2015

Technical Adequacy and Acceptability of Curriculum-Based Measurement and the Measures of Academic Progress

Stacy-Ann A. January
University of Nebraska–Lincoln, sjanuary2@unl.edu

Scott P. Ardoin
University of Georgia, spardoin@uga.edu

Published in *Assessment for Effective Intervention* 41:1 (2015), pp. 3–15; doi: 10.1177/1534508415579095

Technical Adequacy and Acceptability of Curriculum-Based Measurement and the Measures of Academic Progress

Stacy-Ann A. January, PhD[1] and Scott P. Ardoin, PhD[2]

1 University of Nebraska–Lincoln
2 University of Georgia

Corresponding author — Stacy-Ann A. January, Department of Special Education and Communication Disorders, University of Nebraska–Lincoln, 204 Barkley Memorial Center, Lincoln, NE 68583, USA; email sjanuary2@unl.edu

Abstract

Curriculum-based measurement in reading (CBM-R) and the Measures of Academic Progress (MAP) are assessment tools widely employed for universal screening in schools. Although a large body of research supports the validity of CBM-R, limited empirical evidence exists supporting the technical adequacy of MAP or the acceptability of either measure for universal screening. Purposes of the current study were to replicate and extend prior research by (a) examining the extent to which CBM-R performance measures more than word reading skills, (b) evaluating the concurrent validity of MAP with CBM-R, (c) determining the potential benefit of administering MAP with CBM-R for universal screening, and (d) examining teachers' acceptability of MAP and CBM-R. Participants included 802 students in Grades 1 to 5 who were administered three CBM-R probes and the MAP during universal screening and 86 elementary teachers who completed a universal screening assessments survey. Results provide evidence of the concurrent validity of MAP with CBM-R and suggest both measures are acceptable to teachers. Implications for research and practice are discussed.

Keywords: curriculum-based measurement, computer adaptive tests, reading, universal screening, teacher acceptability

Within the past 10 years, greater emphasis on accountability and the widespread adoption of Response to Intervention (RtI) frameworks has revolutionized the nature of assessment within schools. As of 2013, 34 states *allow* schools to consider RtI assessment data to determine special education eligibility and 17 states *require* the use of RtI assessment data, either as the sole method or in combination with other methods (Hauerwas, Brown, & Scott, 2013). One component of RtI models is the universal screening of all students three or four times per year. Purposes of universal screening include identifying those who are at risk for failing to meet academic goals (Mellard, McKnight, & Woods, 2009) and predicting which students are likely to pass high-stakes achievement tests (Hintze & Silberglitt, 2005). By identifying students early within RtI models, schools are able to intervene quickly, thus maximizing the likelihood that students will benefit from intervention and reducing the referral rates for special education services (VanDerHeyden, Witt, & Gilbertson, 2007). Two measurement procedures often employed for conducting screenings are curriculum-based measurement (CBM) and computer adaptive tests (CATs).

CBM in Reading (CBM-R)

The most common CBM instrument used in schools and examined within the research literature is CBM-R (Ball & Christ, 2012). CBM-R is a timed, individually administered assessment of oral reading rate with accuracy. Students read a passage aloud for 1 min while an examiner records errors, utilizing the number of words read correctly per minute (WRCM) as the outcome score. Although researchers originally developed CBM-R more than 30 years ago to assist special education teachers in establishing individual education goals and evaluating individual students' progress toward their goals (Deno, 2003), it is now widely used in universal screening within RtI models (Ball & Christ, 2012). A primary benefit of CBM-R is that it is a general outcome measure; as such, it assesses global proficiency within a curriculum across the year, instead of assessing mastery of a set of hierarchically organized subskills (Fuchs & Deno, 1991). In fact, research suggests that CBM-R requires the integration of the component skills of reading (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Furthermore, a meta-analysis of the literature indi-

cated that CBM-R has a strong relation with norm-referenced measures of global reading achievement and component reading skills (Reschly, Busch, Betts, Deno, & Long, 2009), further supporting its evidence as a general outcome measure.

Despite a long history of support demonstrating its technical adequacy and its wide use within schools for universal screening and progress monitoring purposes, one potential limitation associated with CBM-R is that teachers may perceive it simply as an indicator of word reading skills (Hamilton & Shinn, 2003). However, research suggests that CBM-R is more than merely a measure of students' word reading rates. For example, previous research indicated that fourth-grade students' rate of reading words in context (CBM-R) far exceeds their rate of reading words presented in lists and that their reading comprehension skills explained variance in CBM-R performance beyond that of word lists (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003b). Conversely, for less skilled fourth-grade readers, decoding skills accounted for more variance in CBM-R than for more skilled readers (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003a). In a replication and extension of Jenkins et al. (2003a, 2003b), first- and second-grade students' oral reading rate with accuracy was greater for connected text (i.e., CBM-R) than for sight-word lists (Ardoin et al., 2013). However, in contrast to Jenkins et al. (2003b), a high-frequency sight-word list explained more variance in CBM-R than did the measure of reading comprehension (Ardoin et al., 2013). Together, these results suggest that the relation between CBM-R and component reading skills (such as decoding and comprehension) may change with age and skill level. However, several studies employing different aged participants helped draw this conclusion. To date, researchers have yet to examine the unique contribution of decoding and reading comprehension skills to explain variance in CBM-R performance using a single sample of students across first through fifth grades. Differences in the association between CBM-R and reading comprehension as a function of grade level have implications for use in schools. Providing clear evidence that CBM-R measures more than word reading skill is an important step toward changing teachers' beliefs about CBM-R, use of CBM-R, and use of the resultant data.

CATs

Unlike traditional norm-referenced achievement tests, CATs provide students with a personalized set of multiplechoice test items, presented one at a time, tailored to students' achievement level (Olsen, 1990). By presenting items that are near students' achievement level instead of items that might be too easy or too difficult, resultant data provide a more precise measurement of students' achievement, as well as more detailed information regarding students' strengths and weaknesses. The adaptive nature of CATs typically results in a lower standard error of measurement when compared with traditional group-administered, norm-referenced tests (NRTs; van der Linden & Glas, 2000). Decreased standard error equates to a more accurate assessment and potentially greater practitioner confidence in the scores gleaned from such tests. In addition, because CATs are computer based, and thus procedures are predefined, test administration procedures are likely reliable across students.

The Measures of Academic Progress (MAP; Northwest Evaluation Association [NWEA], 2009) is a CAT based on the Rasch model of Item Response Theory. The Rasch model considers item difficulty and test-taker ability to estimate the probability that a student will be successful on a given item. Items are administered to each test taker, one at a time, adapting to the student's responses until the probability of the student responding accurately to each item is estimated to be 50%. Schools can administer MAP tests multiple times per year for screening and informing instruction. Although MAP data may provide schools with more detailed information regarding students' skills than CBM does (e.g., vocabulary, decoding), drawbacks include the expense per student, resources required (e.g., computers), and limited validity evidence. MAP scores are reported on an equal interval, vertically equated Rasch Unit (RIT) scale, which allows for comparisons of student growth within and across grade levels.

Despite the fact that almost eight million students complete MAP tests annually (NWEA, 2014), currently no peer-refereed study exists examining the validity of MAP with a norm-referenced reading achievement test. Merino and Beckman (2010) did, however, examine the relation between performance on CBM-R and the Reading MAP for students in Grades 2 to 5. Results indicated that spring CBM-R performance demonstrated moderately strong predictive validity with fall MAP scores (rs = .67 to .73), spring MAP scores demonstrated moderate predictive validity with fall CBM-R performance (rs = .62 to .66), and spring CBM-R scores in Grade 2 accounted for more variance in Grade 3 fall MAP scores than in any other grade. Despite both measures being administered at each assessment period, the researchers failed to report the concurrent validity between the two measures. Independent evidence of the concurrent validity of MAP with CBM-R equates to educators and administrators being able to have greater confidence that MAP scores are strong indicators of global reading achievement.

Acceptability of Universal Screeners

Within the applied behavior analysis research, Wolf (1978) contended that interventions must be socially significant and have acceptable procedures and outcomes. Acceptability, as it applies to school-based assessment, may be defined as

> the degree to which an assessment method is appropriate for analyzing a given problem, the fairness of the method, and its intrusiveness into teachers' and students' instructional time . . . [and] the degree to which the measures help in developing treatment plans. (Shapiro & Eckert, 1994, pp. 168–169)

If screeners are to be used effectively, teachers must find them acceptable (Glover & Albers, 2007). Furthermore, prior research indicates that teachers who find CBM procedures more acceptable are more likely to utilize the resultant data (Allinder & Oats, 1997).

Despite the importance of acceptability (Allinder & Oats, 1997) and an increase in the utilization of CBM-R for universal screening (Ball & Christ, 2012; Mellard et al., 2009), there are only two studies that provide relevant information regarding CBM-R's acceptability. The first study, published in 1995 by Eckert and Shapiro, examined teacher acceptability of two assessment methods, curriculum-based assessment (CBA) and NRTs. CBA is a set of assessment methods that measure students' achievement level within a school's curriculum (Tucker, 1985). CBM-R is one type of CBA. In Eckert and Shapiro (1995), teachers read a case summary with assessment data featuring either CBA or NRTs. After reading the case summary, teachers completed a questionnaire regarding their acceptance of CBA or NRTs. Findings indicated that teachers rated CBA as more acceptable than published NRTs.

The second acceptability study (Foegen, Espin, Allinder, & Markell, 2001) examined preservice teachers' beliefs regarding CBM-R. Participants received training on the validity and utility of CBM-R and then completed a 15-item Teacher Belief Survey. Results indicated that preservice teachers believed in the utility of CBM-R but doubted its validity as an estimate of reading comprehension. Unfortunately, these two studies (Eckert & Shapiro, 1995; Foegen et al., 2001) are somewhat dated as both were published more than 13 years ago and predate the emphasis in schools on measuring reading fluency (National Institute of Child Health and Human Development [NICHD], 2000) and the subsequent wide use of CBM-R within RtI models. Therefore, it is possible that teachers' acceptability of CBM-R has changed. In addition to CBM-R's acceptability evidence being outdated, teachers' acceptability of MAP has yet to be examined empirically.

Widespread adoption of RtI frameworks has led to an increase in the use of universal screening assessments. In a review of the literature examining teachers' use of CBM for instructional decision making, findings suggested that when teachers use CBM data to modify their instruction, students make greater gains in achievement than students of teachers who do not use CBM data to inform instruction (Stecker, Fuchs, & Fuchs, 2005). Although teachers' use of CBM-R data improves student outcomes, the extent of teachers' knowledge of the characteristics and benefits of CBM-R has not been examined. Furthermore, despite the broad use of MAP for universal screening, no studies have evaluated teachers' knowledge of MAP or how teachers use MAP data.

Purposes of the Study

To demonstrate utility, universal screeners must be technically adequate (Hosp & Ardoin, 2008), be acceptable to teachers, and have adequate classification accuracy (Glover & Albers, 2007). The current study aimed to replicate and extend the universal screening literature by examining one aspect of technical adequacy (i.e., concurrent validity) and the acceptability of MAP and CBM-R. Despite the wide use of MAP and CBM-R, empirical evidence of the concurrent validity of MAP with CBM-R is limited and published research has not examined this relation with first-grade students (Merino & Beckman, 2010). Furthermore, Merino and Beckman (2010) administered grade-level probes from AIMSweb (Howe & Shinn, 2002), a universal screening and progress monitoring assessment system that provides CBM probes in reading. AIMSweb probes are not equated vertically across grade levels, resulting in the inability to make comparisons across grades (Howe & Shinn, 2002). Therefore, the current study extended prior research by (a) examining the concurrent validity of MAP with CBM-R, (b) employing first-grade students, and (c) using a different set of probes (Formative Assessment Instrumentation and Procedures for Reading [FAIP-R]; Christ, Ardoin, & Eckert, 2010), which are equated both vertically and horizontally, allowing for comparisons across grade levels.

Although CBM-R is widely used for universal screening, teachers may consider it as solely a measure of word reading rates (Hamilton & Shinn, 2003). However, extant research indicates that reading comprehension accounts for a larger portion of the variance in CBM-R scores than word lists (Jenkins et al., 2003b) and that the association between CBM-R and reading skills (decoding, reading comprehension) may differ based on reading skill and age (Ardoin et al., 2013; Jenkins et al., 2003a). Therefore, a second purpose of this study was to examine the unique contribution of reading comprehension (after controlling for decoding skills) in explaining variance in CBM-R performance for students in first through fifth grades.

Some schools conduct universal screening using both CBM-R and MAP, which may have utility to the extent that combined resultant data provide more information about students' reading achievement than each assessment independently. Therefore, the third purpose of the present study was to examine the unique benefit, if any, of administering both CBM-R and MAP when conducting universal screening in third grade. Third grade was selected due to the availability of concurrent standardized test scores at this grade. Finally, not only must universal screening measures be technically sound (Hosp & Ardoin, 2008), but they must also be acceptable to

teachers (Glover & Albers, 2007); however, no published studies examining teacher acceptability of MAP exist. Therefore, the final purpose of the current study was to examine the extent of teachers' knowledge, use, and acceptance of MAP and CBM-R for universal screening.

Method

Participants and Settings

Students. Participants were 802 students in first (n = 158), second (n = 158), third (n = 147), fourth (n = 159), and fifth (n = 180) grade from two public elementary schools in one suburban district located in the Southeast. The two schools were recruited to participate in the study because they utilized both CBM-R and MAP for universal screening. Given that collecting universal screening data was already a part of the schools' typical educational practices, all of the students in first through fifth grades in the participating schools who were present during the winter universal screening period were included as participants. Students participated in this study anonymously; therefore, individual demographic data were not obtained.

School A enrolled approximately 480 students in Grades K–5 and School B, a Title 1 school, enrolled approximately 500 students in Grades K–5. The demographic information available for School A indicated that approximately 81% of the students were White, 6% Hispanic, 5% Asian, 4% Black, and 4% multiracial. Schoolwide, 20% of students qualified for free or reduced-priced meals and 11% were enrolled in special education. Demographic data reported for School B indicated that 88% of the student population were White, 5% Hispanic, 3% Black, 3% multiracial, and 1% Asian. Approximately 26% of students in School B qualified for free or reduced-priced meals and 9% were enrolled in special education.

Teachers. A total of 86 teachers from five elementary schools (inclusive of the two schools described above) located within the same district participated in the current study. All teachers self-reported their demographic information, but not all teachers responded to each question. According to self-report, 89.5% of teachers were female and ranged in age from 23 to 63 years (M = 44 years, SD = 9.5 years). Regarding education, 25.6% (n = 22) held a bachelor's degree, 45.3% (n = 39) earned a master's degree, 22.1% (n = 19) held an educational specialist degree, and 3.5% (n = 3) held doctorate degrees. There were 71 general education teachers who taught first (n = 15), second (n = 13), third (n = 16), fourth (n = 11), fifth (n = 14), or multiple (n = 2) grades. Ten respondents reported teaching special education and three reported teaching both general and special education. On average, teachers had 17 years of experience (range, 1 to 39 years), had used CBM-R for 4.9 years (SD = 3.9 years; range, 1 to 20 years), and used MAP for 2.1 years (SD = 0.7 years; range, 1 to 4 years). Teachers reported receiving an average of 1.2 hr (SD = 1.7 hr; range, 0–8 hr) of training on CBM-R and 5 hr (SD = 6 hr; range, 0–40 hr) of MAP training, which was all provided by the participating district, with 45.3% of teachers reporting being trained in CBM-R by the district.

Measures

CBM-R. The three universal screening CBM-R probes at each grade level from the FAIP-R (Christ et al., 2010) passage set were utilized for the current study. FAIP-R CBM-R probes are unique in that as opposed to relying on readability formulas for passage equating, passage equivalence was based on the field-testing of passages with more than 500 students reading all passages within each of three specified difficulty levels. These three passage levels aligned with the following grades: first grade (Level 1), second/third grade (Level 2), and fourth/fifth grade (Level 3). FAIP-R passages were equated within grade levels and vertically to allow for comparisons of student performances across grade levels. All students, regardless of grade level, read three "anchor" passages prior to reading the passages at their specified grade level. The FAIP-R vertical scale was created by using a common-item (i.e., anchor passages) nonequivalent groups (i.e., across achievement levels) equating design with a linear equating method. Then scores on the anchor passages were used to adjust scores for the remaining passages, which resulted in a vertical scale (Albano & Christ, 2012; Albano & Rodriguez, 2012; Kolen & Brennan, 2004). The median and transformed (vertical scale) WRCM for the probes were used as the dependent scores. The internal consistency of FAIP-R CBM-R probes ranges from .88 to .93 and the alternate form reliability ranges from .87 to .95 across levels (Christ et al., 2014).

MAP. The MAP is a CAT published by NWEA. The two schools from which the data were collected administered the MAP (Grades 2–5) and MAP Primary Grades (MPG; Grade 1) Reading Survey with Goals tests. The MAP Reading Survey with Goals provides an overall composite score and scores for each goal area: (a) foundational skills and vocabulary (e.g., decoding skills), (b) informational texts (e.g., comprehension of expository passages), and (c) literature texts (e.g., comprehension of narrative passages). MPG Reading Survey with Goals includes an overall score and scores for the following goal areas: (a) foundational skills (e.g., decoding skills), (b) vocabulary use and functions, and (c) literature and informational texts (e.g., comprehension of expository and narrative passages). MAP assessments include approximately 42 items with a 1-hr administration time (NWEA, 2009). Students' RIT scores, which may

range from 120 to 250 for the MAP and the MPG, were utilized as the dependent scores. According to the technical manual, the marginal reliability of the MAP ranges from .94 to .95 for students in Grades 2 to 5 and from .94 to .97 for the MPG for students in Grades K–2 (NWEA, 2009). Test–retest reliability ranges from .70 to .85 for students in Grades 2 to 10 and from .71 to .86 for Grade 1. The concurrent validity between the MAP Reading Survey with Goals and various state achievement tests is moderate to strong, ranging from .57 to .79 for students in Grades 2 to 5 (NWEA, 2009).

Iowa Test of Basic Skills (ITBS®). The ITBS (University of Iowa, 2005) is a group-administered, norm-referenced achievement test for students in Grades K–8 published by Riverside Publishing. The participating district administered Form A of ITBS to third-grade students only. For the purposes of this study, the ITBS-Total Reading (ITBS-TR) composite was employed, which includes two subtests that measure students' reading comprehension and vocabulary skills. Students' standard scores for the ITBS-TR composite were utilized as the dependent score. According to the ITBS technical manual, the Total Reading composite has a Kuder Richardson-20 internal consistency of .94 (University of Iowa, 2005).

Teacher survey. A 55-item, investigator-developed survey was used to measure the extent of teachers' knowledge, use, and acceptance of CBM-R and MAP. The survey was adapted from the Teacher Belief Survey (Foegen et al., 2001) and the CBM-Acceptability Scale (Allinder & Oats, 1997) employed in prior research. Of the 55 items, 15 items were True/False, 36 items were Likert-type items, 2 items were multiple-choice questions, and 2 items were openended questions. For the purpose of this study, the acronym ORF (oral reading fluency) was used in the questionnaire (instead of CBM-R), given that ORF was the term used by teachers. The first 15 items aimed to assess teachers' knowledge of MAP and CBM-R. The next 18 items evaluated teachers' use of CBM-R and MAP via a 6-point Likert-type scale. Response options ranged from *never* (1) to *daily* (6). Items on the Use subscale included, "I use [ORF/MAP] data to get a good estimate of students' reading comprehension" and "I use [ORF/MAP] data to determine which students are struggling readers." The remaining 18 Likert-type items measured teacher acceptability of CBM-R and MAP on a 6-point scale, on which response options ranged from *strongly disagree* (1) to *strongly agree* (6). Items on the Acceptability subscale included, "[ORF/MAP] is practical in the amount of time required for administration" and "I like having [ORF/MAP] data for the students I teach." Each completed survey yielded a percent accuracy of knowledge regarding CBM-R and MAP and summed scores for CBM-R and MAP in the domains of use and acceptability. Internal consistency of the survey, as estimated by coefficient alpha, was adequate for the entire survey ($\alpha = .92$), the Use subscale ($\alpha = .91$), and the Acceptability subscale ($\alpha = .87$). Furthermore, internal consistency of the CBM-R items ($\alpha = .92$) and MAP items ($\alpha = .93$) was also adequate.

Procedures

Universal screening data were collected across three sessions. First, classroom teachers administered the ITBS to all third graders following standardized procedures during the second week of the fourth month of the school year. Next, the MAP was administered via the computer to all students during the last week of the fourth month through the first two weeks of the fifth month of the year. Finally, all firstthrough fifth-grade students were individually administered three grade-level universal screening CBM-R probes by graduate and undergraduate examiners, during the second week of the fifth month of the year, in the hallways adjacent to their classrooms. For the first of the three CBM-R probes, examiners read scripted instructions that informed students of the title of the passage, where to start reading (i.e., top left of the page), the direction in which to read (i.e., across the page), and to do their best reading when instructed to begin. For the two remaining probes, students were told the title of the passage and to do their best reading. Passages were scored by subtracting the total number of words read by the number of errors to produce the WRCM. Teachers were recruited for the study during after-school faculty meetings at the participating elementary schools. At these meetings, the details of the study were explained to potential participants, and teachers who agreed to participate received the consent documents and a copy of the survey. Participants completed the survey in 10 to 15 min. All participants were entered into a drawing to win one of two US$25 e-gift cards.

Procedural Integrity and Interscorer Agreement

Third-grade teachers who administered the ITBS were trained in standardized administration procedures by their district. CBM-R probes were administered by graduate students enrolled in a school psychology doctoral program and undergraduate research assistants who all had extensive experience with CBM-R standardized administration procedures. During one half-hour training session led by the primary author, examiners were provided with an overview of the study and a review of administration and scoring procedures. Each CBM-R administration was audio recorded. Approximately, 20% of these audio recordings were randomly selected for the evaluation of interscorer agreement and procedural integrity. Interscorer agreement was calculated by dividing the number of word-by-word agreements by the number of agreements plus disagreements, multiplied

by 100%. Interscorer agreement averaged 99% (range, 93%–100%). Procedural integrity was calculated by dividing the number of accurately completed steps by the number of total steps (17), multiplied by 100%. Procedural integrity averaged 98% (range, 94%–100%) across examiners.

Results

Analytic Method

To examine the relation between CBM-R, MAP, and ITBS-TR within and across grade levels, Pearson product– moment correlations were employed. Fisher's *Z* transformations were conducted to examine potential differences in correlation coefficients between CBM-R and MAP as a function of grade level. Hierarchical multiple regression analyses were utilized to determine the shared and unique contribution of MAP subtests in explaining CBM-R performance and the unique contribution of CBM-R and MAP scores in explaining variance in ITBS-TR performance. For the purposes of making statistical comparisons across grade levels, students' transformed WRCM scores were utilized. To control for Type 1 error, a Bonferroni correction was applied and an alpha level of .006 was used across all eight regression models.

For the survey data, descriptive statistics and frequencies were obtained for each survey question and the sum of responses for each domain. Then, three paired-samples *t* tests were used to examine differences in the summed knowledge, use, and acceptability scores for MAP and CBM-R. For the multiple-choice items, the percentage of respondents who selected each response was calculated. For the open-ended items, participants' responses were recorded and coded (1–9) based on the common themes that were present across responses. Finally, the number of responses for each theme was summed and the percentage of respondents identifying each theme was calculated.

Descriptive Statistics: Universal Screening Data

All variables were examined for normality, kurtosis, and skew; all values were found to be within normal limits. Furthermore, very few instances of missing data existed (0.04%), thus data were assumed to have been missing at random and cases were deleted listwise prior to conducting analyses. Descriptive statistics for CBM-R and MAP are presented in Table 1. Correlational analyses yielded statistically significant correlations among all experimental variables. Within each grade, the magnitude of the correlations between CBM-R and MAP overall RIT scores were high (range, .72–.79). Fisher's *Z* comparisons yielded no statistically significant differences between correlational coefficients, suggesting that the relation between CBM-R and MAP did not differ significantly between grade levels. Across Grades 2 to 5, CBM-R was strongly related to the MAP overall RIT score ($r = .83$) and goal areas of foundational skills and vocabulary ($r = .79$), informational texts ($r = .80$), and literature texts ($r = .78$; see Table 2).

MAP Goal Areas as Components of CBM-R

Hierarchical regression analyses were conducted to examine the unique contribution of MAP goal areas to CBM-R within each grade (see Table 3). For first graders, the foundational skills and vocabulary use and functions subtests were entered in the first block, and the literature/informational texts subtest was entered in the second block. Results indicated that foundational skills and vocabulary accounted for 45% of the variance in CBM-R scores and the additional variance explained by literature/informational texts (4%) approached significance ($p = .035$). Due to differences in MAP goal areas, hierarchical regression analyses for Grades 2 to 5 were slightly different than for first grade. For this set of regressions, CBM-R performance was predicted by entering foundational skills/vocabulary in the first block and the literature texts and informational texts in the second block. Across Grades 2 to 5, foundational skills and vocabulary explained a significant portion of the variance in CBM-R (63%) and, together, informational and literature texts accounted for an additional 6% ($p < .001$) of the variance in CBM-R. Within grade level, findings were consistent with across grade-level results and are presented in Table 3.

MAP and CBM-R as Components of Reading Achievement Among Third-Grade Students

For students in third grade, both MAP overall RIT scores and median CBM-R scores were highly related to the ITBS-TR composite. A statistically significant difference between the CBM-R and ITBS-TR ($r = .74$) and MAP and ITBS-TR ($r = .87$) correlations was observed, $z = –3.25$, $p = .001$. Two hierarchical multiple regression analyses were conducted to evaluate the unique contribution of CBM-R and MAP overall RIT scores in third-grade students' ITBS-TR performance (see Table 4). To determine the unique contribution of each measure in explaining achievement scores, CBM-R was entered first in one regression and MAP was entered first in the second regression. Therefore, results were used to determine whether either measure explained variance beyond the other, allowing for evaluation of whether both should be administered. Together, CBM-R and MAP explained 76% of the variance in the ITBS-TR composite ($p < .001$). When CBM-R scores were entered first, results indicated that CBM-R accounted for 55% of the variance in students' ITBS-TR scores and MAP accounted for an additional 21% of the variance ($p < .001$). When MAP was entered first, it accounted for 75% of the variance in ITBS-TR ($p < .001$), and CBM-R failed to explain any significant unique variance ($p > .05$).

Table 1. Means and Standard Deviations Among Study Variables for Grades 1 to 5.

Grade	First[a]		Second[b]		Third[c]		Fourth[d]		Fifth[e]	
Variables	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Median CBM-R	75.75	(39.32)	109.68	(40.93)	128.59	(36.07)	155.06	(35.48)	148.98	(35.80)
Transformed CBM-R	35.34	(19.17)	92.21	(37.16)	131.57	(38.09)	146.17	(31.99)	159.42	(33.28)
MPG overall	176.87	(13.56)								
MPG foundational skills	176.77	(16.59)								
MPG vocabulary use and functions	177.91	(13.62)								
MPG literature and informational texts	174.30	(16.95)								
MAP overall			190.63	(15.09)	202.41	(14.88)	210.94	(11.20)	214.78	(13.73)
MAP foundational skills and vocabulary			190.41	(14.89)	201.76	(15.21)	210.70	(11.36)	214.05	(13.93)
MAP literature texts			191.58	16.61)	202.96	(15.26)	210.94	(13.26)	214.80	(15.01)
MAP informational texts			189.80	(16.59)	202.25	(16.87)	211.23	(12.27)	215.39	(15.06)
ITBS-TR					188.58	(26.05)				

CBM-R = curriculum-based measurement in reading; MPG = Measures of Academic Progress Primary Grades; MAP = Measures of Academic Progress; ITBS-TR = Iowa Test of Basic Skills–Total Reading Standard Score.

a: $n = 158$; b: $n = 158$; c: $n = 147$; d: $n = 159$; e: $n = 180$

Table 2. Intercorrelations Among Study Variables for Grade 1 and Across Grades 2 to 5.

Variables	1	2	3	4	5
Grade 1					
1. Median CBM-R	—				
2. MPG overall	.73**	—			
3. MPG foundational skills	.65**	.92**	—		
4. MPG vocabulary use and functions	.60**	.90**	.75**	—	
5. MPG literature and informational texts	.64**	.93**	.81**	.79**	—
Grades 2 to 5					
1. Transformed CBM-R	—				
2. MAP overall	.83*	—			
3. MAP foundational skills and vocabulary	.79*	.94*	—		
4. MAP literature texts	.78*	.96*	.85*	—	
5. MAP informational texts	.80*	.96*	.86*	.89*	—

CBM-R = curriculum-based measurement in reading; MPG = Measures of Academic Progress Primary Grades; MAP = Measures of Academic Progress.

* $p < .01$; ** $p < .001$

Acceptability Data

The average percent accuracy for the questions within the knowledge domain is presented in Table 5 and means, standard deviations, and medians for the Likert-type questions within the use and acceptability domains are presented in Table 6. Three one-way ANOVAs were conducted to examine differences in participant responses based on selfreported demographic variables. Results indicated that respondents differed in their responses to the knowledge and use items based on their training on MAP.

Knowledge of universal screening assessments. Teachers' percent accuracy was greater on the MAP items ($M = 92.6\%$, $SD = 14.82$) than on CBM-R items ($M = 58.1\%$, $SD = 21.67$), $t = 12.46$, $p < .001$. Responses to the most frequently missed items indicated that a majority of respondents did not know that CBM-R is a good predictor of reading comprehension or that students can make greater gains in reading if CBM-R is used to evaluate the impact of instruction on student performance.

Use of universal screening assessments. Overall, teachers reported using MAP data ($M = 31.25$, $SD = 7.69$) significantly more than CBM-R data ($M = 24.92$, $SD = 9.41$), $t = 5.88$, $p < .001$. Specifically, teachers most frequently indicated using MAP data for providing differentiated instruction to individual students ($M = 4.28$, $SD = 1.24$) and planning their classroom

Table 3. Summary of Hierarchical Regression Analyses Using MAP Goal Areas to Predict CBM-R Performance for Grades 1 to 5.

Grade	Predictor and steps	R^2	ΔR^2
1	1 Foundational skills vocabulary	.45*	
	2 Literature and informational texts	.47	.04
2	1 Foundational skills and vocabulary	.56*	
	2 Literature texts Informational texts	.60*	.04*
3	1 Foundational skills and vocabulary	.51*	
	2 Literature texts Informational texts	.61*	.10*
4	1 Foundational skills and vocabulary	.39*	
	2 Literature texts Informational texts	.52*	.13*
5	1 Foundational skills and vocabulary	.46*	
	2 Literature texts Informational texts	.55*	.09*

MAP = Measures of Academic Progress; CBM-R = curriculum-based measurement in reading.
*$p < .001$.

Table 4. Summary of Hierarchical Regression Analyses Using MAP and CBM-R to Predict ITBS-Total Reading in Third Grade.

Model	Predictor and steps		R^2	ΔR^2
A.	1	CBM-R	.55*	
	2	MAP	.76*	.21*
B.	1	MAP	.75*	
	2	CBM-R	.76	.01

MAP = Measures of Academic Progress; CBM-R = curriculumbased measurement in reading; ITBS = Iowa Test of Basic Skills.
* $p < .001$

reading instruction ($M = 3.86$, $SD = 1.26$). In contrast, teachers reported using results from CBM-R universal screening most often to identify struggling readers ($M = 3.31$, $SD = 1.18$) and to determine which students were in need of additional instruction/intervention in reading ($M = 3.19$, $SD = 1.15$). In addition, most respondents reported rarely using CBM-R data to obtain an estimate of their students' reading comprehension skills ($M = 1.70$, $SD = 1.23$).

Acceptability of universal screening assessments. Results indicated high degrees of acceptability for both MAP ($M = 37.51$, $SD = 6.95$) and CBM-R ($M = 37.39$, $SD = 8.38$) that did not differ significantly, $t = .198$, $p = .844$. Most respondents reported that they liked having MAP data for their students ($M = 5.15$, $SD = .89$) and would suggest that other teachers also use MAP data to make instructional decisions ($M = 4.74$, $SD = 1.03$). Teachers reported that they think that collecting MAP data is beneficial for students ($M = 4.88$, $SD = .93$) and believe that MAP is appropriate for evaluating their students' academic problems ($M = 4.76$, $SD = .94$), reading skills ($M = 4.72$, $SD = 1.06$), and reading progress ($M = 4.72$, $SD = .99$). Regarding CBM-R, most teachers reported that they like having CBM-R data for their students ($M = 4.48$, $SD = 1.05$), think it is practical in its administration time ($M = 4.67$, $SD = .86$), and think that collecting CBM-R data is beneficial for students ($M = 4.36$, $SD = 1.06$). Teachers were more likely to believe that CBM-R should have a comprehension question at the end ($M = 2.30$, $SD = 1.28$) and perceive it as only a measure of students' decoding skills ($M = 2.91$, $SD = 1.12$).

Multiple-choice and open-ended items. When asked to select one screening instrument (MAP or CBM-R) that best measures comprehension and reading achievement, most teachers reported that MAP best measures students' reading comprehension (95%) and global reading achievement (82%). Teachers also responded to two open-ended questions that asked what they liked best about MAP and best about CBM-R. The majority of respondents (70%) indicated that they most liked the detailed information that MAP provides about students' reading skills. Finally, most respondents (60%) indicated that they liked that CBM-R is a quick assessment.

Discussion

Schools across the country employ CBM-R and MAP in their universal screening efforts as a part of the implementation of an RtI framework (Ball & Christ, 2012; NWEA, 2014). By screening all students, educators can, among other things, identify those students who are "at-risk" and predict students' performance on standardized tests (Hintze & Silberglitt, 2005; Mellard et al., 2009). Some schools administer both MAP and CBM-R to students, despite the fact that the benefit of administering both instruments for universal screening is unknown. The technical adequacy of CBM-R is well established (Reschly et al., 2009; Wayman, Wallace, Wiley, Ticha, & Espin, 2007); however, to date, only one published study (Merino & Beckman, 2010) has evaluated the predictive validity of MAP with CBM-R as a universal screener for students in Grades 2 to 5. Although there is some evidence of CBM-R's acceptability (Eckert & Shapiro, 1995; Foegen et al., 2001), this research is outdated and no evidence exists regarding the acceptability of MAP for universal screening. The lack of acceptability research is also problematic because evidence suggests that teachers with higher degrees of acceptance use assessments more frequently (Allinder & Oats, 1997). Research also suggests that CBM-R measure more than word reading rates (Jenkins et al., 2003b) and that the relation between CBM-R and component reading skills (e.g., decoding, reading comprehension) may change

Table 5. Teachers' Knowledge of CBM-R and MAP.

Item	% accuracy
Using ORF data helps teachers to know if students are benefiting from their reading instruction	68.6
Using MAP data helps teachers to know if students are benefiting from their reading instruction	93.0
ORF is useful for grouping students	60.5
MAP data are useful for grouping students	94.2
When teachers use ORF data to evaluate students' instruction, they often make greater achievement gains	40.7
When teachers use MAP data to evaluate students' instruction, they often make greater achievement gains	94.2
ORF provides me with information about students' general level of reading proficiency	79.1
MAP provides teachers with information about students' general level of reading proficiency	89.5
When teachers use ORF data, it helps them to know if students are making gains in reading achievement	66.3
When teachers use MAP data, it helps them to know if students are making gains in reading achievement	93.0
MAP provides teachers with information about how well students comprehend	91.9
ORF provides me with information about how well a student comprehends	9.3
ORF is not a good predictor of students' comprehension	39.5
ORF provides me with information about how well students are able to decode words	82.6
Assessing students' fluency on word lists is just as informative as assessing ORF	76.7

CBM-R = curriculum-based measurement in reading; MAP = Measures of Academic Progress; ORF = oral reading fluency.

with age and reading skill (Ardoin et al., 2013; Jenkins et al., 2003a). As such, the current study extends previous research by (a) examining the unique contribution of reading comprehension in explaining variance in CBM-R performance, (b) evaluating the concurrent validity of MAP with CBM-R and including first-grade students, (c) examining the unique contributions of MAP and CBM-R in explaining variance in third graders' reading achievement, and (d) evaluating the teacher acceptability of MAP and CBM-R. In general, findings from the current study replicate and extend the CBM-R literature and provide additional empirical evidence of the validity of MAP.

Technical Adequacy of Universal Screeners

Findings from the current study add to the wealth of CBM-R literature supporting its use as a universal screener for elementary school students (e.g., Wayman et al., 2007). Furthermore, regression analyses suggested that CBM-R explained a large portion of the variance in third-grade students' ITBS-TR performance. This study also extends existing CBM-R research suggesting that CBM-R is not merely a measure of students' word reading rates. Consistent with prior research suggesting that CBM-R measures more than word reading rates (Jenkins et al., 2003b), after controlling for foundational reading skills and vocabulary, the two MAP subtests designed to measure components of reading comprehension (i.e., literature texts and informational texts) explained a significant amount of variance in CBM-R both across and within Grades 2 to 5. Despite supporting previous research, the amount of variance explained by reading comprehension in Jenkins et al. (2003b) is greater in magnitude than in current findings, which is likely partially due to differences in the instruments employed. Whereas Jenkins et al. utilized the ITBS reading comprehension subtest, the MAP goal areas employed in the current study measured students' understanding of the craft and structure of texts in addition to measuring reading comprehension. When examining data for the first-grade sample, results were similar to those reported by Ardoin et al. (2013) in that after controlling for basic reading skills MAP's measure of comprehension (i.e., literature and informational texts) failed to explain unique variance in CBM-R.

By employing one sample of students in Grades 1 to 5, the current study also extends prior research suggesting that the relation between CBM-R and reading comprehension varies by grade level (Ardoin et al., 2013; Jenkins et al., 2003a, 2003b). As compared with data reported in previous research, findings suggest that reading comprehension (MAP Literature Texts and Informational Texts) explained more variance in CBM-R for students in Grades 3 to 5 than for students in Grades 1 to 2. Results are also consistent with prior research suggesting that reading comprehension explains less variance in CBM-R for students in the lower elementary grades (Ardoin et al., 2013). Grade-level differences may be due, in part, to the likelihood that students in the upper elementary grades receive less instruction (if any) in foundational reading skills, because it is assumed they have mastered those basic skills by third grade.

Table 6. Teachers' Use and Acceptability of MAP and CBM-R.

	MAP			CBM-R		
Items	*M*	Median	*SD*	*M*	Median	*SD*
Use scale						
Differentiating instruction	4.28	4.00	1.24	2.93	3.00	1.45
Identifying struggling readers	3.33	4.00	1.14	3.31	3.00	1.18
Evaluating reading progress	3.31	3.00	0.96	3.05	3.00	1.25
Planning instruction	3.86	4.00	1.26	2.80	3.00	1.37
Collecting the data if they were not required	3.04	3.00	1.20	2.95	3.00	1.23
Estimating reading comprehension	3.16	3.00	0.98	1.70	1.00	1.23
Identifying strengths and weaknesses	3.65	3.00	1.03	2.59	3.00	1.31
Identifying students in need of intervention	3.57	3.00	1.14	3.19	3.00	1.15
Estimating overall reading achievement	3.31	3.00	1.08	2.62	3.00	1.37
Acceptability scale						
Appropriateness for evaluating academic problems	4.76	5.00	0.93	3.46	4.00	1.31
Appropriateness for evaluating reading skills	4.72	5.00	1.06	3.96	4.00	1.32
Appropriateness for evaluating reading progress	4.71	5.00	0.99	3.86	4.00	1.29
Teachers like the assessment procedures	4.51	5.00	1.06	4.08	4.00	1.13
Teachers like having the assessment data	5.15	4.00	0.89	4.48	5.00	1.05
Assessment data are beneficial for children	4.88	5.00	0.93	4.36	5.00	1.06
Would suggest that other teachers use it for making instructional decisions	4.74	5.00	1.03	3.67	4.00	1.33
Practicality in administration time	4.21	4.00	1.21	4.67	5.00	0.86
CBM-R is a measure of decoding skills	—	— —	2.91	3.00	1.12	
CBM-R should have a comprehension question at the end (reverse scored)	—	—	—	2.30	2.00	1.28

MAP = Measures of Academic Progress; CBM-R = curriculum-based measurement in reading.

Another aim of the current study was to evaluate the concurrent validity of MAP with CBM-R. With the exception of the present study, Merino and Beckman's (2010) was the only study to evaluate the technical adequacy of the MAP. Current findings provide further evidence of the technical adequacy of MAP screening in Grades 2 to 5 and provide initial evidence of MAP's concurrent validity in first grade. The current study also addressed a gap in the literature by examining what benefit there is, if any, to administering both MAP and CBM-R for universal screening. Results indicated that MAP scores explained variance in third graders' ITBS-TR scores above and beyond CBM-R. CBM-R, however, failed to explain variance above and beyond MAP. These findings are not surprising, given shared method variance and the fact that both the MAP and ITBS are comprehensive measures of reading achievement, assessing basic reading skills, vocabulary, and reading comprehension. In contrast, despite CBM-R being described as a general outcome measure (Fuchs & Deno, 1991), it only directly measures students' oral reading rate with accuracy. In sum, current results provide initial evidence that, at least for third graders, MAP better estimates reading achievement than CBM-R, suggesting that MAP may be better to administer for universal screening. However, more evidence (e.g., adequate classification accuracy) is needed to further support this finding.

Acceptability of Universal Screeners

In general, current findings indicate that teachers found both MAP and CBM-R highly acceptable, providing evidence that they meet one of the criteria of universal screening measures suggested by Glover and Albers (2007). However, teachers demonstrated greater knowledge and use of MAP than CBM-R, which is not surprising given that the teachers in this study reported receiving more training on MAP than with CBM-R, despite having used CBM-R for a longer period of time. This interesting finding highlights the importance of teachers receiving initial as well as ongoing training with feedback regarding effective use of assessment data to inform instruction and intervention (Stecker et al., 2005). Alternatively, it could be that because MAP was adopted more recently, it was easier to remember the details of the training.

Acceptability of CBM-R. Findings from the current study replicate and extend previous research as teachers reported high degrees of acceptability of CBM-R. Consistent with prior research

(Allinder & Oats, 1997; Foegen et al., 2001), teachers in the current study thought that CBM-R was practical for assessing students' reading achievement and that the data are beneficial; however, they doubted its validity as an estimate of students' reading comprehension. It is unlikely that one training session is enough to remedy CBM-R's lack of face validity (Foegen et al., 2001); as such, ongoing training may be necessary. Teachers' beliefs about CBM-R may be due to it being a 1-min assessment of oral reading rate with accuracy; thus, it may not be intuitive that a strong relation exists between CBM-R and reading achievement. Teachers in the current study indicated that CBM-R should have a comprehension question at the end; perhaps this is a solution to CBM-R's face validity problem. However, one publisher of CBM-R probes (i.e., Dynamic Indicators of Basic Early Literacy Skills [DIBELS]) currently employs a passage retell after each probe as a measure of comprehension, but the reliability and validity of CBM-R retell data are questionable (Bellinger & DiPerna, 2011; Christ, White, Ardoin, & Eckert, 2013).

The present study also provides up to date information regarding teachers' knowledge and use of CBM-R data. Findings are consistent with previous research (Hamilton & Shinn, 2003) suggesting teachers think CBM-R is simply a measure students' word reading skills. Teachers in the current study also lacked the knowledge that using CBM-R to evaluate and modify their instruction leads to increased achievement gains (see Stecker et al., 2005 for a review) and that CBM-R is a strong estimate of reading comprehension (Jenkins et al., 2003b). Teachers' lack of knowledge regarding CBM-R may be due in part to their reported lack of training. Nonetheless, teachers' use of CBM-R data to identify students in need of intervention in reading was consistent with research-based recommendations (Deno, 2003).

Acceptability of the MAP. The current study is the first to provide evidence of the acceptability of MAP as a universal screening assessment. Provided that MAP is widely used, its high degree of acceptability supports, at least in part, its use as a universal screening assessment (Glover & Albers, 2007). Interestingly, many teachers also reported using MAP data to make instructional decisions, which is not surprising given that MAP publishers advocate for the use of reported scores and provided resources (i.e., the Learning Continuum) to determine which skills students need to be taught next (NWEA, 2015). Unfortunately, to date there is no empirical evidence to support this practice.

Limitations

Findings from the current study add to extant CBM-R and MAP literature; however, some limitations must be considered. First, the student participants were a fairly homogeneous group enrolled in two schools within the same high-performing district and only 86 teachers from five elementary schools in the same district participated. Therefore, the current study should be replicated with a sample more representative of the United States with regard to race, ethnicity, and socioeconomic background, with teachers in different districts, and with lower performing students. Furthermore, given that students are nested within the two participating schools, findings may differ if a greater number of schools were employed and other statistical techniques (e.g., multilevel modeling) were used. A second limitation of the study is that the ITBS was administered only to students in third grade. Therefore, although the current study provides initial evidence of the concurrent validity of MAP with ITBS, these results can only be generalized to third-grade students. Given that MAP is designed to be administered to students in Grades K–12, additional research is necessary to establish the validity of MAP with nationally normed assessments of reading achievement. A third limitation is that the survey employed in this study was developed by the investigator. Nonetheless, it had strong internal consistency, as estimated by coefficient alpha. Fourth, teachers self-reported their use of CBM-R and MAP anecdotally. Results may have differed if the study employed direct measures of teachers' actual use of data instead of their perceptions. A final limitation of this study was that all data were collected during the winter universal screening window; thus, findings may not generalize to other screening time frames.

Directions for Future Research

Current findings add to a growing literature base supporting the validity of MAP for universal screening within schools and, therefore, have important implications for future research. If a school's purpose for universal screening is to identify the lowest performing students, it is essential to determine the extent to which a universal screener accurately identifies students who are "at-risk." Research indicates that CBM-R is technically adequate for identifying at-risk students (Mellard et al., 2009); however, this has yet to be evaluated with MAP. For a universal screener to be appropriate, it must be technically adequate and demonstrate classification accuracy (Glover & Albers, 2007). Thus far, only one study in addition to the current study provides evidence of the technical adequacy of the MAP and no studies have evaluated its ability to identify at-risk students accurately. Therefore, future research is necessary to examine other forms of technical adequacy (e.g., predictive validity) and the classification accuracy of MAP data for universal screening.

Implications for Practice

Given that findings indicate CBM-R measures reading skills beyond word decoding, educators of students in Grades 2 to 5 are

encouraged to use CBM-R with the confidence that the resultant data are a good proxy for measuring students' global reading achievement. This is particularly relevant given that CBM-R is a quick assessment that can be obtained for free and is easy to administer and score. Furthermore, assessing oral reading rate with accuracy is an important task for educators because fluent reading is necessary for adequate reading comprehension (LaBerge & Samuels, 1974) and fluency development is an important component of a comprehensive reading program (NICHD, 2000). Although MAP is also strongly related to reading achievement, it does not assess reading fluency.

Although more research supporting MAP's use for universal screening is needed, present results also indicate that continued use of MAP for universal screening is appropriate. Findings suggest that, if schools are deciding between MAP and CBM-R for universal screening, it might be best to administer MAP, given that CBM-R fails to explain variance in students' reading achievement beyond MAP. By choosing to administer MAP instead of MAP with CBM-R, schools can likely gain the information needed to make data-based decisions and avoid wasting the time and resources used in administering two universal screeners. However, given the expense of the MAP, it may not be feasible for some districts to purchase it for use in their schools. As such, for schools employing an RtI framework, CBM-R is a technically adequate and inexpensive option for conducting universal screening. Furthermore, schools choosing to administer MAP in lieu of CBM-R will likely still need to administer CBM-R prior to providing intervention to establish a baseline for monitoring progress. Regardless of which universal screener schools elect to employ, using a technically adequate screener is a step toward ensuring that struggling readers are identified early and subsequently provided with targeted intervention.

Provided that teachers in the current study received more training on MAP than CBM-R, the finding that teachers differed in their knowledge and use of these measures highlights the importance of increasing teachers' knowledge about the assessments used in their schools and their capacity to use the resultant data to make decisions. It is possible that increased efforts on professional development with ongoing coaching, training, and feedback will increase teachers' capacity to use screening data effectively, which will likely lead to improved outcomes for students. This is particularly important, when considering the increased focus on accountability and student achievement.

Acknowledgments — The research reported here was supported in part by the University of Georgia through the Graduate School Dean's Award in Social Sciences granted to the first author. The opinions expressed are those of the authors and do not represent views of the University of Georgia. This article was based on the doctoral dissertation of the first author.

References

Albano, A. D., & Christ, T. J. (2012). *Curriculum based measurement passages: Scaling, linking and equating* (Tech. Rep. No. 10). Minneapolis: Department of Educational Psychology, University of Minnesota.

Albano, A. D., & Rodriguez, M. C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology*, *50*, 43–59. doi: 10.1016/j.jsp.2011.07.002

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education*, *18*, 113–120. doi: 10.1177/074193259701800205

Ardoin, S. P., Eckert, T. L., Christ, T. J., Morena, L. S., January, S.-A. A., & Hine, J. F. (2013). Examining variance in reading comprehension among developing readers: Words in context (curriculum-based measurement in reading) versus words out of context (word lists). *School Psychology Review*, *42,* 243–261.

Ball, C. R., & Christ, T. J. (2012). Supporting valid decision making: Uses and misuses of assessment data within the context of RTI. *Psychology in the Schools*, *49*, 231–244. doi: 10.1002/pits.21592

Bellinger, J. M., & DiPerna, J. C. (2011). Is fluency-based story retell a good indicator of reading comprehension? *Psychology in the Schools*, *48*, 416–426. doi: 10.1002/pits.20563

Christ, T. J., Aranas, Y. A., Kember, J. M., Kiss, A. J., McCarthy-Trentman, A., Monaghen, B. D., … White, M. J. (2014). *Formative assessment system for teachers technical manual: earlyReading, CBMReading, aReading, aMath, and early-Math*. Minneapolis, MN: Formative Assessment System for Teachers.

Christ, T. J., Ardoin, S. P., & Eckert, T. L. (2010). *Formative assessment instrumentation and procedures for reading (FAIP-R): Oral reading passages*. Minneapolis: University of Minnesota.

Christ, T. J., White, M. J., Ardoin, S. P., & Eckert, T. L. (2013). Curriculum based measurement in reading: Consistency and validity across best, fastest, and question reading conditions. *School Psychology Review*, *42*, 415–436.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*, 184–192. doi: 10.1177/00224669030370030801

Eckert, T. L., & Shapiro, E. S. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review*, *24*, 497–507.

Foegen, A., Espin, C. A., Allinder, R. M., & Markell, M. A. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *The Journal of Special Education*, *34*, 226–236. doi: 10.1177/002246690103400405

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*, 488–501.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239–256. doi: 10.1207/s1532799xssr0503_3

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117–135. doi: 10.1016/j.jsp.2006.05.005

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review*, *32*, 228–240.

Hauerwas, L. B., Brown, R., & Scott, A. N. (2013). Specific learning disability and response to intervention: State-level guidance. *Exceptional Children*, *80*, 101–120.

Hintze, J. M., & Silberglitt, 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.

Hosp, J. L., & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention*, *33*, 69–77. doi: 10.1177/1534508407311428

Howe, K. B., & Shinn, M. M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual describing development and technical features*. https://aimswebtraining.pearson.com/downloads/PassagesTechnicalManual08252002.pdf

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003a). Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. *Learning Disabilities Research & Practice*, *18*, 237–245. doi: 10.1111/1540-5826.00078

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003b). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*, 719–729. doi: 10.1037/0022-0663.95.4.719

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York, NY: Springer.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323. doi: 10.1016/0010-0285(74)90015-2

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice*, *24*, 186–195. doi: 10.1111/j.1540-5826.2009.00292.x

Merino, K., & Beckman, T. O. (2010). Using reading curriculum-based measurements as predictors for the Measures of 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: Reports of the subgroups* (NIH Publication No. 00–4754). Washington, DC.

Northwest Evaluation Association. (2009). *Technical manual for measures of academic progress and measures of academic progress for primary grades*. Lake Oswego, OR.

Northwest Evaluation Association. (2014). *NWEA by the numbers*. Lake Oswego, OR.

Northwest Evaluation Association. (2015). *MAP Learning Continuum*. Retrieved from https://www.nwea.org/assessments/map/learningcontinuum

Olsen, J. B. (1990). Applying computerized adaptive testing in schools. *Measurement and Evaluation in Counseling and Development*, *23*, 31–38.

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. doi: 10.1016/j.jsp.2009.07.001

Shapiro, E. S., & Eckert, T. L. (1994). Acceptability of curriculum-based assessment by school psychologists. *Journal of School Psychology*, *32*, 167–183. doi: 10.1016/00224405(94)90009-4

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. doi: 10.1002/pits.20113

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*, 199–204.

University of Iowa. (2005). *Iowa Test of Basic Skills, Form A*. Rolling Meadows, IL: Riverside Publishing.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225–256. doi: 10.1016/j.jsp.2006.11.004

van der Linden, W. J., & Glas, C. A. (Eds.). (2000). *Computerized adaptive testing: Theory and practice*. New York, NY: Springer.

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

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203–214. doi: 10.1901/jaba.1978.11-203