Summer 2019

A Review of Written Expression Curriculum-Based Measurement with a Focus on English Language Learners

Shelby Montgomery
shelby.montgomery165@topper.wku.edu

REVIEW OF WRITTEN EXPRESSION CURRICULUM-BASED MEASUREMENT WITH A FOCUS ON ENGLISH LANGUAGE LEARNERS

A Specialist Project
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Specialist in Education

By
Shelby Montgomery

August 2019

REVIEW OF WRITTEN EXPRESSION CURRICULUM-BASED MEASUREMENT
WITH A FOCUS ON ENGLISH LANGUAGE LEARNERS

Date Recommended 8-6-2019

Dr. Sarah Ochs, Director of Specialist Project

Carl Myers

Dr. Carl Myers

Shawna Gilbert

Ms. Shawna Gilbert

Cheryl D. Davis 8/7/19

Dean, The Graduate School Date

I dedicate this thesis to my parents, Bob and Ginger Montgomery, who have always encouraged me to work hard and achieve my goals. I would also like to thank Kristen Erps for letting me vent to her all hours of the day about this thesis, and her encouraging comments trying to push me through it.

ACKNOWLEDGMENTS

I would like to sincerely thank Dr. Sarah Ochs for her patience in working with me throughout the past couple of years on this thesis. Without her, I truly would not have been able to complete it. I also would like to thoroughly thank all of the people on my committee including Dr. Carl Myers, of Western Kentucky University and my practicum supervisor, Shawna Gilbert, for agreeing to put in the time to work with me and review my work, and to get me to graduate.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

REVIEW OF WRITTEN EXPRESSION CURRICULUM-BASED MEASUREMENT WITH A FOCUS ON ENGLISH LANGUAGE LEARNERS

Shelby Montgomery August 2019 54 Pages

Directed by: Dr. Sarah Ochs, Dr. Carl Myers, and Ms. Shawna Gilbert

Department of Psychology Western Kentucky University

The purpose of this study was to review the literature on English Language Learners and written expression curriculum-based measurement. In recent years, there has been little research completed in the area of curriculum based measurement for writing for English Language Learners. A systematic review of the literature was conducted and all available literature on the use of written expression curriculum-based measurement since 2006 was identified and reviewed. Given the increasing diversity in our schools, particularly non-native English speakers, this review focused on studies including this group. There were differences in some of the technical features such as sample durations, writing tasks, and scoring procedures within the schools they were conducted in. Remaining gaps in the literature are discussed in addition to future directions and limitations.

Introduction

Screening measures are tools or instruments that can be utilized to help identify students who may need additional supports. They allow school personnel to make educational decisions, quickly and efficiently. With the onset of laws and policies that require annual assessment of students' skills, such as the No Child Left Behind Act of 2001 (NCLB), schools began to focus on early identification of students at-risk for not meeting academic standards (McMaster & Espin, 2007). To identify these students, schools developed stronger response to intervention teams within their schools and began utilizing screeners to help make some of those educational decisions. Not only do they aid in making educational decisions, but the screeners also help with monitoring student progress and performance with ease, as screeners are quick to administer, require few resources to implement, and are a feasible tool for school staff to use.

Curriculum-based measurement (CBM) is one type of popular screening tool that has benefitted many teachers and school personnel. CBMs can be used for mathematics, reading, and writing. Within the past 40 years, there have been countless studies looking at the effectiveness and reliability of CBM measures. However, the majority of those studies have focused on reading and mathematics, with little research looking at written expression CBM.

McMaster and Espin (2007) completed a literature review examining the technical features of written expression CBM (WE-CBM). They highlighted 28 different reports and published articles that centered on the technical adequacy of WE-CBMs and the differences in the writing tasks, durations of the samples used, and scoring procedures across primary and secondary school systems across the United States. A limitation of the McMaster and Espin (2007) study is the lack of research pertaining to English Language Learners. This population of students has grown in the past decade, and in 2016 the percentage of public school students who were English Language Learners was 1.5 percent higher than in 2000 (i.e, 8.1 percent in 2000 and 9.6 percent in 2017; NCES, 2019).

Since the McMaster and Espin (2007) review, more research on WE-CBM has been published. The purpose of this literature review was to provide an update to the McMaster and Espin (2007) study with research from the last decade. More specifically, this specialist project reviewed research pertaining to the use of WE-CBM since 2006 with a focus on studies examining English Language Learners.

Literature Review

Curriculum-based Measurement

Curriculum-based measurement (CBM) was developed by Stanley Deno at the University of Minnesota in the mid-1970s (Hosp, Hosp, & Howell, 2007). This early work by Deno at the Institute for Research on Learning Disabilities, focused on the creation of writing, reading, and spelling measures. The original purpose of CBM was to assist educators in the special education referral process since teacher report of academic behavior was not valid (Hosp et al., 2007). Since then, research and use of CBM has expanded to provide school personnel with an early screening tool to help identify those students who need additional support and to monitor progress toward a goal in reading, writing, or math. CBM provides an aid to teachers in determining the effectiveness of their classroom instruction to students. Decades of research has shown that CBM is economical, time efficient, inexpensive, easy to administer, reliable, and contains high test-retest reliability and interrater agreement (Espin, Shin, Deno, Skare, Robinson, & Benner, 2000; Hosp et al., 2007; Keller-Margulis, Mercer, Payan, & McGee, 2015).

CBM is used in the initial screening of risk and to track progress for students on how they are performing over time (i.e., progress monitoring). The administration typically includes a set of standard directions, a timer, standards for grading or scoring the performance, record forms, and materials such as paper, pencils, and probes. The instructions are direct, and replicate tasks the student would usually do in class (e.g., complete basic math problems, read a story). The student is timed so he or she can be scored based on the number of responses correct and incorrect per minute for that task. The scores are written down by the administrator, and recorded using either a paper graph or computer. CBM allows the teacher to “test what they teach, and teach what thy test” (Hosp et al., 2007, p. 3).

Reading is the most researched area of CBM, with more work now being devoted to math, and little to writing (Codding, Petscher, & Truckenmiller, 2015; Deno, 2003; Hosp et al., 2007). Research has shown that CBM reading scores are related to standardized achievement test scores, students’ age and grade, to teacher’s judgments of reading proficiency, and program placement. Despite the many advantages to using CBM, teachers found them to be time-consuming and lacking face validity (Yell, Deno, & Marston, 1992). Administrators were concerned that there seemed to be an overall resistance to change in practice in many school systems, that CBM appeared to be difficult for effective use by teachers, and whether teachers would be able to use data from CBM to modify instruction.

Written Expression Curriculum-based Measurement (WE-CBM). CBM research contains abundant studies on the effectiveness of their usage, but center primarily around reading measurements. In recent years, written expression CBM (WE-CBM) has gained more extensive data on the effectiveness of it being used as a screening tool. This may be because national writing test data shows that less than 30% of students in fourth and 12th grades as well as less than 40% of eighth-grade students, performed at the proficient level (Persky, Daane, & Jin, 2003).

WE-CBM is a brief measure of writing performance that can be administered individually or in groups. WE-CBM begins with a simple story starter and it involves a teacher orally giving a sentence prompt to a student or group of students, in which they the students then have 3-7 minutes to come up with a written response based on the story starter provided. The standardized instructions can be found in Appendix A. Students are presented with a story starter (e.g., Yesterday, a monkey climbed through the window at school and...), and then given one minute to plan or think, and then three minutes to write. Passages are generally scored using production dependent, production independent, and accurate production indices.

There are three common ways to score production dependent indices including Total Words Written (TWW), Correct Writing Sequences (CWS), and Words Spelled Correctly (WSC). TWW is the total number of words the student is able to produce, regardless of if the word written is a real word or a redundant word. A word is considered any letter or group of letters separated by a space, even if the word is misspelled or a nonsense word. CWS is defined as two adjacent writing units (including words and punctuation) that are acceptable within the context of what the student wrote (Pearson Education, 2014). CWS scoring allows for the consideration of many writing elements, including spelling, grammar, syntax, punctuation, etc. First, you must circle all of the misspelled words. Then, you draw a caret and place it between the words that (a) are mechanically (spelled correctly, appropriate capitalization), (b) semantically, and (c) syntactically correct. CWS is the sum of the carets used. Scoring CWS requires more inferences about what the student intended such as whether a sentence ended when a period was omitted. WSC is calculated by counting the words that would be considered correct if it can stand alone as a word in the English language. Sometimes, there is an additional scoring method used called Correct Punctuation (CP). This scoring includes only punctuation marks that are used correctly. When a written passage includes a punctuation mark, it is scored by whether it is in the correct location (e.g., a period, question mark, or exclamation mark appeared at the end of the sentence and after a subject/verb combination) and whether it is used appropriately (e.g., a question mark is used at the end of a sentence that begins with a question word, such as what or how). It is only counted as correct if the punctuation is at the end of a sentence. For the production-dependent scores, calculation of both WSC and CWS are used and then divided by the TWW. This calculation provides a percentage. Accurate production indices looks at accuracy in production of the responses, using Correct Minus Incorrect Writing Sequences (CIWS) which examines the accuracy and fluency of the response. The incorrect word responses are then subtracted from the correct word sequences (Wright, 1992).

The production-independent indicator of %WSC is calculated by using WSC divided by TWW. The %CWS is calculated using CWS divided by Total Writing Sequences. The accurate-production indicator of CIWS includes the assessment of fluency (production dependent) and accuracy (production independent). CIWS is scored by calculating the difference between CWS and the Incorrect Writing Sequences.

Research on WE-CBM. Early studies found WE-CBM yielded a criterion validity ranging from .41 to .92 for the indicators of Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Writing Sequences (CWS) (Marston, 1989). Recent literature has found WE-CBM to have strong validity and reliability (Codding et al., 2015). Research has found that the validity of these metrics of TWW, WSC, and CWS are within the moderate to strong range (Keller-Margulis et al., 2015). The production-dependent measure of TWW provides less benefit than WSC and CWS, however, it was found that complete sentences and correct capitalizations were moderately correlated with standardized test scores (Gansle, VanDerheyden, Noell, Resetar, & Williams, 2006).

Other studies have researched production-independent and accurate-production variables across grade levels. In the study by Espin et al. (2000), researchers assessed elementary school-age students and the use of production-dependent measures of CWS, TWW, and WSC on WE-CBM, and found that they have moderate to strong correlations based on writing outcomes. However, when looking at secondary school outcomes, production-independent outcomes were more highly correlated with writing outcomes.

Another study looked at the production-dependent, production-independent, and accurate-production using writing samples from second, fourth, and sixth grade students (Jewell & Malecki, 2005). They found that production-independent and accurate-production variables were more highly correlated to achievement test, grades, and writing scoring systems.

The diagnostic accuracy of CBM helps to identify those students who are at-risk academically. The diagnostic accuracy is examined by using cut scores. Using CBMs and examining the students who performed below an identified cut score can help to recognize students who might be at-risk for failing the state test. The students performing below the cut-score on WE-CBM, are at-risk for not reaching proficiency of the curriculum. There are times where students are able to pass a screener such as a CBM, and they still do not perform at a proficient level by another point in time on a different assessment, like the state test in the Spring. Based on diagnostic accuracy statistics, if a student is performing below the cut score (failing) on the screener, then we would assume they are likely to fail the state test as well.

The diagnostic accuracy statistics include sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). Sensitivity is also referred to as the probability of detection, which measures the likelihood that an individual who has some type of condition, is correctly identified as having that condition, by the predictor variable. An example of this would be being able to identify the percentage of individuals who have an academic issue, correctly, by using a predictor such as a CBM. Specificity examines the likelihood that someone without the academic issue is correctly identified by the predictor variable. PPP is the likelihood of an individual having an issue, and the predicting variable is able to predict and confirm that it is true. NPP is the likelihood that an individual who is not identified as having an issue, actually is able to perform well on the criterion measure.

Schools place great importance and expectations on school staff for statewide achievement tests. There are many incentives for school systems to receive high scores on the state-wide tests. In preparation for the state tests administered in the Spring, many teachers put a lot of instructional focus on the topics that might appear on the test. In many schools, after Spring break, the focus is teaching for the state test and reviewing material that might need clarification in hopes that students will be able to know more content on the state test. The writing portion on the test varies upon grades, but typically begins with a prompt and requires a written explanation for the student to produce.

Both the National Center for Education Statistics (NCES, 2012) and the National Commission on Writing (2006) have suggested that students perform below the desired levels in writing. The National Assessment of Educational Progress (NAEP) indicated that 72% of 4th grade students, 69% of 8th grade students, and 77% of 12th grade students performed below proficient level nationally (NCES, 2012), with more recent statistics suggesting that 76% of students in eighth and twelfth grade are not proficient in writing (NCES, 2012).

Fuchs and Fuchs (2004) provided a framework for categorizing the CBM research and determining the needed areas of study that is most relevant to WE-CBM. Fuchs and Fuchs analyzed three stages of CBMs. Stage I research examines the technical adequacy of scores at a single point in time, which is what a heavy body of research has explored. Stage II research, which is less researched, focuses on whether CBM scores can adequately measure growth over time (i.e., slope). Stage III, involves the utility of the CBM data for shaping instruction while making it useful for teachers and administrators, and helps to improve skills in the areas of writing (Fuchs and Fuchs, 2004). Recent research has suggested that production-independent measures of writing performance, are linked to more statewide test outcomes than the conventionally used production-dependent indicators such as TWW, particularly for those students in middle and high school (Mercer, Martínez, Faust, & Mitchell, 2012; Weissenburger & Espin, 2005).

The usefulness of WE-CBM for indexing growth across the school year is an important practical adequacy issue (Stage II; Fuchs and Fuchs, 2004) because normative rates of growth are needed to determine goals and estimate student performance in response to instruction and/or intervention (RTI). Early studies of CBM growth (e.g., Fuchs, Fuchs, Hamlett, & Walz, 1993) have served as the standard for understanding student performance over time and have also shown results that growth over time displayed a linear pattern. However, more recent studies have found that growth is not linear because growth is variable across the fall and spring semesters for skills such as reading and math. Writing skills are thought to have the same variable growth pattern, however there has been little research in the area of annual growth for writing. Several of these recent studies indicating this variability in growth across semesters include R-CBM (Ardoin & Christ, 2008; Christ, Silberglitt, Yeo, & Cormier, 2010) and M-CBM (Keller-Margulis, Mercer, & Shapiro, 2014) but there have been fewer in writing up until 2007, and there are still a fewer amount of research in the area of writing than in reading and mathematics.

The most comprehensive review to date of WE-CBM was conducted by McMaster and Espin (2007). They sought to review the technical features of WE-CBM and identified 28 technical reports and published articles, half of which were early research studies from the Institute for Research on Learning Disabilities (IRLD) during the development of CBMs. They summarized the findings of each article and highlighted differences in technical features (i.e., reliability and validity) across types of writing tasks, times, and scoring metrics. In their discussion, McMaster and Espin (2007) identified a need to understand how WE-CBM works with diverse populations, specifically English learners.

English Language Learners (ELLs) in Public Schools

The percentage of public-school students in the United States who were ELLs was higher in Fall 2015 (9.5 percent, or 4.8 million students) than in Fall 2000 (8.1 percent, or 3.8 million students; NCES, 2018). This increase added 1 million more ELL students in the 15-year time span and is still continuing to increase in upcoming years.

In the Fall 2015, about 4.8 million public school students were identified as ELL, additionally, it is estimated that around 713,000 ELL students were identified as students with disabilities, representing 14.7 percent of the total ELL population enrolled in U.S. public elementary and secondary schools according to the National Center for Education Statistics. Also in the Fall of 2015, the percentage of public school students who were ELLs reached 10 percent or more in eight states. These states included Alaska, California, Colorado, Kansas, Nevada, New Mexico, Texas, and Washington. California reported the highest percentage of ELLs at 21 percent, followed by Texas and Nevada, each at 16.8 percent. Nineteen states and the District of Columbia had percentages of ELL students that were 6.0 percent or higher but less than 10 percent, and 14 states had percentages that were three percent or higher but less than 6 percent. The percentage of students who were ELLs was less than three percent in nine states, with Mississippi (2.0 percent), Vermont (1.6 percent), and West Virginia (1.0 percent) having the lowest percentages.

In 2013, a record 61.8 million U.S. residents (native-born, legal immigrants, and illegal immigrants) spoke a language other than English at home. The number of foreign-language speakers increased 2.2 million between 2010 and 2013. It has grown by nearly 15 million (32 percent) since 2000 and by almost 30 million since 1990 (94 percent). In 2014–2015, around 665,000 ELL students were also identified as students with disabilities (only includes students served under IDEA act). ELL students with disabilities represented 13.8 percent of the total ELL population enrolled in U.S. public elementary and secondary schools.

The Center for Immigration studies used data from the Census Bureau's survey to reflect the population as of July 2013. This study is completed every year and is the largest survey taken by the federal government. This survey includes over two million households. The survey has three questions; first the survey asks whether each person in the household speaks a language other than English at home. The second question is for those who answer "yes" to the first question, which asks what language the person speaks at home. The third question asks how well the person speaks English. Only those who speak a language in the home other than English, are asked about their English skills.

The Center for Immigration used the survey's results and found that the largest increases from year 2010 to 2013 were for Spanish speakers (up 1.4 million, 4 percent growth), Chinese (up 220,000, 8 percent growth), Arabic (up 188,000, 22 percent growth), and Urdu (up 50,000, 13 percent growth). Of the school-age (5 to 17) nationally, more than one in five children speaks a foreign language at home. It is 44 percent in California and roughly one in three students in Texas, Nevada, and New York. But more surprisingly, it is now one in seven students in Georgia, North Carolina, Virginia, Nebraska and Delaware; and one out of eight students in Kansas, Utah, Minnesota, and Idaho. Of those who speak a foreign language at home, 25.1 million (41 percent) told the Census Bureau that they speak English less than very well.

Roughly two million school-age children ages 5–17 speak Spanish in the home (National Center for Education Statistics, 2019). ELLs account for 1 in 10 students in elementary schools, with students who are specifically Spanish speaking accounting for 79.2% of total ELLs (NCES, 2019). Roughly 50% of ELL students fail state assessments required for graduation, compared to the 24% overall students. This large increase of people speaking a foreign language, has caused a tremendous need for English

Language instruction. In order for students to benefit from instruction, instructors must tailor instruction to meet the student's needs. While the success of writing instruction depends on establishing student needs, establishing authentic student needs, depends on collecting and analyzing authentic data. Valid data depends on the source, as well as the methods used to gather the data.

English Language Learners Writing Acquisition. Native English writers acquire English grammatical and linguistic patterns naturally, due to their environment and cultural background. ELL writers must acquire these writing skills deliberately, usually as a result of direct instruction. Meaning that ELL writing samples will probably have more surface-level errors in grammar, spelling, mechanics, vocabulary and other linguistic features than the native English-speaking writers. ELL writers also have a tendency to transfer their native language patterns from first language, to second language when they are learning to write English, especially in the beginning stages (Nelson, 1991). Some of the challenges that ELL students face is the need for scaffolding while learning written expression tasks because the language they know already has different syntactic instruction (e.g., adjective before noun in English, but the noun before adjective in Spanish).

ELL writing can also differ by cultural influence. ELLs may have little to no writing in their culture, the ideas they generate might also be different than western norms, and their organizational structure of writing may appear somewhat abnormal. For those ELL students who have received instruction on writing in their native language, they may have to purposely un-learn their native language writing patterns in order to write in the normal academic American writing form. This re-learning process is very complicated and challenging for many ELL students and adults (Swales, 1990). For many ELL students, they might not have well-developed literacy skills and there can be a lack of formal training on writing, even in their native language. This causes ELL students to be at a great disadvantage because they are forced to unlearn concepts if they have been through proper instruction, while simultaneously learn a new organized way of writing in their non-native language, in a short amount of time.

Due to the cultural influence on writing, one study found that adult ELL students unique form or style of written expression may not easily conform to the expected format, explanation, and organization that American ELL writing classes require. Often, ELL writers' written expression samples may be misunderstood, undervalued, and/or poorly evaluated by teachers who are unfamiliar with ELL students' different styles of rhetorical expression (Hamp-Lyons, 1991).

One study by Lin (2015), researched 20 ELL students to gain insight and understanding of the challenges that ELL students experience in their writing development. Researchers collected data from questionnaires and in-depth interviews. The 20 ELL participants used, had volunteered and were all enrolled in a Spring 2011 composition class at an American Midwest University's Center for English as a Second Language (CESL) program. The CESL program offers a quite demanding English language program designed for students who would like to master English quickly and thoroughly. The participants had 25 hours a week of instruction for 7 weeks. Among the 20 participating ELL students, 11 were male (55%) and 9 females (45%). Participating students originated from seven countries, including China (n = 8, 40%), Taiwan (n = 4, 20%), Saudi Arabia (n = 3, 15%), Angola (n = 2, 10%), Cameroon (n = l, 5%), Thailand

(n = l, 5%), and Korea (n = l, 5%). In this study, the primary criteria that was used to define items into most and least difficult categories were mean scores (M = 3.54) and standard deviations (SD = .44). The participants agreed that word choice (M = 5; SD = .97) was the chief writing difficulty. The second greatest writing difficulty was being able to adjust to American thought patterns (M = 4.96; SD = 1.08). Many participants indicated that they had difficulties adapting from their native ways of writing in their language, to American thought patterns. Students still thought in then-first language and used the rhetorical patterns of their first language to write English essays.

The third greatest writing difficulty was writing fluency (M = 4.79; SD = .88). Students stated that they stop many times throughout writing tasks, to think about what to write when they write English essays. This causes a fluency problem as well, because it takes a higher cognitive load to understand the material in another language, process the material, and then comprehend what is being asked, and then changing the structure of the format, and being able to produce written thoughts in a timely manner.

The fourth greatest writing difficulty was anxiety (M = 4.58; SD = 1.18). All 20 of the participants felt some sort of anxiety-provoking behaviors when it came to writing in English. This anxiety often revealed itself in low self-confidence in writing, whether in regard to how they organized their ideas or whether they could adequately detect their own grammatical errors. Organization, generating ideas, and paragraph conclusion (M = 4.04; SD = 1.08) tied for the 11th greatest writing difficulty. In regards to organization, all 20 of the participants expressed problems with having to follow a specified organizational pattern in presenting their ideas in English writing. Participants shared that they did not have the experience to follow specific American organizational patterns in their first language writing. They felt it was un-natural to conform their writing to a specific organizational design.

Overall, the findings indicated that ELL students perceived their highest writing difficulties in linguistic/cognitive deficiencies (M = 3.89; SD = .50), the next highest in psychological/emotional deficiencies (M = 2.77; SD = 1.18) and the third highest in socio-cultural aspects of writing difficulties (M = 2.48; SD = 1.10). Generally speaking, ELL students' experiences and perceptions of writing, are more focused on performance and product writing. Some of the psychological/emotional or socio-cultural factors of writing may be ignored, by most ELLs, due to the fact that these are not scored and instructors do not give feedback on these components. Therefore, this creates a challenge for ELL students when writing in a Western style, being that the sentence structure and words have to be organized in a way that makes it difficult for ELL students to write. By focusing on more meaningful content instead of things such as punctuation, grammar, and spelling, it makes ELL students appear as if they are not up to benchmark standards.

Purpose of Study

CBMs are important because they function as a net to catch those students who may be at-risk for failing and can also be beneficial to classroom teachers to narrow specific areas in writing their students are struggling with and use those data to help. When data from CBMs are used to help inform intervention decisions, students have greater opportunity to improve performance. The growing number of ELL students in schools has created a need to review our current measurement systems to determine if they adequately assess what skills ELL students have learned and monitor their progress toward goals or skill attainment. Being that many cultures write differently from Western culture, especially in sentence structure, our current procedures for screening and progress monitoring may seem disadvantageous to ELL students.

McMaster and Espin (2007) conducted a systematic review which focused on the technical adequacy features of WE-CBM and highlighted the importance of future research examining the use of WE-CBM with English learners. Since the review published by McMaster and Espin in 2007, no updated review has been conducted to focus on the past decade of research on WE-CBM as well as the need for research that includes the growing population of English language learners in schools. Researchers have expressed the value and importance of utilizing CBMs in schools, but much of the research fails to include a population of students who have English as a second language – a growing percentage of our schools. Therefore, this study sought to update McMaster and Espin's work by conducting a systematic review of the literature published on the use of WE-CBM since 2006.

The purpose of this project was to review the existing research on WE-CBM and determine if, since the last review was published in 2007, there are studies specifically examining WE-CBM use for ELLs. This specialist project was a comprehensive, systematic review of the literature on WE-CBM. Although there are fewer studies available compared to other academic areas, it is important to review and synthesize this literature to better understand the current state of the field and possible implications for schools with high ELL populations, as well as for future research areas.

Method

Procedure

This specialist project examined existing literature regarding use of written expression curriculum-based measurements. Articles were identified using the following search engines: EBSCOhost and ERIC (ProQuest). Within EBSCOhost, the following databases were selected: Academic Search Complete, ERIC, PsycARTICLES, PsycINFO, Psychology and Behavioral Sciences Collection. Only peer-reviewed articles were included using the keywords in the following combinations: curriculum-based measurement/CBM, written expression curriculum-based measurement/WE-CBM. Studies were only included if they were published between the years of 2006 and 2019. 2006 was selected due to the lag time between conducting and publishing research. That is, although McMaster and Espin's review was published in 2007, we identified at least one article (Gansle et al., 2006) not included, suggesting the review itself was conducted prior to 2007, though this was not explicitly stated. In addition to the written review of the literature related to the written expression-curriculum-based measurements, studies are also presented in a table. The number of articles found in the initial search was 22. After reviewing the abstracts, four additional articles were removed because no study was conducted, rather they described recommended uses for WE-CBM. One additional article was removed because it did not use WE-CBM as a measure, only the traditional scoring procedures (e.g., TWW) to score a different type of writing. This resulted in 17 empirical studies being reviewed.

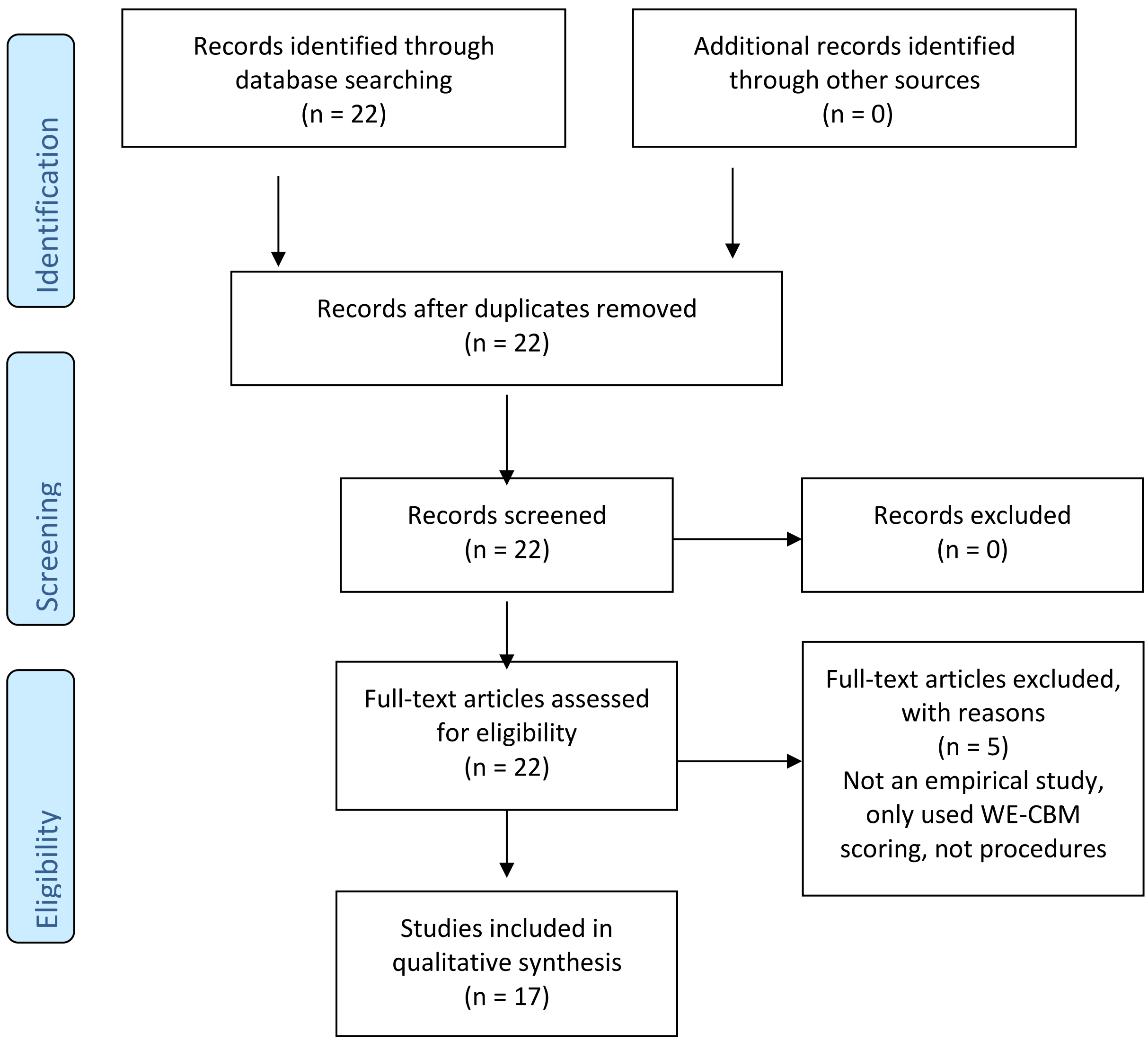

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analyses: The PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009)

Results

McMaster and Espin (2007) reviewed the previous research on WE-CBM from the initial Institute for Research on Learning Disabilities studies conducted during the development of CBM, with the most recently reviewed article from 2005. These results include 17 studies published from 2006 through mid-year 2019. Studies are summarized below and also within Table 1. They have been organized into three broad categories: measurement studies, intervention studies, and studies exploring diverse student groups (i.e., gender, middle and high school students, students with disabilities, and ELL students).

Measurement Studies

Several studies identified examined the reliability and/or validity of WE-CBM using novel statistical approaches, scoring procedures, or designs. These four studies summarized below examine the use of WE-CBM but do not include an analysis of diverse student groups.

Gansle and colleagues (2006) analyzed the reliability and criterion validity of WE-CBM and a trait rubric by using 538 writing samples of students in first through fifth grade. Despite teachers' positive beliefs about the instrument's quality, test-retest reliabilities of the 6+1 model were not acceptable and the validity coefficients were small, resulting a conclusion that the model seemed to lack technical adequacy and was not appropriate to be used with WE-CBM (Gansle et al., 2006).

The study by Codding et al. (2015) researched CBM utility for reading, writing, and mathematics using 249 seventh grade students and investigated patterns and growth rates to examine their predictive validity to a high-stakes state test. The sample

Table 1

Characteristics of Studies Examining WE-CBM

	Sample			Writing Measure		Criterion Validity
Study	Total N (N Special Education)	Grade(s)	Language Status	Time	Scoring Procedures	Criterion Measure
Measurment Studies						
Ganselle, VanDerHeyden, Noell, Resetar, & Williams (2006)	538	1-5	NES	3 min	TWW, CWS, WSC, CP, CC, Complete sentences, Words in complete sentences	Stanford-9
Codding, Petscher, & Truckenmiller (2015)	249 (47)	7	NES, ESL	3 min	TWW, CWS, CIWS, %CWS	MCAS ELA
Furey, Marcotte, Hintze, & Shackett (2016)	109	4	N/A	3 min, 10 min	TWW, CWS, %CWS, CIWS	MCAS ELA
Keller-Margulis, Mercer, & Thomas (2016)	145 (9)	2-5	NES, ELL	7 min	TWW, CWS, WSC, & CIWS	
Mercer, Keller-Margulis, Faith, Reid, & Ochs (2019)	144 (9)	2-5	NES, ELL	7 min	TWW, WSC, CWS, %WSC, CIWS, & %CWS	--
Intervention Studies						
Stotz, Itoi, Konrad, & Alber-Morgan (2008)	3 (3)	4	NES	3 min	TWW, CWS	N/A
Truckenmiller, Eckert, Codding, & Petscher (2014)	133 (20)	3	--	3 min	TWW, CWS	TOWL-3, CBM-M

(Table 1 Continued)

Hier and Excert (2014)	103	3	Excluded students in special education and ELL students	3 min	CWS	TOWL-3, paragraph copying task
Alitto, Malecki, Coyle, & Santuzzi (2016)	Study 1: 114 (11) Study 2: 106 (13)	Study 1: 4 Study 2: 5	Study 1: NES, LEP Study 2: N/A	3 min	TWW, WSC, CWS, CIWS, %WSC, %CWS	--
			Diverse Populations			
Diercks-Granese, Weissenburger, Johnson, & Christensen (2009)	82 (8)	10	--	10 min	ICWS, CPM, ADV, ADJ	WKCE, holistic ratings
Lopez & Thompson (2010)	83	6-8	--	3 min	CWS	AIMS
Amato & Watkins (2011)	447 (66)	8	NES	3 min	TWW, WSC, %WSC, CWS, %CWS, CWS-IWS, SEN, CC, TPM, CP	TOWL-3
Mercer, Martinez, Faust & Mitchell (2012)	163 (17)	10	NES	5 min	TWW, CWS, %CWS, & CIWS,	Indiana Statewide Testing for Educational Progress Plus End-of-Course Assessment in English
Fearrington, et al., (2014)	1,240 (181)	3-8	--	3 min	TWW, CWS	--

(Table 1 Continued)

Hosp, Hensley, Huddle, & Ford (2014)	41	Post-secondary	--	3 min	TWW, WSC, CWS, CIWS, WS Accuracy	WJ Ach III
Keller-Margulis, Mercer, Payan, & McGee (2015)	672 (54)	2-5	NES, LEP	3 min	TWW, CWS, CIWS, %CWS	--
Weiss, Brinkley, & Bock (2019)	324 (23)	3-4	N/A	3 min	TWW, WSC, CWS, CPM, LW, DW, %WSC, %CWS	--
			English Language Learners			
Keller-Margulis, Payan, Jaspers, & Brewton (2016)	139	4	NES, ELL, Monitored	3 min	TWW, CWS, WSC, CP, %CWS, %WSC, CIWS	STAAR Writing

Note. NES = Native English Speaker, TWW = Total Words Written, CWS = Correct Writing Sequence, WSC = Words Spelled Correctly, CP = Correct Punctuation, CC = Correct Capitalization, ESL = English Second Language, CIWS = Correct Minus Incorrect Writing Sequence, MCAS ELA = Massachusetts Comprehensive Assessment System in English Language Arts, ELL = English Language Learner, TOWL = Test of Written Language, CBM-M = Curriculum based measure – Math, ADV = Adverbs, ADJ = Adjectives, WKCE = Wisconsin Knowledge and Concepts Examination, AIMS = Arizona Instrument to Measure Skills, SEN = Sentences, TPM = Total Punctuation Marks, WS = Writing Sequence, WJ Ach = Woodcock Johnson Tests of Achievement, LW = Long Words, DW = Different Words, STAAR = State of Texas Assessments of Academic Readiness contained 58% males, 42% females. 19% of the students in the sample received special education services, and 7% of the students in the sample were English language learners. Examiners used WE-CBM story starters and administered the probes following AIMSWeb protocol. They administered the probes at three different time points in the year. The MCAS-English Language Arts test scores was used as the criterion measure and was completed in the Spring. The MCAS-ELA were composed of three strands: comprehension of reading and literature, composition, and comprehension language. When researchers compared all screening session scores, they noted that all students displayed gains from the fall to spring across all CBM measures. Average increases included 14.71 for TWW, 16.33 for CWS, 1.84 for IWS, and 15.08 for CIWS. Although the findings appeared to be promising, when scores were compared to the criterion construct of MCAS-ELA, there was only a moderate correlation.

Similarly, Furey, Marcotte, Hintze, and Shackett (2016) examined how well WE-CBM could accurately predict writers in the fourth grade who were below proficient level on the state test and those who were above. Fourth grade students (N = 109) completed a story starter and wrote for three minutes in one session, and 10 minutes in another session. The participants also completed the English Language Arts Composition subtest of Massachusetts, a Comprehensive Assessment System (MCAS). Although 53.2% (n= 58) of participants were females, other demographic information was not able to be obtained. When researchers calculated correlations between total scores from each WE-CBM writing task sample with the state achievement test (MCAS), they found that TWW had the lowest correlation with the criterion measure, which was the state achievement test. When researchers increased the WE-CBM sampling time from three minutes to ten minutes, correlation increased as well from .40 to .50. additionally, the correlation between the each of the other scoring indices and the criterion measure also increased with the longer time sampling. When the results were transformed through the Fisher's Z transformation however, the results indicated increased correlations from the longer time sample, and were not statistically significant for any of the metrics.

At three minutes for TWW, in comparison to MCAS, was .40. CWS at three minutes was .57. %CWS at three minutes was .45. CMIWS at three minutes was .60. At 10 minutes, TWW was .50, CWS was .61, %CWS was .50 and CMIWS was .65. Researchers then wanted to test the accuracy of distinguishing between proficient and non-proficient writers by using a cut score and analyzing each index on the state assessment using discriminant function analysis and Receiver Operating Characteristic (ROC) curves. The cut score used, held sensitivity at above .90 specificity for each metric increased when the ten minute probes were used. Researchers found that the 10-minute sample for CWS more accurately identified those students who were at-risk in the center of the distribution. They found that writers that were in the middle of the distribution were more difficult to classify than those students who are struggling and the ones who are successful in writing, making WE-CBM difficult to be used as a screening tool.

Despite the potential of WE-CBM to measure student growth and identify weakness, a study by Keller-Margulis, Mercer, and Thomas (2016) examined the use of WE-CBM as a universal screening tool and its limitations for reliably predicting student growth. There were 145 participants in the study spanning from the second through fifth grade. The sample included 54% female students and 49% of the sample were White students, 23% African American, 17% Hispanic, 8% Asian, and 3% of the students identified themselves as 2 or more races. 8% of the sample were ELLs and 6% of the sample received special education services.

Researchers administered WE-CBM probes to students during three different time points in the school year. The directions were altered and researchers asked students to write for seven minutes instead of three minutes. Students were asked to swap their pencil to a different colored pencil after each minute of writing. By switching pencils, researchers were able to score probes for three, five, and seven minutes. Three probes were administered during each of the three time points. Researchers scored the students samples on TWW, CWS, WSC, and CIWS. Within each grade, CIWS average scores tended to increase across the school year, and as grade increased. Nonsystematic variance accounted for more than half of the total variance for CIWS in all grades. Nonsystematic error ranged from 35% to 44% for TWW, 36% to 57% on WSC, and 40% to 45% on CWS.

The results showed that one 3-minute writing sample did not produce scores that were reliable for the purpose of screening. Writing samples of longer duration, like the 5- and 7-minute writing samples, produced higher levels of reliability for both low and high-stakes decision making.

Despite the various uses of WE-CBM, there are limitations in its technical adequacy, construct representation, and scoring feasibility (Mercer, Keller-Margulis, Faith, Reid, & Ochs, 2019). Mercer et al. (2019) addressed possible solutions to these limitations by examining whether the use of scoring metrics from an automated scoring system (i.e., Coh-Metrix) improved the technical adequacy of traditional WE-CBM. A sample of second to fifth grade students (N = 144) completed 7-minute writing narratives in the fall and winter. Coh-Metrix and traditional WE-CBM scoring were used to predict the holistic writing quality as a means to measure structural validity.

Automated scoring using Coh-Metrix resulted in higher structural validity coefficients than traditional WE-CBM (.77 vs. .70 at fall; .65 vs. .59 at winter). External validity coefficients were not significantly different across the models. Results suggest that not only did Coh-Metrix better predict quality of the writing samples (i.e., structural validity) when compared to traditional WE-CBM scoring methods, but also that Coh-Metrix could be feasible and valid alternative to hand-scoring WE-CBM in the future.

WE-CBM in Intervention Studies

Several studies in the last decade use CBM as a tool in intervention studies. Its utility for both screening and progress monitoring, as well as its feasibility to implement, makes it a popular choice to researchers to include in their protocol. The following studies highlight how CBM is used as a measure in various interventions.

A study by Stotz, Itoi, Konrad, and Alber-Morgan (2008) researched the effects of WE-CBM self-graphing for three fourth-grade students who had high incidence disabilities. All three students had IEP goals that addressed written expression. The researchers looked at TWW and CWS for all three students. Data and intervention took place four days a week, over the course of 30 days. Each of the students were below the benchmark for the fourth grade according to AIMSweb norming data. These students had the story starter repeated by the interventionist several times throughout the three-minute writing session. The students were then taught by the interventionist how to score their TWW and then graph it on paper by bar graph format. The interventionist determined the CWS without the students present, and reviewed the TWW to ensure they were counting correctly. By the end of the 30 days, student one's TWW from a baseline of 29, had increased 51.5. As for CWS, student one at baseline was 14.3 and went to 34.6 average. Student two's TWW at baseline was initially 44.4 and increased to 58.8 average. Student two's CWS at baseline was 29.0 and at maintenance phase went to 42.6 average. Student three's TWW at baseline was at 36.6 and went to 61.7 average at maintenance phase. Student three's CWS at baseline was at 21.6 and increased to 32.0 average at maintenance phase. All three students increased in both TWW and CWS. The examiners also noted that teaching self-graphing for the CBM indices may have benefitted the students because it encouraged them to be more responsible for their writing goals and it allowed them to track and monitor their own progress.

A study by Truckenmiller, Eckert, Codding, and Petscher (2014) researched 133 third grade students on writing fluency growth in response to instructional practices, sex differences, and individual students level of writing. The primary aim of this study was to examine the impact of basic instructional practices (i.e., practice, performance feedback, and typical instructional practices) on commonly used CBM-WE outcome metrics (i.e., TWW and CWS) during a brief intervention period (i.e., progress monitoring period). There were three conditions that students were randomly assigned to. Group A, which was individual performance feedback condition (n = 46), group B, a practice-only condition (writing practice, n = 39), and group C, control condition (n = 48). The students remained in these groups for a total of eight weeks.

Group A gender breakdown was as follows: 59% females (*n* = 27), 41% males (*n* = 19). The racial breakdown: 76% African American (*n* = 34), 0 Native American, 18% Caucasian (*n* = 8), 7% Hispanic (*n* = 3). Educational breakdown had 17% (*n* = 8) receiving special education services.

Group B gender breakdown was as follows: 46% females (*n* = 18), 44% males (*n* = 21). The racial breakdown was 80% African American (*n* = 32), 5% Native American (*n* = 2), 15% Caucasian (*n* = 6), 0% Hispanic (*n* = 0). The educational breakdown had 15% (*n* = 3) receiving special education services. Group C gender breakdown had 56% females (*n* = 27), 44% males (*n* = 21). The racial breakdown had 71% African American (*n* = 35), 0 Native American, 20% Caucasian (*n* = 10), 8% Hispanic (*n* = 4). The educational breakdown had 18% (*n* = 9) receiving special education services.

The students completed a writing prompt using a story starter, but were also given the Test of Written Language (TOWL-3), and scored on TWW and CWS. Interscorer agreement were within in a range of acceptability between scorers. The average performance on the TOWL-3 Spontaneous Writing Quotient (*M* = 84.31, *SD* = 32.05) fell more than one standard deviation below the mean. However, the results of the CBM-WE probe indicated that students wrote an average of 31.50 TWW in 3 min (*SD* = 12.44) and 26.86 CWS in 3 min (*SD* = 12.22). The students' baseline performance in this sample corresponded to the 50th percentile for third-grade students (Pearson Education, Inc., 2009). Students' performance on each of these baseline measures (TOWL-3 Spontaneous Writing Quotient, CBM-WE, and CBM-M) was significantly correlated with one another (range *r* = .33 to .91). One of the significant differences in performance on WE-CBM was found between boys and girls, with the girls significantly outperforming the boys in TWW, $F(1,131) = 5.15$, $p = .025$ and CWS, $F(1,131) = 5.85$, $p = .017$.

The results highlighted the significance that performance feedback improved writing fluency at a statistically significant level with a moderate effect size, but it also demonstrated clinical significance in their study. In just 7 weeks, the students assigned to the feedback condition in the current study reached 41.6 TWW, a level of writing fluency that exceeds the designation for instructional level for third grade (i.e., 37 TWW in 3 min; Mirkin et al., 1981).

Similarly, Hier, and Eckert (2014) also examined the impact of performance feedback on writing fluency. This study looked at the generalization of past studies on performance feedback and maintenance of writing fluency developed through interventions. Researchers randomly assigned 103 third grade students in their sample to one of two conditions. Male and female distribution between conditions were about equal. The racial breakdown consisted of 13.6% ($n = 14$) students being identified at Hispanic or Latino, and 86.4% ($n = 89$) students identified as being not Hispanic or Latino. 94.2% ($n = 97$) of the students were in the general education classroom and 5.8% ($n = 6$) of the students were in special education. The first condition was a practice-only condition and the other condition was a performance feedback condition. Researchers predicted that students in the performance feedback condition would demonstrate greater writing fluency improvement than those in the practice-only condition, immediately following treatment. Researchers had 10 sessions for students to complete one WE-CBM probe per session. Students writing performance was measured using a t-test between the TOWL-3 and the WE-CBM scores. The performance feedback condition scored higher than the practice only condition, on-average gaining 2.62 CWS per week, whereas the PO condition only gained an average of 0.35 CWS per week and remained at the Frustrational level, according to Aimsweb norming data for third grade students. Students who were at Mastery level during baseline, made less improvements on CWS, averaging 0.93 CWS per week. In the PO condition who had students at Mastery level at baseline decreased by an average of 2.21 CWS per week. Overall, the performance feedback condition showed better improvements for immediate effects of writing fluency, but was limited for long-term effects.

Alitto, Malecki, Coyle, and Santuzzi (2016) replicated a previous study by looking at the effects of adult and peer mediated goal setting and feedback interventions for writing. They completed two different studies. The first study was conducted to replicate an existing study that supports the utility of goal setting and teacher given performance feedback as an intervention using CBM. The study had 114 4^{th} grade students 49% male (56), 51% female (58). 103 were in general education classrooms, 11 (10%) in special education services. Over 96% of the participants were white. Each of the students were given a packet at baseline in both the experimental and control group. Both groups were given a story starter and then asked to write for 3 minutes. The experimental group packet contained a cover sheet, the scored probe from the previous week, a TWW chart, a %CWS chart, a page with a stop sign, and a new probe. Students in the control group had a packet that contained a cover sheet, a humidity chart, a temperature chart, and a page with a stop sign, and a new probe. The intervention group had individual goals listed in their packet for TWW and %CWS and they received feedback toward goals. A study from Jewell (2003) found that a growth of .80 words written per week on the TWW index and .38 on the %CWS index for students in the third grade, so ambitious goals were set based upon this study's findings. TWW was increased by two words per week and %CWS increased by one percentage point per week. If the student did not reach their goal from the week prior, the goal from the week prior remained the same the following week. Long term goals were set up by adding 10 and twenty and ten to each students TWW and %CWS baseline performance. The students in the experimental group also received feedback from a computer-generated bar graph that listed previous week's performance, individual goals and a graph of ten weekly performance data points. If the student met their goal, the words, 'Great job! You met your goal!' were added below the performance information. The control group had the distractor charts related to weather conditions.

The results from the first study on TWW showed that the intervention group produced significantly more words than control group (*b* = 6.32, *p* = .001). The slope for time differed by group such that intervention group showed significantly steeper positive slope (*b* = 1.12, *p* < .001) compared to the control (*b* = 0.64, *p* < .001). The control group averaged about .64 words each additional time point, the intervention group showed increases of 1.12 words each time, with the effect size of .12 for the slope based on the proportion of variance. As for WSC, the intervention group produced significantly more words than the control group (*b* = 6.39, *p* = .002) the interaction between group and time was also statistically significant suggesting that the rate of change for WSC was greater for the intervention group as compared to control (*b* = 0.48, *p* = .035). The effect size was .12 for the slope based on the proportion of variance.

On CWS, significant for groups after the intervention ($b = 6.48$, $p = .003$). The intervention group produced 6.48 more correct word sequences on average than the control group. Linear increases in CWS were different between groups. The effect size was .28 for the slope based on the proportion of variance. Significant interaction between group and time suggested that the intervention group had an increase of 1.19 CWS for each time point ($b = 1.19$, $p < .001$) compared to the 0.66 CWS increase for the control group. As for IWS, the results revealed no significant IWS differences between groups after the initial intervention or in their linear changes over time.

In the second study, researchers had 106 5^{th} grade students. 93 of the participants (88%) were in the general education classroom and 13 (12%) in special education. Racial background: 42% Caucasian, 12% African American, 37% Hispanic, 1% Asian/Pacific Islander and 8% multi-racial. Limited English proficiency was not provided in this sample, but the overall, the students in the school who were limited in English proficiency were 7.3%. This study was set up similarly to the first study, the difference being that peers gave feedback on other student's TWW goals after they completed the writing assignment. The control group did not receive any feedback. The strategies were to enhance student's writing with allowing frequent feedback on their writing performance, teaching students how to evaluate their own writing and monitoring their own progress over time. The results found a similar pattern observed in Study 1, which concluded there was a significant difference between the intervention group and the control group. After the intervention, for TWW ($b = 14.56$, $p < .001$), such that the intervention group wrote 14.56 more words than the control group. Moreover, there was also a significant interaction between group and time that indicated that students in the Intervention Group showed a significantly steeper increase in words written across the intervention period relative to the Control Group. The interaction yielded an effect size of .21 for the slope based on the proportion of the variance. The Control Group showed a non-significant increase over time ($b = 0.20$, $p = .19$) whereas the intervention group showed a significant increase of 0.92 words with each time point ($b = 0.92$, $p < .001$).

As for WSC, the intervention group yielded significantly more words spelled correctly (12.86 words) after the intervention than the Control Group, $b = 12.86$, $p < .001$. The Intervention Group also showed a significantly steeper increase in words spelled correctly across the intervention period, as indicated by the significant interaction between Group and Time. The interaction yielded an effect size of .15 for the slope based on the proportion of the variance. The Control Group showed a non-significant increase of 0.21 correctly spelled words per time point ($b = 0.21$, $p = .09$), the intervention group yielded an increase of 0.76 words spelled correctly per time point ($b = 0.76$, $p < .001$).

As for CWS, after the intervention, the intervention group showed 9.72 more correct word sequences than the Control Group; this difference was statistically significant, $b = 9.72$, $p = .001$. Contrary to expectations, students in the Intervention Group showed only slightly steeper increases in correct word sequences ($b = 0.51$, $p < .001$ relative to the Control Group ($b = 0.27$, $p = .03$) as indicated by the marginal interaction between Group and Time ($b = 0.31$).

ICWS results differed from the results from Study 1. The students in the Intervention Group showed significantly more incorrect word sequences as compared to the Control Group ($b = 5.62$, $p = .001$). Also, the interaction between Group and Time was significant ($b = 0.39$, $p = .001$). This yielded an effect size of .20 for the slope based on the proportion of the variance. The students in the Control Group showed nearly no change in incorrect word sequences with each time point ($b = 0.02$, $p = .84$). However, the Intervention Group showed an increase of 0.41 incorrect sequences per time point ($b = 0.41$, $p < .001$). Although the number of correct word sequences seemed to increase in both groups, the number of incorrect word sequences also seemed to increase across the intervention period for the intervention group.

The finding from this study indicate that when peers are scoring work, the ICWS increases, possibly because students are trying to write more words to increase their TWW goal, so they make more mistakes.

WE-CBM with Diverse Populations

Several studies in the last decade explored differences in use of WE-CBM for different groups including different genders, individuals with disabilities, different grade levels, or ethnicities.

Gender differences in WE-CBM. Fearrington et al. (2014) researched gender differences on WE-CBM for students in the third through eighth grade. They looked at production dependent measures of WE-CBM including TWW and CWS at each benchmark (Fall, Winter, and Spring) to determine if there were any differences across genders. The study had 1,240 general and special education students in third through eighth grades, which were distributed fairly evenly across each grade. The sample included 600 boys (48%) and was predominantly Caucasian (94%). Additional ethnicities represented in the sample were African American (1.4%), Hispanic (3.4%), Asian American (0.5%), and American Indian (0.3%). The percentage of students receiving special education services was 14.6%.

After data collection, researchers found that for TWW, using a series of two-way, repeated measures ANOVAs, with time of year (fall, winter, spring) serving as the within- subjects variable and gender serving as the between-subjects variable, revealed a significant main effect for time for Grade 3, $F(1, 193) = 58.87$, $p < .001$. the effect of time was not significant for grades 4 and 5. The main effect of gender was significant in favor of girls for Grade 3, $F(1, 193) = 34.99$, $p < .001$. The effect sizes were medium for Grades 4 and 8 (ES = 0.103 and 0.121, respectively). There was a significant interaction between gender and time for Grade 3, $F(1, 193) = 16.17$, $p < .001$. The effect size for this interaction was 0.08, which is considered medium. No significant interaction effects were found for the remaining grades.

As for CWS, a series of two-way repeated-measures ANOVAs, with time of year serving as the within-subjects variable and gender serving as the between-subjects variable, revealed a significant main effect for time in Grade 3, $F(1, 193) = 53.06$, $p < .001$. Effect sizes for these differences were large for Grades 3 (ES = 0.216). Effect sizes were moderate for Grade 5 and small for Grade 4 (ES = 0.072 and 0.049, respectively). Significant main effects for gender were found in Grade 3, $F(1, 193) = 29.37$, $p < .001$. Effect sizes for these differences were medium for Grades 3, 4, and 8 (ES = 0.312, 0.118, and 0.13, respectively) and large for Grades 5, 6, and 7 (ES = 0.166, 0.153, and 0.242, respectively).

A significant interaction between gender and time was found for Grade 3, $F(1, 193) = 24.09$, $p < .001$ and Grade 7 $F(1, 207) = 5.25$, $p < .05$. The effect size for Grade 3 was 0.111(medium). The effect size for Grade 7 was small (ES = 0.025). There was not a significant interaction between gender and time for Grades 4, 5, 6, and 8.

In conclusion, girls' scores increased significantly more than did boys' scores from fall to spring benchmarks, within each grade sample. This study's findings demonstrate support for sociocultural theories of gender differences in the classroom.

Similarly, a study by Keller-Margulis et al. (2015) examined the annual growth and gender differences in WE-CBM when used as a universal screening tool. The study's sample consisted of 672 students between the second and fifth grade. Demographic information was only accessible for 93% of the sample, but of what was obtained, 50% of the sample were females and 50% were males. Racial breakdown contained 55% Hispanic, 36% Black, 5% White, 4% Asian, and <1% American Indian. 19% of the sample were identified as Limited English Proficient and 8% were eligible for special education services.

The students were all given WE-CBM story starters and were scored on TWW, CWS, CIWS, and %CWS. They were all grouped by grade level for accurate analysis purposes. For TWW, growth patterns were found to be curvilinear for each grade, and increased in curvilinearity as grade increased. There was only statistically significant variance in second and third grade, but in the fourth and fifth grade, there were fewer individual differences in slope. As for CWS, second grade growth pattern was found to be linear, but increased in curvilinearity in the higher grades. Researchers found that for CIWS, second through fourth grade, growth was linear, but curvilinear for the fifth grade. For %CWS growth was linear in second and third grade, but curvilinear for fourth and fifth grade. Growth was found to be inconsistent in the fourth grade, declining in the fall semester, and increasing in the spring semester in the fifth grade. In terms of gender differences, boys had statistically lower scores on initial WE-CBM assessments than girls. The boys experienced less growth than girls on CWS in third grade, but boys experienced more growth than girls on %CWS in all grades.

Research from the past decade highlights the various uses of WE-CBM. One study by Weiss, Brinkley, and Bock (2019) focused on how WE-CBM could be utilized to reveal more about a student's strengths and weaknesses, than about what was actually written by the student in a sample, such as a student's vocabulary choice. The authors wanted to investigate student performance over time, as well as the characteristics that may be related to the student's score such as grade, disability, gender, and handwriting. The study consisted of 324 third and fourth grade students who completed WE-CBMs, during three time points: fall, winter, and spring. Of the 154 third graders, 51.9% (n = 80) were male and of the 170 fourth graders, 56.5% (n = 96) were male. About 7% of all students received special education services.

Researchers conducted t-tests with the fall WE-CBM probes to see if there were any differences based on gender, hand writing, grade level, or disabilities. Both third and fourth grade students demonstrated similar performances in the number of words they produced, but the fourth graders produced many more CWS than the third-grade students. Female students on average also produced more TWW and CWS than the male students. Females wrote more correctly spelled words, longer words, and different words than the males. Related to special education classification, students who were receiving special education services produced less words on average than those in the general education classroom. This study demonstrated differences in performance on WE-CBM based on gender and special education status.

Students with disabilities. One study by Hosp, Hensley, Huddle, and Ford (2014) examined the effectiveness of CBMs with 41 post-secondary students that had intellectual and developmental disabilities. All of the students met criteria for intellectual, developmental, or multiple disabilities and were enrolled in a 2-year certificate program for students with disabilities. Intellectual disability was ranging from mild to moderate. The program had coursework that consisted of mathematics, writing, and reading. Nineteen of the students in the study were identified as having cognitive disability, nine with ASD, and thirteen with multiple disabilities. Fourteen (34%) of the participants were female, 27 (66%) male. Thirty-seven (90%) students were Caucasian, and the other races/ethnicities were not noted in the study.

Each of the students were given a story starter and then asked to write for three minutes. They were scored on TWW, CWS, WSC, and CIWS. They were then given several subtests from the Woodcock Johnson-III test of achievement. Researchers used bivariate correlations to calculate each CBM and the content-appropriate criterion measures. They found that all four of the CBM metrics for writing had similar patterns with the WJ-III Broad Written Language cluster (r = .531-.669) with higher correlations with Broad Written Language subtest than spelling or writing fluency. The only significant differences were that CWS had high correlations with Broad Written Language than TWW or WSC (z = 2.138 and 2.131).

CWS is a broader metric of writing than TWW and WSC, which was expected. These results emphasize the strength and stability that CBM measures performance.

Middle and high school studies. CBMs are commonly used in elementary grades and occasionally in middle school, but much less often in high school; however, one group of researchers examined the technical adequacy of WE-CBM for use with high school students (Diercks-Gransee, Weissenburger, Johnson, & Christensen, 2009). General (*n* = 74) and special education (*n* = 8) students in 10th grade completed two narrative writing samples during a winter time point. Students wrote for 10 minutes in response to a provided story starter. Writing samples were scored for incorrect word sequences (ICWS), and the number of correct punctuation marks (CPM), adjectives (ADJ), and adverbs (ADV). Performance on the language arts statewide assessment and a holistic score generated from a teacher rating were used as criterion measures.

Students in special education wrote significantly more ICWS and general education students wrote significantly more CPM. CPM and ICWS were also related to the holistic score given by the teacher (.62-.71, respectively) whereas ADJ and ADV demonstrated weak and insignificant relationships with the holistic score. Similar results were found for the state test. Students who produced more ICWS tended to have lower state test scores (-.51). CPM demonstrated a weak correlation (.28) and ADJ and ADV were not significantly related to state test scores. Lastly, when the 20th percentile was applied as a cut score, ICWS correctly identified 87.5% of students classified as having a disability and 56% of students who scored below 50 (out of 99) on the state test. These results suggest, that while additional research is needed, ICWS and potentially CPM could be viable scoring options for using WE-CBM with high school students for predicting state test performance or screening for a learning disability.

Another study conducted by Lopez and Thompson (2010) examined the accuracy of WE-CBM using middle school students in sixth, seventh, and eighth grade and their performance on statewide achievement test in Arizona. The cut score that was used for this study looked at CWS that was one standard deviation below the mean. The CWS score for the sample that was selected. Researchers were able to establish a cut score of 40 CWS for all three grade levels by using this one standard deviation below the mean method. The score of 40 CWS was able to propose high levels of accuracy for performance on the state test, although direct comparisons across grade levels were unable to be determined.

Amato and Watkins (2011) examined the predictive validity of CBM writing indices for monitoring 8^{th} grade students writing proficiency. They looked at the size of the overall relationship of criterion variable and set a predictor variable. The population consisted of general (n = 381) and special education (n = 66) students, ages 12-16. They were given a story starter and wrote for three minutes. WE-CBMs were scored for TWW, WSC, %WSC, CWS, %CWS, CIWS, number of sentences, correct capitalization, punctuation marks, and correct punctuation marks.

They were then compared to the criterion variable, the Test of Written Language-3^{rd} edition (TOWL-3). Three variables were omitted (WSC, CWS, and number of punctuation marks) and so seven of the predictor variables remained. The three were omitted due to current research on the high correlation for those variables. The remaining 7 together accounted for only 44% of the variance on the TOWL-3 scores. Only three of the seven variables exclusively contributed to prediction of TOWL-3 scores which were the percentage of correct word sequences, correct punctuation marks, and correct capitalizations. CWS had the strongest bivariate correlation with TOWL-3 r =.61 and also contributed to the most unique variance (37%) in prediction of TOWL-3 scores. In upper grades, accuracy was a better predictor of writing performance than fluency. The study suggested that educators may wish to rely on other qualitative and quantitative measures for student writing, including norm-referenced tests for instructional or high stakes decisions.

Mercer et al. (2012) also examined the predictive validity of CBM to high-stakes exams; however, they narrowed their scope on the validity of the written measures. This study examined the differences between criterion-related validity of production dependent (TWW and CWS), accurate-production (CIWS), and production-independent (%CWS) scoring on writing probes in relation to state mandated writing assessment. Students included 163 10^{th} graders from a rural high school in the Midwest, of whom 52.8 % were female. All of the participants were native English speakers. Based on school records, 93.9% of participants identified as Caucasian, 2.5% as Asian, 1.8% as multiracial, 1.2% as Hispanic, and 0.6% as African American. In addition, 10.4% met eligibility for special education and 6.1% of all students were classified with a specific learning disability.

The students completed the WE-CBM and wrote for five minutes. Halfway through the time they were all told as a group an encouraging comment by the teacher, telling students to write the best story they could. The participants then completed the criterion referenced measure, which was The Indiana Statewide Testing for Educational Progress Plus End-of-Course Assessment in English. The English 10 ECA consists of three item types: selected-response (i.e., multiple choice), constructed response (i.e., short answer), and prompted writing samples. Approximately 73% of the ECA total score represents performance in Reading comprehension, with the remainder reflecting performance in writing applications. Passing scores are determined by a statewide committee of teachers using the bookmark procedure for standard setting (Mitzel, Lewis, Patz, & Green, 2001).

The results of this exploratory study found that for 10th grade students, narrative probes scored using more complex indicators of WE-CBM such as CWS, CIWS, and % CWS— explained more variance in the primary criterion measure (i.e., ECA writing) than WE-CBM scores on expository probes. The better performance of these complex indicators, for the narrative probes, is consistent with previous research that suggests CWS and CIWS have larger validity coefficients for secondary students (Espin et al., 2008; Jewell & Malecki, 2005). The study also showed that when passage copying speed was considered, the %CWS on the narrative probes explained the greatest amount of variance in the state writing assessment. When speed was not taken into account, CIWS on the narrative probes explained the most variance on the state writing assessment. However, the study found that with or without the copying speed as an additional predictor, none of the WE-CBM scores had explained enough variance in the state writing assessment to be considered useful when predicting performance on high-stakes exams.

English Language Learners. Only one study specifically examined differences in WE-CBM performance across diverse language groups. Keller-Margulis, Payan, Jaspers, and Brewton (2016) examined the technical adequacy of WE-CBM for ELLs in elementary grades. The study involved 139 fourth-grade elementary students from a culturally and linguistically diverse school in the southeastern United States. Of those, 89 were NES, 31 were monitored, and 19 were ELL. In this study, monitored means that a student was previously in ELL but had transitioned out and were in their first or second year of being monitored. Each of the participants completed the WE-CBM probes in English. A total of 81.29% of the fourth grade students in the study passed the State of Texas Assessments of Academic Readiness (STAAR) Writing test, including 76.40% of the native English-speaking students, 78.90% of ELLs, and 96.80% of Monitored students. Researchers determined an appropriate cut-score (.70) and compared each of the students WE-CBM scores to their STAAR score. Diagnostic accuracy of three groups of WE-CBM indices (production-dependent, production-independent, and accurate-production) were examined as related to the statewide test scores. Results determined that production-dependent indicators (e.g., TWW) were valid only for non-ELL students and none of the coefficients for production-dependent indices were significant for ELL students. For production-independent indices, only CIWS was significant across different groups of students. Researchers hypothesized that the reason they received these results could be due to the unique nature of the language instruction and development for those students who are still learning English.

Discussion

CBMs are a tool that can be used for screening, progress monitoring, and supporting educational decision making. Although popular in schools, little is known on the usefulness of CBM for diverse populations, specifically WE-CBMs. In McMaster and Espin's (2007) initial review on the technical adequacy of WE-CBMs, they highlighted the need for future research on the utility of WE-CBM with English learners. Given the growing number of ELL students served in schools, it was important to provide an updated review and see if this increased number of ELLs was reflected in the types of studies found in the literature. This study sought to serve as an update on McMaster and Espin's (2007) review and examine research on the use of WE-CBM for ELLs.

Overall, after reviewing 22 identified articles for eligibility, 17 were included in this review. These articles were grouped into four broad categories: (a) articles examining measurement approaches to studying WE-CBM (n = 4), (b) intervention studies using WE-CBM (n = 4), and (c) the use of WE-CBM with diverse populations (n = 9), and (d) use of CBM with ELL students. Only one study specifically examined the use of WE-CBM with ELL students, though several studies did include students from diverse language backgrounds within their sample. It is noteworthy that in over a decade, only one article within the literature was identified that specifically focused on the use of WE-CBM for ELLs. Without having any studies examining the usefulness of WE-CBMs for culturally and linguistically diverse students, it is impossible to determine their effectiveness for this population. Without evidence to support our assessment decisions for ELLs, we may be making inaccurate decisions, inaccurately and wastefully allocating resources, or hindering academic progress.

Implications for Research and Practice

Results from this study have implications for both research and practice. WE-CBM has become an increasingly common tool used within the schools, and articles within this literature review highlight both the usefulness and the various ways in which WE-CBM can be used.

As mentioned previously, there are various uses for CBMs. Not only is WE-CBM useful for screening, but it also aids in progress monitoring as well as, educational decision making. Several studies in this literature review highlighted the use of CBMs to measure student growth and potential weaknesses. Because of the little resources required to implement CBMs, these measures are a quick, feasible way for school staff to identify student progress and identify areas in need of growth. Moreover, CBM can be used within schools for treatment planning and educational decisions. One study in this review examined the usefulness of CBM to aid in IEP writing and decisions. School professionals are able to use information obtained from CBMs to center student’s individualized treatment and goal setting.

In addition to the various uses of CBM highlighted in the literature review, several articles highlighted the predictive validity of CBMs. This information implies that performance on CBMs are able to predict later performance on high-stakes testing (e.g., statewide testing). Not only this, but studies from this literature highlighted the importance of CBMs as a screener to identify at-risk students. Similar to its predictive power for performance on high-stakes testing, poor performers on CBM measures can inform school administration and staff of students who may need additional support, and in what areas.

Limitations and Future Directions

One limitation of the study itself was the low number of articles actually obtained. Because the researcher only had access to the University's database, there was a limited number of articles that met inclusion criteria of this study. As a result of this limited access to research articles, it is likely that there are additional articles that were missed that could have been included. More specifically, it is possible that articles pertaining to the research question (i.e., the use of WE-CBM with ELLs) were excluded due to this lack of access. Similarly, a decision was made to not include dissertation or theses, which may have also limited the scope of findings.

Future research should expand upon the information found from this literature review and address limitations seen in this study. Future studies should look at writing intervention decisions and progress monitoring based off of these WE-CBM screeners, and examine whether WE-CBM can be sensitive to instructional changes, particularly for diverse groups of students such as ELLs. Most intervention studies currently published are not specific to writing, and more research is needed to examine the usefulness of WE-CBM in intervention planning and maintenance.

Education is an ever-changing field and both researchers and practitioners need to stay up-to-date on current policy and research and use assessment tools, such as WE-CBM that are appropriate for the current population and decision to be made. Recently, Texas, a state with the second highest number of ELLs, passed a bill (HB3906) that will eliminate the stand-alone state writing assessment and require all state assessments to transition to online formats. Given these changes, which will likely look similar in other states, it is important to continue to research the necessity and appropriateness of our screeners within a specific educational context for a specific and unique population.

In addition to using WE-CBM for intervention decision, future research should seek to confirm the challenges and difficulties seen with WE-CBM before making it a reliable, decision-based universal screening tool for all populations.

Finally, it is imperative for future researchers to examine WE-CBM with the ELL population. This research can be conducted by comparing WE-CBM scores of ELL students to NES students' scores, or another criterion measure, such as the state assessment. With the growing population of ELLs within the school system, it is vital to ensure that measures used are technically adequate and psychometrically sound for diverse populations. With legislation such as NCLB requiring annual screening and assessment for students, school personnel must ensure that the measures used are appropriate for all students, and as the United States becomes increasingly diverse, it is a challenge that we must constantly work to overcome. Based on the results of this study, the field has made very minimal progress toward including and targeting ELLs within our work on WE-CBMs.

References

Alitto, J., Malecki, C., Coyle, S., & Santuzzi, A. (2016). Examining the effects of adult and peer mediated goal setting and feedback interventions for writing: Two studies. *Journal of School Psychology, 56*, 89-109.

Amato, J., & Watkins, M. (2011). The predictive validity of CBM writing indices for eighth- grade students. *The Journal of Special Education*, *44*(4), 195-204.

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from tri-annual universal screenings. *School Psychology Review, 37,* 109–125.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-Based Measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39,* 447– 462.

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. http://dx.doi.org/10.1037/a0037520

Deno, S. L. (2003). Curriculum-Based Measures: Development and perspectives. *Assessment for Effective Intervention*, *28*, 3–12.

Diercks-Granese, B., Weissenburger, J. W., Johnson, C. L., & Christensen, P. (2009) Curriculum-based measures of writing for high school students. *Remedial and Special Education,30*(6), 360-371.

Espin, C., Shin, J., Deno, S., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *Journal of Special Education*, *34*, 140–153.

Fearrington, J., Parker, P., Kidder-Ashley, P., Gagnon, S., McCane-Bowling, S., & Sorrell, C. (2014). Gender differences in written expression curriculum-based measurement in third- through eighth-grade students. *Psychology in the Schools*, *51*(1), 85-96.

Fuchs, L. S., & Fuchs, D. (2004). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. *Assessment for Effective Intervention, 29,* 25–37.

Fuchs, L. S., & Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993) Formative evaluation of academic progress-how much growth can we expect? *School Psychology Review*, *22*, pp. 27-48.

Furey, W., Marcotte, A., Hintze, J., & Shackett, C. (2016). Concurrent validity and classification accuracy of curriculum-based measurement for written expression. *School Psychology Quarterly, 31*(3), 369-382.

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review, 35*(3), 435-450.

Hamp-Lyons, L. (1991). *Assessing second language writing in academic contexts.* Norwood, NJ: Ablex.

Hier, B., & Eckert, T. (2014). Evaluating elementary-aged students' abilities to generalize and maintain fluency gains of a performance feedback writing intervention. *School Psychology Quarterly, 29*(4), 488-502.

Hosp, J., Hensley, K., Huddle, S., & Ford, J. (2014). Using curriculum-based measures with postsecondary students with intellectual and developmental disabilities. *Remedial and Special Education, 35*(4), 247-257.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: 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*, 27–44.

Keller-Margulis, M. A., Mercer, S. H., & Shapiro, E. S. (2014). Seasonal differences in growth on math curriculum-based measures using tri-annual benchmarks. *Assessment for Effective Intervention, 39*, 146–155.

Keller-Margulis, M. A., Mercer, S. H., Payan, A., & McGee, W. (2015). Measuring annual growth using written expression curriculum-based measurement: An examination of seasonal and gender differences. *School Psychology Quarterly, 30*(2), 276-288.

Keller-Margulis, M., Mercer, S., & Thomas, E. (2016). Generalizability theory reliability of written expression curriculum-based measurement in universal screening. *School Psychology Quarterly, 31*(3), 383-392.

Keller-Margulis, M. A., Payan, A., Jaspers, K., & Brewton, C. (2016). Validity and diagnostic accuracy of written-expression curriculum-based measurement for students with diverse language backgrounds. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, *32*, 174-198.

Lin, S. M. (2015). A study of ELL students' writing difficulties: A call for culturally, linguistically, and psychologically responsive teaching. *College Student Journal, 49*(2), 237-250.

López, F. A., & Thompson, S. S. (2010). The relationship among measures of written expression using curriculum-based measurement and the Arizona Instrument to Measure Skills (AIMS) at the middle school level. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *27,* 129 – 152.

Marston, D. (1989). A curriculum-based approach to assessing academic performance: What it is and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children,* (pp. 18-78). New York, NY: Guilford Press.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*(2), 68-84.

Mercer, S. H., Martínez, R. S., Faust, D., & Mitchell, R. R. (2012). Criterion-related validity of curriculum-based measurement in writing with narrative and expository prompts relative to passage copying speed in 10th grade students. *School Psychology Quarterly, 27,* 85–95.

Mercer, S., Keller-Margulis, M., Faith, E., Reid, E., & Ochs, S. (2019). The potential for automated text evaluation to improve the technical adequacy of written expression curriculum-based measurement. *Learning Disability Quarterly, 42*(2), 117-128.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., The PRISMA Group (2009). *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement.* PLoS Med 6(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097

National Center for Education Statistics. (2019). *The condition of education: English Language Learners in Public Schools (NCES2012- 470)*. Washington, DC: Institute of Education Sciences, U. S. Department of Education. Retrieved from nces.ed.gov/programs/coe/indicator_cgf.asp.

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

Nelson, M. W. (1991). *At the point of need: Teaching basic and ESL writers.* Portsmouth, NH: Heinemann.

Pearson Education. (2014). *AIMSweb national norms: WE- CBM.* Retrieved from http://www.aimsweb.com

Persky, H. R., Daane, M. C., & Jin, Y. (2003). The nation's report card: *Writing 2002*. (NCES 2003–529). Washington, DC: U.S. Government Printing Office.

Stotz, K. E., Itoi, M., Konrad, M., & Alber-Morgan, S. R. (2008). Effects of self-graphing on written expression of fourth grade students with high-incidence disabilities. *Journal of Behavioral Education, 17*(2), 172-186.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge, MA: Cambridge University Press.

Truckenmiller, A. J., Eckert, T. L., Codding, R. S., & Petscher, Y. (2014). Evaluating the impact of feedback on elementary aged students' fluency growth in written expression: A randomized controlled trial. *Journal of School Psychology*, *52*(6), 531-548.

Weiss, S., Brinkley, J., & Bock, J. (2019). Patterns of growth: Cluster analysis of written expression curriculum-based measurement. *Education and Treatment of Children*, *42*(2), 161-184.

Weissenburger, J. W., & Espin, C. A. (2005). Curriculum-based measures of writing across grade levels. *Journal of School Psychology*, *43*, 153–169.

Wright, J. (1992). *Curriculum-based measurement: A manual for teachers.* Retrieved September 23, 2011, from: http://www.jimwrightonline.com/pdfdocs.

Yell, M., Deno, S. L., & Marston, D. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique,*18(1), 99-105.