the technical quality of CBM-R should be more closely investigated to determine differences with regard to this as well as other subgroups such as grade and socioeconomic level.

Summary

In this chapter, I reviewed the 10 selected studies that evaluated whether CBM-R correlated with other standardized measures of reading achievement for students in grades K-6. Findings of these quantitative studies are summarized in Table 1 on the next page, which presents the authors, participants, setting, methods, and general findings of each study. These findings are discussed in Chapter III.

Table 1

Summary of Chapter II Studies

Author (Date)	Participants/ Setting	Procedure	General Findings
Colon and Kranzler (2006)	50 fifth-grade students in North Central Florida	As participants read aloud reading probe in the 1-min period, the examiner counted the number of mispronunciations, omissions, and reversals with three different instructions. They were also administrated the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests of WJ-III in the school year.	1-year study. Specific instructions had a significant impact on CBM-R outcomes, and correlations between words read correctly per minute and WJ-III reading subtests were statistically significant.

Table 1 (Continued)

Author (Date)	Participants/ Setting	Procedure	General Findings
Shapiro, Keller, Lutz, Santoro, and Hintze (2006)	1,048 students in grades 3 to 5 from 1 urban and 1 suburban school districts in Eastern Pennsylvania	AIMSweb CBM data were collected for reading and math. PSSA was administered for third and fifth grades for both districts in the school year. In District 1, MAT-8 data in reading and math were obtained in Grade 4, and SDRT reading data were collected in Grade 5. In District 2, SAT-9 data were collected in reading and math in Grades 2 and 4.	1-year study. CBM-R had moderate to strong correlations with midyear assessments in reading and mathematics and both types of standardized tests across school districts. The data suggest that CBM-R can be one source of data that could be used to potentially identify those students likely to be successful or fail the high stakes assessment measure.
Silberglitt, Burns, Madyun, and Lail (2006)	5,472 students in grades 3, 5, 7, and 8 from 5 rural or suburban school districts in Minnesota	All students in grades 3, 5, 7, and 8 take one or more of MCA-R in Grade 3, 5, and 7or BST-R in Grade 8. Trained school personnel also administered CBM-R and Maze-CBM and collected CBM data in the same grades.	1-year study. The relationship between CBMs and state accountability test scores were moderate to strong correlations. No significant grade differences in relationship magnitude were found between the coefficients for state test scores to CBM-R and to Maze.
Uribe-Zarain (2006)	652 third-grade students from nine *Reading First* schools in Delaware	Data were obtained from both the DIBELS-ORF subtest and the reading portion of DSTP. Although the ORF was administered three times a year, the winter ORF score was selected because the reading DSTP was taken in March. The ORF outcomes were classified, and reading DSTP scale scores were classified into five performance levels.	2-year study. A significant correlation as a reliable predictor of reading performance existed between the measures of ORF in the DIBELS and reading DSTP. Although all the correlations among six groups were significant, the correlation coefficients varied widely.

Table 1 (Continued)

Author (Date)	Participants/ Setting	Procedure	General Findings
McIntosh, Graves, and Gersten (2007)	111 first-grade ELs in 4 classrooms from 3 schools in a large urban school district in southern California	Four different first-grade classrooms were observed during a 2.5 hour reading/ language arts between five and seven times from Year 1 to Year 2 using ELCOI. After observations, each teacher was interviewed for 30 mins and DIBELS was administered to measure ORF, and the Passage Comprehension subtest of WRMT-R was administered as a measure of comprehension.	2-year study. A strong correlation between the number of students below ORF benchmark thresholds at the end of first grade and the teacher rating on the amount of instruction provided for low performers was examined. Follow-up data at the end of third grade in ORF and comprehension indicated moderate correlations to first-grade scores.
Riedel (2007)	1,518 first-grade students in 26 schools from a large urban school district in Memphis, Tennessee	Sample completed the DIBELS tests at the beginning, middle, and end of their first-grade year and were assessed with a measure of reading comprehension at the end of first grade, GRA+DE and second grade, TerraNova Reading during the school year, and TerraNova was administered during the next school year to determine second-grade reading test outcomes.	2-year study. The relationship between DIBELS and reading comprehension at the end of first grade and second grade was correlated. DIBELS-ORF was a better predictor of comprehension than the remaining subtests. The relation between ORF and comprehension was stronger for ELL students than for non-ELL students.
Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, and Beck (2008)	Four cohorts of about 2,400 students in grades 1 to 3 from 34 *Reading First* schools in Oregon	Data were collected for two consecutive years. DIBELS-ORF was administered as part of benchmark testing during the year, and median scores was used as the representative performance scores. Two standardized tests were administered SAT-10 and OSRA.	2-year study. The correlation between ORF and SAT/ OSRA were consistently associated. Results supported the use of ORF in the early grades to screen students for reading problems, monitor reading growth over time, and predict performance on reading tests.

Table 1 (Continued)

Author (Date)	Participants/ Setting	Procedure	General Findings
Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008)	35,207 third-grade students in *Reading First* schools in Florida	Data were obtained from PMRN. The SAT-10 was administered at the end of each grade, whereas the FCAT-SSS and DIBELS-ORF were given only to third graders. Students in both samples were administered ORF four times per year by district-based assessment teams and also completed the FACT-SSS and SAT-10 tests during the same academic year.	1-year study. The validity of DIBELS-ORF for predicting performance on the FCAT-SSS and SAT-10 reading comprehension measures were proved. The recalibrated risk-level cut scores predicted performance on the FCAT-SSS equally well for students from different socioeconomic, language, and race/ethnicity
Merino and Beckman (2010)	376 students in grades 2 to 5 in Nebraska	Participants were administered AIMSweb (e.g., CBM-R, Maze-CBM), and MAP reading composite tests in the spring and fall of 2009. The MAP reading composite test measured word recognition, decoding, and comprehension via a computer-based and multiple-choice reading assessment.	1-year study. ORF and the combination of ORF and Maze scores significantly predicted MAP reading scores. The evidence suggested that ORF could be used to monitor students' progress and target students who were at-risk of failing the high-stakes reading tests.
Marchand and Furrer (2014)	563 students in grades 3 to 5 in 6 schools from a large, urban, southwestern school district in Nevada	Two subtests from AIMSweb were used to measure reading competence: CBM-R and Maze-CBM. The CRT reading scores were obtained as a standardized assessment. Trained testers collected individual CBMs data with group-administered procedures. District personnel administered CRT reading tests, and scores were obtained from district records.	1-year study. The relationship among CBMs and CRT reading scores as a standardized assessment were significantly correlated. The research demonstrated that CBM-R predicted student performance on year-end standardized reading tests.

Chapter III: Conclusions and Recommendations

Curriculum-based measurement of oral reading (CBM-R) is perhaps the most widely used and researched intervention for monitoring students' reading progress. This Starred Paper investigated how CBM-R correlated with scores from other standardized measures of reading achievement in various K-6 school settings and examined associated variables. This is important because reading is an essential skill for all students to learn in life, and reading fluency is a strong predictor of reading comprehension for students in grades K-6. In the first chapter, I discussed the importance of progress monitoring and curriculum-based measurement for instructional data-based decision making, problem solving shift, and response to intervention. In Chapter II, I presented the reviews of 10 empirical studies that examined the statistical correlation between CBM-R and other standardized measures of reading achievement. In this chapter, I discuss study findings and present recommendations for future research and implications for current practice.

Conclusions

Although the relationship between CBM-R and other standardized tests of reading achievement varied as a function of demographics and setting, all 10 research studies reported moderate to strong correlations between oral reading fluency (ORF) scores of CBM-R probes with those derived from other standardized tests of reading achievement (see Table 2). In addition to these correlation results, 4 of the 10 research studies supported the positive use of ORF in the K-6 level to identify students for reading problems (Baker et al., 2008; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). The findings of these studies provide support for the use of CBM-R with students in grades K-6.

Five of the 10 studies reviewed in Chapter II reported CBM-R could predict student performance on other standardized reading tests (Baker et al., 2008; Marchand & Furrer, 2014; McIntosh, Graves, & Gersten, 2007; Merino & Beckman, 2010; Roehrig et al., 2008). These studies reported moderately strong correlations between CBM-R outcomes and other standardized tests of reading achievement. Table 2 data show correlation coefficients ranged from .49 - .83, indicating CBM-R was a reasonable indicator of how well students are likely to perform across a wide range of standardized reading achievement tests.

Table 2

Reading Measure Correlations

Author (Year)	Type of CBM-R	Standardized Test	Reading Score Correlations (*r*)
Colon and Kranzler (2006)	CBM-R	WJ-III [a]	Comprehension: .465 - .813 Total Reading Score: .805 - .832
Shapiro, Keller, Lutz, Santoro, and Hintze (2006)	CBM-R	PSSA [b] SAT-9 [a] MAT-8 [a] SDRT [a]	Comprehension: .65-.67 Total Reading Score: .62-.72
Silberglitt, Burns, Madyun, and Lail (2006)	AIMSweb	MCA-R [b] BST-R [b]	Total Reading Score: .65-.68 (Eighth graders *r* =.51, third graders *r* =.71 for CBM-R and eighth graders .49, seventh *r* =.54 for Maze.)
Uribe-Zarain (2006)	DIBELS	DSTP [b]	Total Reading Score: .61 (Hispanic *r* = .723, White *r* = .624, African American *r* = .536, higher socioeconomic status levels *r* = .643, the group with no special education placement *r* = .579, $p < 0.01$)
McIntosh, Graves, and Gersten (2007)	DIBELS	WRMT-R [a]	Comprehension: .51 - .73
Riedel (2007)	DIBELS	GRA+DE [a] TerraNova [a]	Total Reading Score: .49 - .67 (ELL students *r* = .72 to .80, non-ELL students *r* = .49 to .67)

Table 2 (continued)

Author (Year)	Type of CBM-R	Standardized Test	Reading Score Correlations (*r*)
Baker et al. (2008)	DIBELS	SAT-10 [a] OSRA [b]	Total Reading Score: .58 - .82
Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008)	DIBELS	SAT-10 [a] FCAT-SSS [b]	Total Reading Score: .70 - .71
Merino and Beckman (2010)	AIMSweb	MAP [b]	Total Reading Score: .624 - .720
Marchand and Furrer (2014)	AIMSweb	CRT [b]	Total Reading Score: .70-.79

Note. [a] = Statewide standardized test, [b] = Nationwide standardized test.

[a] Statewide standardized test: PSSA = *Pennsylvania System of School Assessment*; MCA-R = *Minnesota Comprehensive Assessments-Reading*; BST-R = *Basic Standards Test-Reading*; DSTP = *Delaware Student Testing Program*; OSRA = *Oregon Statewide Reading Assessment,* FCAT-SSS = *Florida Comprehensive Assessment Test - Sunshine State Standards*; MAP = *Nebraska Measure of Academic Progress,* CRT = *Nevada State Criterion Referenced Tests).*

[b] Nationwide standardized test: WJ III = *Woodcock Johnson III* (Woodcock, McGrew, & Mathew, 2001); SAT = *Stanford Achievement Test* (Madden, Gardner, Rudman, Karlsen, & Merwin, 1973); MAT = *Metropolitan Achievement Test* (Prescott, Balow, Hogan, & Farr, 1984); SDRT = *Stanford Diagnostic Reading Test* (Karlsen, Madden, & Gardner, 1976); WRMT-R = *Woodcock Reading Mastery Test–Revised* (Woodcock, 1987); GRA + DE (Williams, 2001); *TerraNova* (CTB/McGraw-Hill, 2003).

Seven studies were conducted for 1 academic year, and four studies were conducted across 2 academic years (Baker et al, 2008; McIntosh et al., 2007; Riedel, 2007; Uribe-Zarain; 2006). The authors of these studies could not ascertain different outcomes between the 1- and 2-

year studies. Although there are no distinct differences between a 1-year and 2-year studies besides those among different grade levels (e.g., Silberglitt et al. (2006) and Baker et al (2008), further investigation is needed across more than two years.

Different assessment measures were used throughout the studies. Three different CBM-R measures were used to assess oral reading fluency and maze reading comprehension. Five studies used the DIBELS as a CBM-R probe (Baker et al., 2008; McIntosh et al., 2007; Riedel, 2007; Uribe-Zarain, 2006; Roehrig et al 2008), four studies selected AIMSweb (Shapiro et al., 2006; Silberglitt et al., 2006; Merino & Beckman, 2010; Marchand & Furrer, 2014), and one study created their own CBM-R probes (Colon & Kranzler, 2006). As indicated in Table 2, a variety of standardized reading tests including statewide standardized tests were used to evaluate the association with CBM-R. Regardless of whether it was a national or state standardized reading test, they correlated well with CBM-R as a consistent indicator of reading achievement.

Seven of 10 studies suggested CBM-R can be used to screen students at risk for reading failure or who may need additional services, monitor their progress, and adjust teachers' instruction to meet students' needs. Four of the 10 studies supported the positive use of ORF in the K-6 school level to identify students for reading problems (Baker et al., 2008; Riedel, 2007; Roehrig et al., 2008; Shapiro et al., 2006). Five of the 10 studies reported CBM-R could predict student reading performance on other standardized reading tests (Baker et al., 2008; Marchand & Furrer, 2014; McIntosh et al., 2007; Merino & Beckman, 2010; Roehrig et al., 2008). Thus, CBM-R can be part of comprehensive assessment for the purpose of making a range of decisions about students' reading growth.

In two studies (Baker et al., 2008; Silberglitt et al., 2006), significant differences were found among different grade levels. Stronger correlations were reported among CBM-R and lower grade groups than higher grade groups who were assessed on more specific grade-level standards. Silberglitt et al. (2006) indicated the coefficients between CBM-R and state test scores were significantly larger for third and fifth graders than those for eighth graders. Baker et al. (2008) indicated that the reason why ORF might provide a stronger index of overall reading proficiency in Grade 2 than Grade 3 was that the nature of reading development may be different in these grades. They speculated that reading fluency as an overall index of reading proficiency may diminish over a period of time. These changes were more apparent when the grade difference was larger than 1 year, and it implies that ORF should be administered once or twice a month for progress monitoring

In four studies, evidence did not support predictive bias derived from student demographic characteristics that included ELL, IEP, free or reduced-priced lunch status, and race/ethnicity (Marchand & Furrer, 2014; McIntosh et al., 2007; Roehrig et al, 2008; Uribe-Zarain, 2006). They examined correlations between CBM-R and other standardized tests of reading achievement across student demographic characteristics: Hispanic, White, African American, socioeconomic status level, and special education placement. Roehrig et al. (2008) and Uribe-Zarain (2006) found no evidence of predictive bias according to free or reduced-priced lunch status, English language learner status, and race/ethnicity. That is, CBM-R predicted performance on state-standardized test equally well for students with different demographic variables.

Marchand and Furrer (2014) and McIntosh et al. (2007) examined variables related to teacher instruction and classroom engagement. McIntosh et al. found a strong correlation between the number of students below DIBELS benchmark thresholds at the end of first grade and teacher rating results of the amount of instruction provided for low performers. Marchand

and Furrer (2014) reported a significant correlation between students' classroom engagement and higher reading scores, particularly for those with higher reading competence. Thus, the amount of teacher instruction and the degree of student classroom engagement were crucial to improve reading achievement.

The findings of the 10 selected studies in Chapter II are meaningful when one considers the low cost in terms of time and financial resources required to administer tests, monitor progress, and screen those who have reading problems or are likely at risk of failing high-stakes reading achievement goals. Overall, the literature showed consistent results in the relationship between CBM-R and other standardized measures of reading achievement across time, tests, sample size, state, and research term.

Recommendations for Future Research

After reviewing the literature regarding the relationship CBM-R and other standardized tests of reading achievement, further research should be focused on three limitations that require cautious interpretation of findings. First, the range of correlation coefficients between .49 - .83 warrants further investigation. Additional research in needed to determine the cause for lower correlation coefficients and the factors responsible for the lack of correlational consistency in research findings.

Another limitation is a possible bias for students of varying socioeconomic, language, and racial-ethnic backgrounds. Given the growing subgroup population in schools, it is important to replicate these findings with other subgroups of students in the future because CBM-R passages are likely to be biased not only in K-6 curriculum, but also in passages of commercially developed probes such as DIBELS, AIMSweb, or EasyCBM. For example, if a racial, ethnic, or cultural bias was in the curriculum, it could possibly exist in CBM probes and another measures of assessment would need to be considered together. Thus, the generalization issue from potential biases of the association between CBM-R and other standardized measures of reading achievement should be addressed to examine predictive bias and accumulate correlational evidences across other languages and academic areas.

Lastly, CBM studies should be continued to develop research to other academic areas and other languages because much research has focused on CBM-R in English to support its effectiveness. Thus, further research in other languages or countries is beneficial on whether the relationship between another version of CBM-R and other standardized tests of reading achievement is consistent in other languages. A focus should be placed on other unfocused areas that have not been considerably researched.

Implications for Current Practice

Teachers have various ways of monitoring students' achievement in reading. The results of this empirical research implies that CBM-R should be considered as one way to monitor students' reading performance because it is a research-based, curriculum-based method that is time- and cost-efficient. Research has also shown CBM-R to be effective in identifying at-risk students who need additional supports. It provides progress monitoring to assess students' academic performance, track their own progress, improve education accountability and effectiveness, and communicate with parents or other professionals about students' progress in general and special education at the K-6 grade level. The use of CBM-R continues to show students' improvement over time and as a potential measure of performance on standardized test measure. At present, it has become a common practice to provide preventive services for at-risk students who may need additional supports and successful progress with all students by implementing CBM-R in coordination with RTI.

Internet-based CBMs have been developed to reduce administration time, conduct data analysis, and provide specific information for teachers that would facilitate IEP progress monitoring. Technology-integrated CBMs help teachers to administer tests and analyze student performance. At the same time, instructional modifications and special education services must be made based upon individual student data and its cautious analysis. Data-based decision making for those who are in need of additional supports or special education services will not always result in student achievement unless the data are analyzed and appropriate instructional modifications are evaluated for individual students.

Summary

Over the past 40 years, research has focused more attention on the use of CBM-R as a scientifically based progress monitoring tool to test students' growth. The purpose of this literature review was to investigate the relationship of CBM-R as an indicator of reading achievement as compared with other standardized tests of reading achievement. Research results have shown a significant and strong overall correlation between CBM-R and other standardized tests of reading achievement. As a result, CBM has had an influence on general and special education fields to assess students' performance, monitor their progress frequently across time, and make decisions to improve instructional plans that will increase individual student progress. Based on the conclusions of the literature I reviewed and the limitations and recommendations, CMB-R is a powerful indicator of students' overall reading achievement, and the use of CBM is also expected to expand into other educational settings. It is also strongly recommended that CBM-R on student reading achievement and progress be implemented with instructional modifications by educators when making a quality decision are considered.

References

Christine A., E., Kristen L., M., Susan, R., & Miya Miura, W. (2012). *A Measure of Success: The Influence of Curriculum-Based Measurement on Education*. Minneapolis, MN: University of Minnesota Press.

Armbruster, B., Lehr, F., & Osborn, J. (2001). *Put Reading First: The Research Building Blocks for Teaching Children to Read K-3*. Washington, DC: The National Institute for Literacy.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'Enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*(1), 18.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., & Tilly III, W. D. (2005). *Response to intervention*. Alexandria, VA: National Association of State Directors of Special Education.

Burns, M. K., & Gibbons, K. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. New York, NY: Routledge.

Colon, E. P., & Kranzler, J. H. (2006). Effect of instructions on curriculum based measurement of reading. *Journal of Psychoeducational Assessment*, *24*, 318−328.

CTB/McGraw-Hill. (2003). *TerraNova second edition: California Achievement Tests technical report*. Monterey, CA: Author.

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

Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36*(2), 5-10.

Deno, S. L. (1997). Whether thou goest… Perspectives on progress monitoring. *Issues in educating students with disabilities, 16*, 213-235.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Remedial and Special Education, 37*, 184-192.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41,* 93-99.

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

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, *58,* 45-58.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, 39(5), 14-20.

Fuchs, L. S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading.* Washington, DC: National Center on Student Progress Monitoring. Retrieved from http:// www. studentprogress.org

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

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. N. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research & Practice, 22*(1), 13-24.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.

Greenwood, C. R., Kratchowill, T., & Clements, M. (2008). *Schoolwide prevention models: Lessons learned in elementary schools.* New York, NY: Guilford Press.

Greenwood, C. R., Carta, J. J., Baggett, K., Buzhardt, J., Walker, D., & Terry, B. (2008). Best practices in integrating progress monitoring and response-to-intervention concepts into early childhood systems. In A. Thomas, J. Grimes, & J. Gruba (Eds.), *Best practices in school psychology V* (pp. 535–548). Washington, DC: National Association of School Psychology.

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*(4), 541-556.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2012). *The ABCs of CBM: A practical guide to curriculum-based measurement.* New York, NY: Guilford Press.

Huba, M. E., & Freed, J. E. (2000). Learner centered assessment on college campuses: Shifting the focus from teaching to learning. *Community College Journal of Research and Practice, 24*(9), 759-766.

Hudson, R., Pullen, P., Lane, H., & Torgesen, J. (2009). The complex nature of reading fluency: A multidimensional view. *Reading and Writing Quarterly, 25*, 4-32.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly, 2*, 81-92.

Jenkins, J. R., Hudson, R. F., & Lee, S. H. (2007). Using CBM-reading assessments to monitor progress. *Perspectives on Language and Literacy, 33*, 11-18.

Justice, L. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech & Hearing Services in Schools, 37,* 284-297.

Karlsen, B., Madden, R., & Gardner, E. F. (1976). *Stanford Diagnostic Reading Test*. San Antonio, TX: Psychological Corporation.

Lai, M. K., & Hsiao, S. (2014). Developing data collection and management systems for decision-making: What professional development is required? *Studies in Educational Evaluation, 42*, 63-70.

Madden, R., Gardner, E. R., Rudman, H., Karlsen, B., & Merwin, J. C. (1973). *Stanford Achievement Test*. New York: Harcourt Brace.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. Educational Psychologist, 47, 71-85.

Marchand, G. C. & Furrer, C. J. (2014). Formative, informative and summative assessment: The relationship between curriculum-based measurement of reading (CBM-R), classroom

engagement, and reading performance. Psychology in the Schools, 51, 659-676.

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.

McIntosh, A. S., Graves, A., & Gersten, R. (2007). The effects of response to intervention on literacy development in multiple-language settings. *Learning Disability Quarterly, 30*, 197−212.

McMaster, K. L., & Wagner, D. (2007). Monitoring response to general education instruction.

In S. R. Jimerson, M. K. Burns, A. M. VanDerHeyden, S. R. Jimerson, M. K. Burns, A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 223-233). New York, NY, US: Springer.

Mercer, C. D., Mercer, A. R., & Pullen, P. C. (2011). *Teaching students with learning problems*. Upper Saddle River, NJ: Pearson.

Merino, K., & Beckman, T. O. (2010). Using reading curriculum-based measurements as predictors for the Measure Academic Progress (MAP) standardized test in Nebraska. *International Journal of Psychology: A Biopsychosocial Approach*, *6*, 85-98.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office. Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/pages/smallbook.aspx

Overton, T. (2006). Assessing learners with special needs: An applied approach. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Prasse, D. P., Breunlin, R. J., Giroux, D., Hunt, J., Morrison, D., & Thier, K. (2012). Embedding multi-tiered system of supports/response to intervention into teacher preparation. *Learning Disabilities--A Contemporary Journal*, *10*, 75-93.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1984). *Metropolitan Achievement Tests (Mat-6)*. San Antonio, TX: The Psychological Corporation.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to intervention. *The handbook of school psychology*, *4*, 434-460.

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

Riedel, B.W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*, 546−567.

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.

Sanetti, L. H., & Collier-Meek, M. A. (2015). Data-driven delivery of implementation supports in a multi-tiered framework: A pilot study. *Psychology in the Schools, 52,* 815-828.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24,* 19−35.

Shinn, M. R. (2007). Identifying Students at Risk, Monitoring Performance, and Determining Eligibility Within Response to. *School Psychology Review, 36*(4), 601-61.

Shinn, M. (2012). *The relation of AIMSweb, curriculum-based measurement, and the Common Core Standards: All parts of meaningful school improvement.* Austin, TX: Pearson Education.

Silberglitt, 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.

Spinelli, C. G. (2011). *Classroom Assessment for Students in Special and General Education* (3rd ed.). *Upper Saddle River, NJ: Pearson*.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15*, 128-134.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure, 52*, 48-58.

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17*, 223-237.

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRN Education, 2013*.

Tucker, J. (1987). Curriculum-based assessment is not a fad. *The Collaborative Educator, 1*, 4-10.

Uribe-Zarain, X. (2006). *Relationship between performance on DIBELS Oral Reading Fluency and performance on the Reading DSTP*. Newark, DE: Delaware Education Research & Development Center, University of Delaware.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning disabilities research & practice, 18*(3), 137-146.

Williams, K. T. (2001). *Technical manual: Group Reading Assessment and Diagnostic Evaluation.* Circle Pines, MN: American Guidance Service.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests, Revised*. Circle Pines, MN: American Guidance Services Inc.

Woodcock, R. W., McGrew, K. S., & Mathew, N. (2001). *Woodcock–Johnson III Test of Achievement*. Itasca, IL: Riverside Publishing.

Yell, M. L., Shriner, J. G., & Katsiyannis, A. (2006). Individuals with disabilities education improvement act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus on exceptional children*, *39*(1), 1-24.