Technical Features of Slopes for Curriculum-Based Measures of Secondary Writing

John Elwood Romig

Amanda A. Olsen

University of Texas at Arlington

John Elwood Romig, Curriculum and Instruction, University of Texas at Arlington; Amanda A. Olsen, Curriculum and Instruction, University of Texas at Arlington.

The authors have no competing interests or conflicts of interest to report.

Correspondence concerning this article should be addressed to John Romig, Curriculum and Instruction, University of Texas at Arlington, Arlington, TX 76013. Email: john.romig@uta.edu

Abstract

Compared to other content areas, there is a dearth of research examining curriculum-based measurement of writing (CBM-W). This study conducted a conceptual replication examining the reliability, stability, and sensitivity to growth of slopes produced from CBM-W. Eighty-nine (*N* = 89) eighth grade students responded to one CBM-W probe weekly for 11 weeks. Probes were scored using four different CBM-W scoring procedures, words written, words spelled correctly, correct word sequences, and correct minus incorrect word sequences. We found limited evidence for reliability and stability of these slopes. Further, the slopes were not sensitive to growth, as defined as a slope significantly greater than zero. We recommend caution when using CBM-W to progress monitor eighth grade students. Future research should examine the same technical features with a sample of students who are low-performing or at-risk for writing failure.

**Technical Features of Slopes for Curriculum-Based Measures of Secondary Writing**

In 2011, the most recently reported National Assessment of Educational Progress (NAEP) found that 74% of eighth grade and 73% of twelfth grade students in the United States scored at or below the basic level in writing, indicating only partial mastery of grade level standards (National Center for Education Statistics, 2012). This poor performance in writing is consistent with past NAEP assessments. In 2007, 70% of eighth graders and 75% of twelfth graders scored at or below basic in writing (National Center of Education Statistics, 2008). These national results make two things clear. First, nearly three-fourths of eighth and twelfth grade students consistently perform below grade level standards in writing. Second, the percent of students failing to meet grade level standards has remained relatively unchanged between eighth and twelfth graders, indicating a clear need for improvement in writing outcomes in secondary grades.

**Progress Monitoring**

To improve instruction, teachers and schools need high-quality measures to determine whether instruction is helping students make adequate progress towards a goal (Deno, 1985). If writing measures indicate a student is not making adequate progress towards a goal, an instructional change can and should be made. Instructional changes could include increasing the dosage (e.g., frequency or duration) of a writing intervention, increasing the intensity of a writing intervention (e.g., reducing class size, providing more opportunities for the student to respond, providing more feedback), or increasing the explicitness of instruction (Fuchs et al., 2017). This process of using assessment data to make instructional changes is known as progress monitoring (Fuchs & Fuchs, 2006). In a progress monitoring framework, progress monitoring assessments are given to students frequently (e.g., weekly). Students’ progress is graphed and compared to an annual goal line. When students appear off-target for reaching the annual goal, teachers should make an instructional change (Hosp et al., 2016).

***Curriculum-based Measurement***

Curriculum-based measurement (CBM) is a set of measurement and evaluation procedures designed for teachers “to make decisions about whether and when to modify a student's instructional program” (Deno, 1985, pg. 221). CBM can be used as a screening tool, identifying students at-risk of academic struggle or in need of additional support. Although many assessment tools can serve as a screening measure, “CBM’s unique contribution resides in its capacity to model learning (over time) and to inform instruction” (Fuchs, 2004, pg. 191).When used to guide instructional decisions in a progress monitoring framework, CBM has a long history of improving reading and mathematics outcomes (Jung et al., 2018; Stecker et al., 2005), with preliminary evidence suggesting the potential to improve writing outcomes (Jung et al., 2018).

Deno (1985) designed CBM to be reliable and valid; simple, efficient, and inexpensive to score and administer; and easily understood by all stakeholders. As these qualities indicate, a primary goal for CBM is to help classroom teachers when making instructional decisions (Deno, 1985). Deno designed CBM to be a global outcome measure. That is, CBM was intended to be an indicator of significant annual learning goals. For example, oral reading fluency (number of words read per minute) is perhaps the most common type of CBM. This fluency measure of reading has strong associations with reading comprehension (Shin & McMaster, 2019) – the most significant learning goal for reading.

CBM of writing (CBM-W) has three key features: prompt type, timed administration, and scoring procedures. In the literature, CBM-W researchers have examined three types of writing prompts with students in sixth through twelfth grades: story starters, expository prompts, or picture prompts (Romig et al., 2020). For timed administration, CBM-W researchers typically examined writing durations between 1.5 minutes and 10 minutes with the 3-minute duration being most common for students in sixth through twelfth grades (Romig et al., 2020). Scoring procedures are perhaps the most unique feature of CBM-W as they differ from other writing assessments such as standardized rubrics or ratings scales. CBM-W writing samples are scored for discrete, observable characteristics. Although there are dozens of CBM-W scoring procedures, the most commonly-researched scoring procedures are: (a) a count of the number of words written (WW); (b) the number of words spelled correctly (WSC); (c) the number of correct word sequences (CWS); (d) the number of correct minus incorrect word sequences (CIWS; Romig et al., 2017).

**CBM-W - Stages of Research.** To support the use of CBM as a progress monitoring tool, Fuchs (2004) described three stages of sequential research necessary before using CBM in this way. Stage 1 research examines technical properties of static data (e.g., reliability and validity), stage 2 research examines technical properties of slope data (e.g., reliability, stability, sensitivity to growth), and stage 3 research examines the instructional utility – or effectiveness – of using CBM to guide instruction. Technical properties at each stage must be established before moving into subsequent stages (Fuchs, 2004). To illustrate, it would be inappropriate to make instructional decisions (stage 3) based on measures that lack validity (stage 1). The paragraphs below describe CBM-W research in each of these stages.

***Stage 1 Research.*** Deno et al. (1980) began development of CBM-W by conducting three different studies examining the criterion validity of three different writing prompts (story starter, topic sentence, and picture stimulus) when scored for the number of WW and WSC for students in third through sixth grades. In general, when compared to scores on the *Test of Written Language,* they found moderate to high correlations for both scoring procedures, indicating adequate criterion validity (WW *r* = .63 - .82; WSC *r* = .67 - .88). Videen et al. (1982) examined criterion validity of the CWS scoring procedure and found moderate to high correlations with broad measures of writing quality (*r* = .69 with *Test of Written Language*; .85 with a holistic rating).

These initial studies appeared promising for the criterion validity of all scoring procedures. However, later studies were not as positive for the WW and WSC scoring procedures. Gansle et al. (2002) found a criterion validity coefficient of .10 and .15 for the WW scoring procedure and .18 and .24 for WSC. Other studies have found similarly low criterion validity coefficients for these scoring procedures (e.g., Coker & Ritchey, 2010; Jewell & Malecki, 2005), suggesting these scoring procedures may not be meaningfully correlated with broad measures of writing quality (Romig et al., 2017).

However, for students in sixth through eighth grades, Romig et al. (2017) found the CWS and related scoring procedure – CIWS – had criterion validity similar to what was expected for other writing assessments. Specifically, for students in sixth through eighth grades, the CWS scoring procedure had an average weighted criterion validity coefficient of .50 [.40-.59], and the CIWS scoring procedure had an average weighted criterion validity coefficient of .59 [.49, .67]. Prompt type (e.g., story starters, expository prompts, or picture prompts) appeared to have little effect on criterion validity for students in sixth through twelfth grades (Romig et al., 2020).

The studies described above all used a criterion validity approach to examine the technical properties of CBM-W. Codding et al. (2016) used a novel approach – latent growth modeling – to examine the relationship of CBM-W scoring procedures with a state achievement test. Although criterion validity studies have for the most part found the WW scoring procedure to have limited criterion validity for middle school students (Romig et al., 2017), Codding et al.’s model retained the WW and CIWS scoring procedures, indicating some usefulness for the WW scoring procedure as a screening measure.

***Stage 2 Research.*** Stage 2 in CBM is a broad area of research that can include several different research questions, reflecting the many purposes for using CBM. These research questions can include alternate week reliability, longitudinal multi-level modeling, pre- post-test comparisons, and many other technical properties. Although stage 2 CBM-W research in middle school grades is limited, there are some stage 2 studies that include secondary students and inform the current study. Much of this research base is focused on sensitivity to growth.

With increasing specificity, Marston et al. (1981), Malecki and Jewell (2003), McMaster and Campbell (2008), and Fearrington et al. (2013) all demonstrated that CBM-W can be sensitive to growth across an academic year in middle school grades. Marston et al. (1981) first explored CBM-W in middle school grades by including sixth grade students in their sample. Malecki and Jewell (2003) extended the work of Marston et al. (1981) by including students in first through eighth grades. Both studies analyzed all grade levels together. Fearrington et al. (2003) analyzed and reported results for students in sixth, seventh, and eighth grades individually. McMaster and Campbell (2008) included seventh grade students in their examination of sensitivity to growth. These studies indicated that CBM-W can be sensitive to growth of students in various middle school grades over a semester or year. Specifically, studies found support for the sensitivity to growth of the WW scoring procedure (Deno et al., 1980; Fearrington et al., 2013; Marston et al., 1981), the WSC scoring procedure (Deno et al., 1980; Marston et al., 1981), the CWS scoring procedure (Fearrington et al., 2013; Malecki & Jewell, 2003 and the CIWS scoring procedure (Malecki & Jewell, 2003; McMaster & Campbell, 2008).

However, other technical properties aside from sensitivity to growth are important to consider when determining feasibility of CBM-W. For example, Parker et al. (2011) found that CBM-W was sensitive to growth for first grade writers, but their CBM-W prompts differed in growth patterns (linear vs. quadratic). In other words, they found that first graders responding to a picture-word prompt (i.e., writing a novel sentence when presented a word and accompanying picture) produced rapid growth early in the data collection with growth slowing as data collection progressed. Growth was not uniform across the data collection period. In contrast, their sentence copying prompt (i.e., students copying assigned text) produced a linear growth pattern. Technical findings like this are important to understand how CBM-W may be used to guide instruction in a progress monitoring context.

Likewise, aside from being sensitive to growth over the course of a year or a semester, CBM-W need to be sensitive to growth over a much short time period if they are used to guide instructional decisions. Additionally, although sensitivity to growth is an important technical consideration for stage 2 research, sensitivity alone is insufficient support for CBM scores. Before examining sensitivity to growth, slopes produced from CBM scores must be reliable and stable. If growth slopes are not reliable and stable, their sensitivity cannot be trusted.

Researchers have not examined the reliability and stability of CBM-W slopes at the secondary level; however, early elementary researchers have provided a model for secondary researchers to follow. McMaster et al. (2011) examined the stability, reliability, and sensitivity to growth for CBM tools used to assess first grade students. They administered a variety of weekly writing tasks for twelve weeks and found their measures to be stable and reliable after 8 or 9 weeks of data collection. The measures were sensitive to growth in a much shorter time period (4 weeks or fewer); however, due to the lack of reliability and stability in the early weeks, they cautioned against making instructional decisions until at least 8-9 weeks of data had been collected. Based on their results, this time frame yielded CBM slopes that were reliable, stable, and sensitive to growth.

***Stage 3 Research.*** Stage 3 research of CBM-W is limited. A meta-analysis by Jung et al. (2018) only found one study that used CBM to guide writing instruction in early elementary grades with none found at the secondary level. That study (i.e., Jung et al., 2017) found a large positive effect on first through third grade students’ CBM scores and the *Woodcock-Johnson III*. Although Jung et al. (2017) indicated CBM-W has the potential to improve students’ writing outcomes, more stage 2 questions need to be answered with secondary grade levels to inform how CBM-W should be used at the stage 3 level.

**Purpose and Research Questions**

Writing is an area of struggle for a significant majority of 8th grade writers (National Center on Education Statistics, 2012). Progress monitoring with CBM-W is one potential avenue for identifying struggling writers and evaluating instruction for these students. If technically sufficient CBM-W tools are identified, teachers could implement a data-based decision-making framework for writing instruction to improve students’ writing outcomes like what has been done successfully in reading and mathematics [see Jung et al., (2018) for a review).

Previous research has demonstrated the ability of CBM-W to be sensitive to growth over an academic year for middle school students (Codding et al., 2016; Fearington et al., 2013; Jewell & Malecki, 2003; McMaster & Campbell, 2008). However, previous research has not examined the reliability and stability of CBM-W slopes. Likewise, researchers have not examined the number of weeks necessary for CBM-W to be sensitive to growth in secondary grades. To address these gaps in the CBM-W literature, we conducted a conceptual replication (Coyne et al., 2016) of McMaster et al. (2011). A conceptual replication changes one or more features from an original study while answering the same or similar research questions (Coyne et al., 2016). We used the same research questions as McMaster et al. (2011), but some features of the study were modified. Modifications are explicitly highlighted in the methods section. The following research questions guided our study:

1. Do CBM-W prompts produce reliable, stable slopes of eighth grade student writing achievement*?*

 We hypothesized that like McMaster et al. (2011) CBM-W prompts would produce reliable, stable slopes after collecting eight weeks of data.

2. Are CBM-W measures sensitive to growth for eighth grade students?

 We hypothesized that CBM-W measures would be sensitive to growth over a shorter period of time (i.e. 3-4 weeks) as they were in McMaster et al. (2011).

**Method**

**Setting and Participants**

 This study took place in one middle school in a midsize suburban school district in the mid-Atlantic region of the United States. The school served 574 students in sixth through eighth grades. Sixty-six (66.2%) percent of students were White; 10.0% were Black; 9.8% were Asian; 9.1% were Hispanic, and 4.7% identified with two or more races. Eighteen percent (17.6%) of students received free or reduced lunch, 12.4% of students received special education services, and 8.1% were English language learners. Specific demographic data for the participating students were not provided by the school.

 The two teachers who taught eighth grade English Language Arts classes at this school were recruited to participate in this study. Although one teacher declined due to scheduling conflicts, the other teacher agreed to participate with all four sections of her English Language Arts classes (*N* = 89). Recruiting eighth grade participants was the first and most significant departure from the methods of McMaster et al. (2011).

**Measures**

Students responded to a weekly CBM story starter writing prompt that was randomly drawn from www.InterventionCentral.org (i.e., Wright, n.d.). Intervention Central is a teacher resource website that provides 20 free story starter prompts and a writing prompt generator. This website was chosen because it is easily accessible to teachers and provides other progress monitoring tools, like graphing resources that can be used to facilitate instructional decisions.

Each story starter had the opening line of a story followed by an ellipses mark. See Figure 1 for the list of story starters. Story starters are listed in the order they were presented to students. Examining writing prompts exclusively was the second departure from McMaster et al.’s (2011) procedures. McMaster et al. examined story starters, sentence copying prompts (students copy text provided for them), and picture-word prompts (students are given a word with illustrating picture and directed to write a sentence using the word). Of these three prompts, story starters seemed most appropriate for eighth grade students. Also, Romig et al. (2020) found story starters were the most frequently researched CBM-W prompt for students in sixth through twelfth grades.

***Scoring Procedures***

Each writing prompt was scored using rules from Hosp et al. (2016) for the four CBM scoring procedures: words written, words spelled correctly, correct word sequences, and correct minus incorrect word sequences. See Figure 2 for an illustration of scoring procedures.

 ***Words Written.*** WW was defined as any two adjacent letters offset by a space on either side plus single letter words *I* and *a*. Words did not have to be spelled correctly, make sense in the sentence, be capitalized correctly, or be punctuated correctly. In a meta-analysis of CBM-W scoring procedures, Romig et al. (2017) found WW had an average (weighted by sample size) criterion validity coefficient of .32 for students in sixth through eighth grades.

 ***Words Spelled Correctly.*** WSC was defined as any correctly spelled word in the English language. The word did not have to be the correct form of the word or be used correctly in the sentence (much like how a spell-checker would consider correct spelling). Incorrectly spelled words were marked by circling them. WSC had an average weighted criterion validity coefficient of .39 for students in sixth through eighth grades (Romig et al., 2017).

 ***Correct Word Sequences.*** CWS was defined as a count of the number of correct sequences – or, links – between words. For a sequence to be correct, words on either side of the sequence had to be spelled, punctuated, and capitalized correctly. Additionally, both words in the sequence had to be syntactically and semantically appropriate for the sentence. Correct sequences were marked with a caret (^). Incorrect word sequences were marked with a downward caret (**˅**). CWS had an average weighted criterion validity coefficient of .50 for students in sixth through eighth grades (Romig et al., 2017).

 ***Correct Minus Incorrect Word Sequences.*** CIWS was defined as the number of incorrect sequences subtracted from the number of correct sequences. Correct and incorrect sequences were defined and marked the same as the CWS procedure defined them. Romig et al. (2017) found the CIWS scoring procedure had an average weighted criterion validity coefficient of .65 for students in sixth through eighth grades.

**Procedures**

This study was approved by the appropriate university’s IRB. Due to the minimal risk of harm to participants, the IRB did not require documented consent. A letter was sent home with each student to notify their parents about the study and included the option of opting out. All students participated in the study.

Using a longitudinal study research design, students responded to one probe per week for 11 weeks. McMaster et al. (2011) administered probes weekly for 12 weeks. We were only able to administer probes for 11 weeks due to constraints with the school’s academic calendar (e.g., academic testing schedules). CBM-W prompt administration occurred at the beginning of the class period. The first CBM-W administration included an introduction and explanation of the purpose of the writing assessments. All subsequent CBM-W administrations were quite brief and included reading the probe directions, providing 1 minute for students to prepare a response, and 7 minutes for students to write a response.

The writing duration was the final departure from McMaster et al.’s (2011) procedures. They allowed students to write for five minutes; we allowed seven minutes for writing. We chose the longer duration because Romig et al. (2020) found the 7-minute duration had higher criterion validity for students in sixth through twelfth grades than the 5-minute duration. Also, we thought the longer writing time might allow more room for students to demonstrate growth over the 11 weeks.

***Administration of Writing Prompts***

The first author, who had extensive training and experience with administering CBM probes, administered all CBM probes for the duration of the study. Prior to the study, the first author developed two scripts for administering the CBM probes and used these scripts when administering the probes. The first script was used only for the first administration and included an explanation of the purpose for the writing prompts (i.e., to help the teacher know how to improve instruction) and the schedule to anticipate for the rest of the semester (i.e., writing on the same day each week at the beginning of the period).

The second script was used for all subsequent prompt administrations. This script read, “Today, I am wanting you to begin writing a story. You may not have time to finish it, and that is ok. I will read the prompt to you. Then, you will have a minute to think about your response. When I say, ‘please start writing,’ you may begin. You will have 7 minutes to write.”

After reading the directions, the first author started a timer. At the end of the allotted time, students were directed to stop writing and turn in their responses. The first author collected students’ responses.

***Fidelity*.** All CBM administrations were audio recorded to ensure fidelity of administration. See Figure 3 for the fidelity checklist items. One undergraduate research assistant listened to all audio recordings and scored each administration using the fidelity checklist. Fidelity was 100%.

***Scoring and Reliability.*** A team of undergraduate research assistants scored all CBM-W probes. All research assistants had a one-hour training session where the CBM-W scoring procedures were explained. Research assistants received a handout with scoring guidelines to facilitate accurate scoring. After the training, research assistants practiced on sample CBM probes until they achieved 100% agreement with the first author.

Twenty percent of the probes were scored by two research assistants. Intraclass correlations were calculated to determine scoring reliability. The intraclass correlation coefficient for WW was 1.00; WSC was 1.00; CWS was 1.00; and CIWS was 1.00. These intraclass correlations indicated a high degree of reliability between the raters.

**Data Analyses**

 To address the research question regarding reliability of slopes, OLS regression was used to calculate individual student slopes for each CBM-W. The CBM-W scores were then averaged by student and across the eleven weeks of the study. The reliability of the regression slopes was analyzed using alternate-slope reliability where the average CBM-W scores of the even weeks (slope A) were correlated with the average CBM-W scores of the odd weeks (slope B). Then, OLS regression was used to compare CBM-W slopes from the first three weeks to the full 11-week slope; we then added weeks individually to the comparison slope (e.g., weeks 1-4, weeks 1-5, etc.). To test whether there were differences between high achieving and low achieving students, mean differences in the average slopes across the CBM-W measures were investigated using independent samples t-tests, where students who scored in the highest quartile during week 1 were compared against students who scored in the lowest quartile. Quartiles were created by cutting student scores for each of the CBM measures into four equal groups, where quartile 1 represented high scoring students and quartile 4 represented low scoring students.

 To address the research question regarding stability of slopes, we examined the standard error of each slope beginning with weeks 1-3 and then gradually added an additional week. To address the research question regarding the sensitivity to growth, we examined the slopes. We defined sensitivity to growth as McMaster et al. (2011) did – slopes significantly greater than zero.

**Results**

**Descriptive Statistics**

Table 1 provides the means and standard deviations for each of the four CBM scoring procedures (i.e., WW, WSC, CWS, and CIWS) across all eleven weeks of the study. Figure 4 presents a line graph with the average CBM scores by week. Based on the directionality of the CBM line graph, it is evident that the average student CBM scores for the different writing measures did not increase over time.

**Slope Analysis**

***Reliability***

Results from the OLS regressions used to calculate the average slopes for slope A (odd weeks 1, 3, 5, 7, 9, 11) and slope B (even weeks 2, 4, 6, 8, 10) across CBMs found that only CWS and CIWS had statistically significant correlation coefficients, *p*<.04, although the correlations were weak at .21 and .31, respectively. See Table 2 for means, standard errors, standard deviations, and correlation coefficients for slope A and slope B.

 We also examined the correlations between slopes from weeks 1-3 and weeks 1-11 (full data) by subsequently adding one week to the comparison slopes (e.g., weeks 1-4, weeks 1-5, and so on). This analysis indicated that the reliability of slopes increased as each week was added to the comparison slopes. Figure 5 presents the graphed correlation coefficients for the incremental durations of data collection and the overall slopes.

In addition, we conducted one analysis that was not included in McMaster et al. (2011) to compare students who scored well to students who scored poorly. Results from an independent samples t-test found that across the eleven weeks of the study, there were no differences between the slopes of those students who scored in the first quartile compared to the fourth quartile, *p*>.05. This suggests that higher achieving students based on week 1 CBM scores did not have a higher slope average across the eleven weeks compared to students who scored poorly during week 1 (see Table 4).

***Stability***

To calculate overall stability of slopes, we calculated *SE*(b) for each scoring procedure (see Table 5). This analysis provided an estimate of individual students’ slope deviations from the mean slope. Individual variability around the mean slopes were indicated by the magnitude of *SE*(b) in relation to the slope. Figure 6 indicated that *SE*(b) decreased as the number of weekly data points increased.

To determine individual variation in slopes, we calculated the percentage of slopes that fell within the 95% confidence interval of the slopes produced from all eleven weeks (see Figure 7). These results suggested a very high level of individual variability within the dataset.

***Sensitivity to Growth***

Table 2 displays the slope, standard error, and t-value for each scoring procedure. Slopes for all four scoring procedures were negative and were not significantly different than zero, indicating none of the scoring procedures detected growth over the eleven weeks. Because the

slopes were not statistically significant, we did not complete further analyses to determine how early the measures detected growth as done by McMaster et al. (2011).

**Discussion**

The purpose of this study was to conduct a conceptual replication of McMaster et al. (2011) to examine the reliability, stability, and sensitivity to growth of CBM-W slopes for eighth grade students. Eighty-nine eighth grade students responded to one story starter prompt weekly and each student’s writing was scored for WW, WSC, CWS, and CIWS. The data were analyzed using a series of OLS regressions to measure reliability, stability, and sensitivity to growth.

**Reliability of Slopes**

In examining the reliability of slopes as determined by the correlation of slopes from even and odd weeks, this study found all four scoring procedures to have limited reliability (.01-.31). CWS and CIWS were the only scoring procedures to have statistically significant correlations between the odd and even week slopes, *p<*.05. However, although statistically significant, Pearson’s *r* value for the slopes indicated a low degree of reliability. Only one scoring procedure (i.e., CIWS; *r =* .31) was within the range of correlation coefficients reported by McMaster et al. (2011), but even this *r* value indicated a low level of reliability. One possible explanation for the limited reliability of the alternate slopes is the number of data points in each slope. Previous research in CBM of reading indicated longer measurement periods increase reliability (e.g., Christ, 2006; Hintze & Christ, 2004). McMaster et al. (2011) also recognized the number of data points in each alternate slope as a possible explanation for weak slope reliability.

 To address this lack of data points, we replicated the analysis of McMaster et al. (2011) by comparing the first three weeks of data points to the full 11-week slope. We incrementally added a data point to the comparison slope and correlated the new slope to the 11-week slope. This analysis indicated that reliability of slopes increased as more data was added. At week 8, slopes for all four scoring procedures had a correlation of ~.70 with the full 11-week slope. McMaster et al. (2011) found that at least 9 weeks of data collection were necessary to produce sufficiently reliable slopes (*r* ~ .70*).* They cautioned against over-interpreting the slope reliability of weeks 10 and 11 (their study collected 12 weeks of data) because these slopes were based on essentially the same data as the full 12-week slope. Similarly, our slopes at week 8 reached an acceptable level of reliability, although caution should be used when interpreting the correlations between slopes from weeks 9 and 10 as they use data similar to the full 11-week slope.

**Stability of Slopes**

 McMaster et al. (2011) found that standard errors decreased across the 11 weeks of data collection, possibly indicating an increase in stability of slopes. However, when examining individual slope variability, we found a very high percentage of individual slopes that fell outside the 95% confidence interval, indicating a high degree of variability in student performance from week to week. Comparable to McMaster et al., the percentage of individual slopes within the 95% confidence interval increased over time for all scoring procedures; however, unlike McMaster et al., the measures in this study never reached an acceptable threshold of stability, as measured by a percent of individual slopes within the 95% confidence interval of the overall slopes. Even when including all 11 weeks in the analysis, only ~20-25% of individual slopes fell within the 95% confidence interval for the overall slope. This result suggests that making instructional decisions (a main purpose of collecting progress monitoring data) based on these CBM scoring procedures may be difficult due to the high level of individual variability.

**Sensitivity to Growth**

This study did not find the CBM scoring procedures were sensitive to growth – as determined by a slope significantly different than 0 – over an 11-week period. Slopes for all four scoring procedures were negative and were not significantly different than 0. One limitation of this study (discussed below) was the lack of demographic information available for participating students. Specifically, we did not have access to the disability, English language learner status, grade point average, or other information to indicate whether any students were at-risk for writing struggle in the study. To address this limitation, we grouped students into quartiles based on the first week’s CBM score and compared slopes for quartile 1 (highest performing students) and quartile 4 (lowest performing students). Although this quartile analysis was a departure from the analyses done by McMaster et al. (2011), results from this analysis indicated there were no differences in slopes for students from the highest and lowest quartiles for any of the scoring procedures, suggesting that the CBM-W measures were not sensitive to growth for even the lowest performing students in this study (see Table 3 for results from quartile analysis). CBM tools typically view score increases as growth. Although not statistically significant, all slopes in this study were negative – counter to the expected indicator of growth.

There are several factors that can possibly explain the negative slope trend. First, a researcher, not the classroom teacher, administered all CBM probes. Therefore, students knew that the probes had no educational implications for them, potentially decreasing their motivation to write as the 11-week period progressed. Second, writing quantity may not necessarily reflect writing quality. In fact, often the opposite is true – brief, concise writing may reflect higher quality writing, possibly suggesting the negative slopes found in this study may actually indicate higher quality writing by more mature writers. However, without other data sources indicating writing quality, future research is needed to investigate whether negative slopes may be an indicator for writing quality for some populations.

**Limitations**

This study should be viewed in light of some limitations. First, participant demographic data could not be obtained for this study. Understanding student characteristics – specifically, disability and English learner status – is an important consideration for CBM-W research. Furthermore, CBM-W is typically conducted with low-performing students. Lacking a significant number of low-achieving students could explain why the CBM scores failed to demonstrate growth over the eleven weeks. Codding et al. (2015) found low-performing seventh graders demonstrated more CBM-W growth than high performing peers. Second, the participating students were aware that the CBM-W were part of a research study and may have inferred that the measures were inconsequential – that the classroom teacher was not using the data to guide instruction. This awareness could explain why students’ performance declined over the eleven weeks. Third, an important assumption of stage 2 research is that the writing prompts be equivalent in terms of difficulty, interest, and necessary background knowledge. This research has not been conducted with eighth grade students on the prompts used in this study.

**Implications for Practice**

CBM was developed to provide teachers with practical tools for instructional decision-making (Deno, 1985). Teachers and other school personnel can use CBM for several types of instructional decisions, including screening students to determine academic risk status, determining the effectiveness of interventions, and making instructional changes based on student progress (Deno & Fuchs, 1987). However, it is important that these decisions are based on assessments with acceptable technical properties. Previous research has not examined the reliability and stability of CBM-W growth slopes for secondary students. The results from this study did not support the use of CBM-W to monitor the progress of eighth grade students. However, this study also does not provide evidence against the use of CBM-W. Given the limitations of this study, we also cannot make strong claims against the use of CBM-W for progress monitoring of eighth grade students. The authors recognize that practitioners have very few alternatives for progress monitoring tools in written expression, nevertheless, professionals who choose to use CBM-W to progress monitor secondary students should be aware of the potential liabilities of CBM-W raised in this study. Teachers may consider using genre-specific rubrics to supplement CBM-W data.

**Future Research**

First, future research should continue to examine the reliability, stability, and sensitivity to growth of CBM-W slopes for secondary grade level students. If CBM-W slopes continue to demonstrate inadequate technical properties, two potential lines of future research should be explored. Researchers should consider alternative CBM-W administration procedures. Extending the duration of CBM-W may give students more time to demonstrate growth. Also, CBM-W tasks that are more closely aligned to the curriculum for secondary students (e.g., Truckenmiller et al., 2019) may produce CBM-W slopes with better technical properties. Additionally, researchers should consider innovating alternative tools for progress monitoring secondary students’ writing secondary students. It is possible that fluency measures may become less relevant progress monitoring tools as students grow older and develop more skill as writers. However, this level of nuance is not something that would be detected in a writing fluency measure like CBM-W. Researchers may consider examining genre-specific traits rubrics (i.e., rubric specific to persuasive, narrative, or explanatory writing) as progress monitoring tools. Genre-specific rubrics may be more sensitive to students’ growth than CBM-W. However, assessing writing through rubrics can require a significant amount of teachers’ time and may be more subjective, leading to limited interrater reliability.

 Second, future research could consider analyzing CBM-W data through multi-level modeling or structural equation growth modeling. Codding et al. (2015) found that seventh graders followed a different pattern of growth on CBM-W than on CBM for reading or mathematics. However, ordinary least squares analysis would not identify these nuances. We chose to use ordinary least squares because we were replicating the analysis of McMaster et al., (2011) and because ordinary least squares are most closely aligned to the practical uses of CBM-W as a progress monitoring tool.

**Conclusion**

 This research study is the first to examine weekly CBM-W administration for eighth grade students, analyzing technical features of slopes. The results raised questions about CBM-W’s ability to monitor progress weekly and guide instruction. Previous research demonstrated CBM-W to be sensitive to growth of middle school students when administered at three points in the year. Until research has demonstrated CBM-W measures to be sensitive over shorter periods of time, teachers should avoid making significant instructional changes based on short-term CBM-W performance.

References

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*(1)*,* 128–133. <https://doi.org/10.1080/02796015>

Codding, R. S., Petscher, Y., & Truckenmiller, A. (2015). CBM reading, mathematics, and written expression at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. *Journal of Educational Psychology, 107*(2), 437-450. [https://dx.doi.org/10.1037%2Fa0037520](https://dx.doi.org/10.1037/a0037520)

Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework for systematic, conceptual replications. *Remedial and Special Education, 37*(4), 244-253. [https://doi.org/10.1177%2F0741932516648463](https://doi.org/10.1177/0741932516648463)

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232. <https://doi.org/10.1177/001440298505200303>

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*(1), 1-16. <https://doi.org/10.17161/foec.v19i8.7497>

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). *Relationships among simple measures of written expression and performance on standardized achievement tests* (Report No. 22). University of Minnesota: Institute for Research on Learning Disabilities.

Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2013). Gender differences in written expression curriculum-based measurement in third- through eighth-grade students. *Psychology in the Schools, 51*(1), 85-96. <https://doi.org/10.1002/pits.21733>

Fuchs, L. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192. <https://doi.org/10.1080/02796015.2004.12086241>

Fuchs, L. S., Fuchs, D., & Malone, A. S. (2017). The taxonomy of intensive intervention. *Teaching Exceptional Children, 50*(1), 35-43. [https://doi.org/10.1177%2F0040059917703962](https://doi.org/10.1177/0040059917703962)

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*(4), 477-497. <https://doi.org/10.1080/02796015.2002.12086169>

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*(2), 204–217. <https://doi.org/10.1080/02796015.2004.12086243>

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement* (2nd Edition)*.* The Guilford Press.

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34*(1), 27-44. <https://doi.org/10.1080/02796015.2005.12086273>

Jung, P. G., McMaster, K. L., & delMas, R. C. (2017). Effects of early writing intervention delivered within a data-based instruction framework. *Exceptional Children, 83*(3)*,* 281-297. [https://doi.org/10.1177%2F0014402916667586](https://doi.org/10.1177/0014402916667586)

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice, 33*(3)*,* 144-155. <https://doi.org/10.1111/ldrp.12172>

Lembke, E., McMaster, K., Smith, R., Allen, A., Brandes, D., & Wagner, K. (2018). Professional development for data-based instruction in early writing: Tools, Learning, and Collaborative Support. *Teacher Education and Special Education, 41*(2), 106-120. [https://doi.org/10.1177%2F0888406417730112](https://doi.org/10.1177/0888406417730112)

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40*(4), 379-390. <https://doi.org/10.1002/pits.10096>

Marston, D., Lowry, L., Deno, S., & Mirkin, P. (1981). *An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study* (Report No. 49). University of Minnesota: Institute for Research on Learning Disabilities.

McMaster, K. L., & Campbell, H. C. (2008). New and existing curriculum-based measures: Technical features within and across grades. *School Psychology Review, 37*(4), 550-566. <https://doi.org/10.1080/02796015.2008.12087867>

McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children, 77*(2)*,* 185-206. [https://doi.org/10.1177%2F001440291107700203](https://doi.org/10.1177/001440291107700203)

National Center for Education Statistics. (2008). *The Nation’s Report Card: Writing 2007* (NCES 2008-468). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Education Statistics (2012). *The Nation’s Report Card: Writing 2011* (NCES 2012-470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Parker, D. C., McMaster, K. L., Medhanie, A., & Silberglitt, B. (2011). Modeling early writing growth with curriculum-based measures. *School Psychology Quarterly, 26*(4), 290-304. <https://psycnet.apa.org/doi/10.1037/a0026833>

Romig, J. E., Miller, A. A., Therrien, W. J., & Lloyd, J. W. (2020). Meta-analysis of prompt and duration for curriculum-based measurement of written language. *Exceptionality.* Advance online publication. <https://doi.org/10.1080/09362835.2020.1743706>

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education, 51*(2), 72-82. [https://doi.org/10.1177%2F0022466916670637](https://doi.org/10.1177/0022466916670637)

Shin, J., & McMaster, K. (2019). Relations between CBM (oral reading and maze) and reading comprehension on state achievement tests: A meta-analysis. *Journal of School Psychology, 73,* 131-149. <https://doi.org/10.1016/j.jsp.2019.03.005>

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: A review of research. *Psychology in the Schools, 42*(8)*,* 795-819. <https://doi.org/10.1002/pits.20113>

Truckenmiller, A. J., McKindles, J. V., Petscher, Y., Eckert, T. L., & Tock, J. (2019). Expanding curriculum-based measurement in written expression for middle school. *Journal of Special Education.* Advance online publication. [https://doi.org/10.1177%2F0022466919887150](https://doi.org/10.1177/0022466919887150)

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Report No. 84). University of Minnesota: Institute for Research on Learning Disabilities.

Wright, J. (n.d.). *Writing probe generator*. Retrieved August 31, 2016 from https://www.interventioncentral.org/teacher-resources/curriculum-based-measurement-probes-writing

|  |  |
| --- | --- |
| Table 1*Weekly Means and Standard Deviations for the CBM Story Prompt* |  |
|  | Week |   |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | Average |
| WW | Mean | 100.70 | 101.89 | 106.62 | 113.40 | 98.05 | 96.05 | 97.11 | 93.76 | 106.38 | 97.77 | 100.80 | 101.14 |
|  | SD | 34.72 | 32.41 | 32.19 | 38.39 | 39.05 | 35.47 | 39.04 | 39.47 | 48.35 | 38.11 | 37.91 | 37.74 |
| WSC | Mean | 99.65 | 97.88 | 104.32 | 111.30 | 96.20 | 92.78 | 94.31 | 92.76 | 104.68 | 94.48 | 99.42 | 98.89 |
|  | SD | 34.66 | 32.02 | 31.75 | 37.64 | 38.19 | 34.70 | 38.39 | 39.40 | 48.08 | 37.05 | 37.64 | 37.23 |
| CWS | Mean | 108.84 | 96.53 | 101.82 | 109.52 | 91.74 | 90.76 | 94.95 | 96.59 | 101.52 | 91.12 | 98.96 | 98.40 |
|  | SD | 38.23 | 31.92 | 31.11 | 36.87 | 36.55 | 34.40 | 37.82 | 41.58 | 46.44 | 36.94 | 37.35 | 37.20 |
| CIWS | Mean | 105.37 | 90.00 | 95.80 | 104.74 | 85.07 | 83.69 | 89.20 | 92.34 | 96.38 | 83.28 | 96.14 | 92.91 |
|   | SD | 37.73 | 32.41 | 30.89 | 35.94 | 34.23 | 33.62 | 36.98 | 41.32 | 44.96 | 37.19 | 36.93 | 36.56 |
| Note: CBM-W Curriculum-based measurement writing; WW=words written, WSC=words spelled correctly; CWS=correct word sequences; CIWS=correct minus incorrect word sequences |

Tables

|  |
| --- |
| Table 2*Means, Standard Deviations, and Correlations Between Slope A and B* |
|   | Slope A |  | Slope B |  |
|   | Mean | SDE | SD |  | Mean | SDE | SD | Pearson's r |
| WW | 0.04 | 2.90 | 3.78 |  | -1.51 | 3.70 | 5.98 | 0.01 |  |
| WSC | 0.04 | 2.86 | 3.77 |  | -1.66 | 3.83 | 5.62 | 0.08 |  |
| CWS | -0.64 | 2.86 | 3.84 |  | -1.63 | 3.74 | 5.62 | 0.21 | \* |
| CIWS | -0.55 | 2.87 | 3.78 |  | -1.07 | 3.82 | 5.58 | 0.31 | \*\* |
| Note: Weeks 1, 3, 5, 7, 9, 11 (Slope A), Weeks 2, 4, 6, 8, 10 (Slope B), \**p<*.05; \*\**p*<.01 |

|  |
| --- |
| Table 3 |
| *Independent Samples t-tests of Slopes from Quartile 1 and 4 Based on Week 1 Data* |
| CBM Type | Quartile  | Mean Score(SD) | Mean Slope (SD) | *t-*value | df | *p* |
| WW | Quartile 1 | 72.47 (21.98) | 0.03 (2.20) | 0.91 | 35.28 | 0.37 |
|  | Quartile 4 | 133.68 (29.41) | -0.82 (3.84) |  |  |  |
| WSC | Quartile 1 | 69.50 (23.81) | -0.35 (2.73) | 0.42 | 41.86 | 0.67 |
|  | Quartile 4 | 129.82 (29.25) | -0.76 (3.71) |  |  |  |
| CWS | Quartile 1 | 70.31 (22.15) | 0.20 (2.54) | 1.64 | 39.23 | 0.11 |
|  | Quartile 4 | 129.60 (28.40) | -1.34 (3.67) |  |  |  |
| CIWS | Quartile 1 | 64.92 (22.82) | 0.30 (2.21) | 1.88 | 40.37 | 0.07 |
|   | Quartile 4 | 114.68 (33.85) | -1.25 (3.31) |   |   |   |
| Note. If Week 1 data was missing, the next available data point was used |  |

|  |
| --- |
| Table 4*Means, Standard Deviations, and SE(b) for Slopes from Incremental Durations* |
|  | Weeks |
|  | 1-3 | 1-4 | 1-5 | 1-6 | 1-7 | 1-8 | 1-9 | 1-10 | 1-11 |
| Measure | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) |
| WW | Slope | 6.56 | (19.31) | 5.80 | (13.10) | 0.72 | (7.12) | -0.85 | (6.24) | -0.93 | (5.03) | -1.32 | (4.57) | -0.65 | (4.52) | -0.81 | (3.98) | -0.59 | (3.25) |
|  | *SE*(b) | 11.87 | (10.35) | 9.15 | (8.12) | 7.03 | (4.69) | 5.21 | (3.13) | 4.40 | (2.37) | 3.67 | (2.15) | 3.26 | (1.84) | 2.80 | (1.46) | 2.44 | (1.22) |
| WSC | Slope | 6.63 | (22.99) | 6.17 | (18.20) | 0.87 | (7.23) | -0.88 | (6.12) | -0.99 | (5.09) | -1.25 | (4.37) | -0.56 | (4.36) | -0.78 | (3.81) | -0.55 | (3.03) |
|  | *SE*(b) | 11.78 | (10.51) | 9.18 | (8.04) | 7.12 | (5.68) | 5.24 | (3.41) | 4.37 | (2.47) | 3.60 | (2.27) | 3.24 | (1.86) | 2.79 | (1.53) | 2.43 | (1.26) |
| CWS | Slope | -0.49 | (19.73) | 2.06 | (13.81) | -2.00 | (7.04) | -2.77 | (5.97) | -2.15 | (4.96) | -1.65 | (4.68) | -1.04 | (4.55) | -1.30 | (3.98) | -0.90 | (3.30) |
|  | *SE*(b) | 12.44 | (11.38) | 9.73 | (8.52) | 7.16 | (4.70) | 5.21 | (3.00) | 4.35 | (2.29) | 3.67 | (2.04) | 3.21 | (1.72) | 2.77 | (1.37) | 2.42 | (1.18) |
| CIWS | Slope | -1.97 | (19.46) | 1.41 | (14.46) | -2.52 | (7.02) | -3.27 | (5.70) | -2.48 | (4.87) | -1.78 | (4.64) | -1.07 | (4.45) | -1.44 | (3.92) | -0.91 | (3.16) |
|   | *SE*(b) | 13.46 | (11.94) | 9.92 | (9.00) | 7.46 | (4.72) | 5.28 | (3.06) | 4.41 | (2.22) | 3.71 | (1.90) | 3.23 | (1.57) | 2.77 | (1.26) | 2.43 | (1.10) |
| Note. WW=words written, WSC=words spelled correctly; CWS=correct word sequences; CIWS=correct minus incorrect word sequences |  |

|  |  |  |
| --- | --- | --- |
| Table 5*Slope, Standard Error, and t-value for the CBM Story Prompt* |  |  |
|   | Slope | SDE | *t*-value |  |  |
| WW | -0.59 | 2.44 | -0.13 |  |  |
| WSC | -0.55 | 2.43 | -0.17 |  |  |
| CWS | -0.90 | 2.42 | -0.36 |  |  |
| CIWS | -0.91 | 2.43 | -0.41 |  |  |
| Note: Individual slopes averaged across all 11 weeks |  |  |