The Pennsylvania State University

The Graduate School

EXPANDING THE USE OF CURRICULUM-BASED MEASUREMENT FOR WRITTEN EXPRESSION: A COLLEGE FRESHMEN NORMING STUDY

A Dissertation in

School Psychology

by

Victoria L. Buser

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2020

The dissertation of Victoria L. Buser was reviewed and approved* by the following:

Cristin M. Hall
Assistant Professor of Education (School Psychology)
Dissertation Adviser
Chair of Committee

Shirley A. Woika
Professor of Education (School Psychology)

Katie E. Hoffman
Teaching Professor of Education (Special Education)

Anne Elrod Whitney
Professor of Education

James C. Diperna
Professor of Education (School Psychology)
Head of the Department of School Psychology

ABSTRACT

Strong writing skills are essential for students from primary through post-secondary education, as there is evidence that writing skills are related to both academic performance and success in college and the workplace. One common method used to assess students' current writing skills, monitor students' progress related to writing skills, and aid in the identification of learning disabilities in written expression is Curriculum-Based Measurement in Written Expression (CBM-WE). Currently, CBM-WE norms only exist through Grade 8; however, CBM-WE norms could have utility with older students as well. The purpose of the current study was to develop CBM-WE norms based on college freshmen. Potential uses for these college CBM-WE norms are: a) providing a benchmark of the writing skills of college freshmen and b) identifying college students with learning disabilities related to written expression.

Keywords: curriculum-based measurement, written expression, norms, college students

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEGEMENTS

This dissertation could not have been completed without the efforts of countless individuals and groups of people. First, I would like to express my appreciation to each and every member of my doctoral committee. I am extremely grateful for my dissertation adviser and committee chair, Dr. Cristin Hall, who has shared her professional and personal knowledge and experiences with me and motivated me to expand my skills as both a researcher and practitioner; for Dr. Shirley Woika, who has been a consistent source of support throughout all aspects of my graduate career and who helped me develop my current research goals; for Dr. Katie Hoffmann, who has shared her positivity and desire to continually learn, as well as her CBM expertise with me; and for Dr. Anne Elrod Whitney, who encouraged me to take a multidisciplinary perspective throughout the research process and shared her expertise regarding college writing with me.

Next, I would like to thank my family, who has provided unending love and support to me, in every sense of the word, throughout my graduate experience. They have joined in celebrating my successes, picked me up and encouraged me to persevere when I had my doubts, and always believed in me throughout every endeavor I have ever pursued.

I would additionally like to thank the professors and students from the Penn State College of Education, who allowed me to come into their freshmen writing seminar courses to discuss my research and collect data for this project. Your support for student research makes advancements within the field of education possible.

Lastly, I would like to thank the Association of School Psychologists of Pennsylvania (ASPP) for providing me with a mini-grant to help fund the current research project. I am incredibly grateful for ASPP's devotion to developing research-practitioners within the field.

CHAPTER 1

Problem Statement

Introduction

Throughout the 21st century, there has been increased attention among educational policymakers and practitioners in the United States (U.S.) regarding writing skills due to low writing performance among students (McMaster & Campbell, 2008; McMaster & Espin, 2007). Specifically, according to the National Center for Education Statistics (2003), 86% of Grade 4 students, 85% of Grade 8 students, and 74% of Grade 12 students were below “proficient” in writing in 2002. As a result of this low student performance, the National Commission on Writing issued a report in which it urged professionals within the education field to focus their efforts on improving writing instruction and related student outcomes (McMaster & Campbell, 2008).

It is essential that schools and institutions, from elementary schools to colleges and universities, have a system in place for: a) identifying students at-risk for writing difficulties and b) monitoring these students’ progress, as writing skills can have a lifelong impact. Writing skills not only affect the academic performance of school-age individuals, but they are also related to success in both higher education and the workplace (Huot, 2002; Kellogg & Raulerson, 2007; Troia & Olinghouse, 2013). For example, Geiser and Studley (2002) found that Scholastic Aptitude Test (SAT) writing scores were a strong predictor of college freshmen grade point averages (GPAs). As another example, the National Commission on Writing for America’s Families, Schools, and Colleges (2004) considers writing to be a “threshold skill,” or a critical skill taken into account when employers are hiring or promoting professionals because the majority of salaried occupations in the U.S. include writing responsibilities.

Background of Curriculum-Based Measurement

One way to identify students who are at-risk for not meeting academic standards and to monitor improvement in student academic performance is through curriculum-based assessment (CBA; McMaster & Espin, 2007). CBA is a broad term used to refer to measurement of student performance that is: a) based on the curriculum and b) used to evaluate instruction (Hintze, Christ, & Methe, 2006). CBA can be divided into two categories - specific subskill mastery measurement and general outcome measurement (Hintze et al., 2006). An example of a specific subskill mastery measurement would be if a teacher gave students a math test assessing their skills related to reducing fractions. On the other hand, an example of a general outcome measurement would be a quiz that included a sampling of all of the math skills the students learned throughout the year. One of the most widely used types of general outcome measurements is Curriculum-Based Measurement (CBM; Stecker, Fuchs, & Fuchs, 2005).

CBMs were first developed by Stanley Deno as a way to measure the general academic skills (i.e., reading, math, and written expression) of a student at a certain point in time (Espin, McMaster, Rose, & Wayman, 2012; Hintze et al., 2006). CBMs were developed to be brief, inexpensive, and easy to administer and interpret, while also remaining technically adequate (Espin et al., 2012; Merrigan, 2012; Stecker, et al., 2005). These qualities allow CBMs to be given to students on a frequent basis in order to monitor their progress while taking minimal time away from learning (Espin et al., 2012; Hintze et al., 2006).

Currently, CBM is one of the most widely researched methods of progress monitoring, and research has provided evidence that using CBMs for progress monitoring has led to greater improvements in student performance compared to other methods used to measure student progress (Espin et al., 2012; McMaster & Campbell, 2008; McMaster & Espin, 2007; Stecker et al., 2005). Progress monitoring has several uses, including measuring current academic performance, setting goals, measuring progress toward those goals, and determining when changes need to be made to an individual student's instruction (i.e., when sufficient progress has not been made; McMaster & Campbell, 2008). Throughout this process, educators are able to identify students who are falling behind academically, as well as those who may be eligible for special education and related services (Espin et al., 2012; Fewster & Macmillan, 2002; Stecker et al., 2005).

Since CBM is a general outcome measure, progress monitoring assesses student progress toward long-term goals and objectives rather than individual subskills (Hintze et al., 2006; Stecker et al., 2005). Progress monitoring usually occurs once or twice weekly, and scores are graphed in order to display performance and predict whether students are likely to meet the long-term goals that have been set for them (Stecker et al., 2005). Probes for each grade are similar in difficulty, which then allows each student's rate of learning to be observed each week (Espin et al., 2012; Stecker et al., 2005)

CBM in Written Expression

CBMs have wide utility, as they can be used to measure student performance and progress across multiple academic disciplines. While there is a large research base evaluating the usefulness of CBMs for student outcomes in reading, there is a gap in the research related to math, and even more so for writing (Fewster & Macmillan, 2002; Merrigan, 2012; Runge, Menko, Young, Paulson, McCracken, & Meiss, 2017). CBM for written expression (CBM-WE) was developed at the University of Minnesota's Institute for Research on Learning Disabilities (McMaster & Campbell, 2008). Some advantages of CBM-WE probes are that they are easy to administer and can be administered in an individual or group format (Runge et al., 2017).

A disadvantage of CBM-WE probes, on the other hand, is that they are more difficult and time-consuming to score than reading or math probes (Runge et al., 2017). Further, the technical adequacy of CBM-WE is lower than that of reading; however, research has shown that the overall reliability and validity of CBM-WE is moderate, which is high enough to still be considered acceptable and useful for practitioners (Fewster & MacMillan, 2002; McMaster & Campbell, 2008). For example, Fewster and Macmillan (2002) compared the reading and writing CBM scores of students in Grades 6 and 7 to the same students' year-end grades in Grades 8 to 10. The results of the study indicated that current CBM scores were related to future academic performance, providing evidence for the predictive validity of CBMs.

Although the overall technical adequacy of CBM-WE is moderate, the exact reliability and validity of probes differ based on the age or grade of the student, the amount of time given to complete the task, the type of probe, and the type of scores utilized. Typically, students are given 3 to 7 min to respond because this amount of time has resulted in reliable and valid results (Espin et al., 2012; McMaster & Espin, 2007). In most cases, students are provided with 1 min to plan and think about what they would like to write and 3 min to write (Merrigan, 2012). Probes can come in the form of a passage copying task (students copy letters, words, or sentences), story starter (students are given a sentence that begins a story and they must continue the story), picture prompt (students write letters, words, or sentences in response to a picture), narrative prompt (students write a sequential story in response to a given prompt), or expository prompt (students write an informational or descriptive piece in response to a given prompt; Campbell, Espin, & McMaster, 2012; Hosp, Hosp, & Howell, 2016; Merrigan, 2012). In most cases, narrative prompts are used with younger students, while expository prompts are used with older students (McMaster & Campbell, 2008; Runge et al., 2017).

The types of scores commonly used, from least to most complex, include: Total Words Written (TWW), Words Spelled Correctly (WSC), Correct Word Sequences (CWS), Incorrect Word Sequences (IWS), and Correct minus Incorrect Word Sequences (CIWS). One criticism of these scores is that they only take microskills into account (e.g., spelling, conventions, and grammar), and they do not focus on macroskills (e.g., content and knowledge about the writing process; Runge et al., 2017). In general, TWW and WSC have high reliability at all grades, but they have more validity at the elementary grades (Runge et al., 2017). In contrast, CWS, IWS, and CIWS have adequate reliability at all grades, but they have more validity at the secondary level (Espin et al., 2012; McMaster & Espin, 2007; Runge et al., 2017).

McMaster and Campbell (2008) compared several different CBM-WE tasks that ranged in time allotted, type of task, and type of scores in order to determine which variables led to the highest technical adequacy among students in Grades 3, 5, and 7. The authors found that passage copying yielded reliable scores for students in all grades, but only valid scores for third graders. Across all grades, picture prompts with a 3-min response period and narrative prompts with a 5-min response period produced reliable and valid scores, but only for CWS and CIWS scores. Finally, expository prompts produced sufficient reliability and validity for fifth- and seventh-grade students. Overall, more complex scoring procedures had the strongest technical adequacy, and narrative prompts seemed to have the best technical adequacy across all grades.

In another study examining the reliability and validity of a variety of CBM-WE probes, Merrigan (2012) found that CIWS scores from a narrative prompt, in combination with a 3-min response period, resulted in the highest technical adequacy among middle school students. However, for Grade 10 students, Espin et al. (2008) demonstrated that reliability coefficients are highest with a 7-min response period, compared to 3- or 5-min periods. Reliability did not increase significantly when students were given a 10-min response period. Thus, these various results indicate that using more complex scoring procedures (i.e., CWS, IWS, CIWS) for narrative or expository prompts, along with a 7-min response period is likely to produce the most technically adequate results for older students. Once the scoring procedure, type of probe, and response time is decided upon for a given student, practitioners must then consider how the student's CBM-WE score will be used to assess his or her performance and make decisions.

CBM Norms

One way that practitioners often assess a student's performance is by comparing the student's CBM score to a set of norms. Ricks (1971) defines norms as "percentile or standard score conversions derived from a distribution of scores earned by an identified group" (p. 1). Norms are important because they allow an individual's performance or score on a given assessment to be compared to a relative group, which provides the student's relative status or standing (Elliott & Bretzing, 1980; Ricks, 1971). This gives meaning to assessment scores that might otherwise be seen as an arbitrary number (Ricks, 1971). When CBMs are given to school age students, their performance is most often compared to a national normative sample of same-age or -grade students (Elliott & Bretzing, 1980). For instance, AIMSweb®, a web-based program used by many schools to screen and progress monitor students using CBM, provides fall, winter, and spring national written expression norms for Grades 1 through 8 for WSC, CWS, and TWW (Pearson, 2014). Thus, for any first- through eighth-grade student, a practitioner could determine the exact percentile rank of the student's WSC, CWS, or TWW score for a CBM-WE probe.

Although national norms are used most frequently with CBMs, there are times when schools may instead wish to compare individual student performance to local norms, which can be based on a given classroom, grade, school, or district (Elliott & Bretzing, 1980; Kamphaus, 1984; Ricks, 1971). Certain communities present much differently than the national sample based on their geographic region or socioeconomic status (Kamphaus, 1984). As a result, local norms often look much different than national norms (Ricks, 1971). For example, the average CBM scores of a low-income school in the middle of a large city will likely be much lower than the average scores of students in the U.S. as a whole. Thus, Elliott and Bretzing (1980) posit that local norms are most useful because they compare a given student to a population similar to the student.

CBM-WE for Secondary and Post-Secondary Students

At this time, national CBM norms only exist through Grade 8 (Runge et al., 2017); however, the development of CBM norms for high school and college students would have utility for this age group as well. Just as CBMs are used to monitor progress and hold educators accountable for their students' continued learning at the elementary and middle school level, CBMs have the potential to serve these same valuable purposes, among others, for high school and college students. Several studies have looked at the utility of using CBMs with college students (Bean & Lane, 1990; Hosp, Ford, Huddle, & Hensley, 2017; Hosp, Hensley, Huddle, & Ford, 2014; Larson & Ward, 2006; Lewandowski, Codding, Kleinmann, & Tucker, 2003), but none have used CBM-WE with the overall general college student population.

For instance, Lewandowski et al. (2003) found that there were no current norms for the reading rate of college-aged students. Therefore, the researchers administered reading CBM probes, along with several other measures of reading fluency, to a convenience sample of college students in order to determine their average reading speed. Bean and Lane (1990) also administered reading CBM probes to postsecondary students, while Larson and Ward (2006) administered vocabulary CBMs to this age group. Hosp and colleagues (2014, 2018) additionally administered CBMs to postsecondary students, but they were specifically looking at the use of reading, math, and writing CBMs with students who had been diagnosed with intellectual and developmental disabilities. These studies indicate that reading CBMs have been shown to be useful with the general college population, and reading, writing, and math CBMs have been useful with college students that have developmental disabilities. Thus, it is plausible that CBMs for written expression would have utility with both the general college population, as well as college students with disabilities.

Purpose and Rationale

The purpose of the current study was to create a set of CBM-WE local norms based on freshmen college students at the Pennsylvania State University (Penn State). These norms potentially have two primary uses at the high school and college level: a) to provide a benchmark for the writing skills of Penn State freshmen and b) to help identify Penn State students with learning disabilities.

First, these particular norms can serve as a benchmark of writing skills for students who hope to attend Penn State or other similar schools. High school students who wish to attend Penn State or another comparable college or university can compare their current writing skills to those of freshmen at Penn State. If the student’s written expression skills are below those of the typical college freshman at Penn State, a goal can be set for the student to work up to a score that reaches at least the 50th percentile, for example. CBM-WE probes can then continue to be used to monitor the student's progress toward this goal.

Second, having access to writing norms for college students may be helpful to school psychologists and school psychology clinics, such as Penn State's Edwin L. Herr clinic, in the process of identifying learning disabilities in written expression among college students. In fact, there has been an increase in the number of college students with disabilities, with learning disabilities being the most common (Lewandowski et al., 2003). CBM-WE scores can be used as a measure of writing skills in addition to traditional achievement batteries (e.g., Woodcock-Johnson Tests of Achievement, Wechsler Individual Achievement Test), as it is a best practice within writing assessment to use various methods to collect data on a student's current academic skills (McMaster & Espin, 2007; Runge et al., 2017; Troia & Olinghouse, 2013).

CHAPTER 2

Literature Review

In a report about higher education, the U.S. Department of Education (2006) asserted, "Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a 'value-added' basis that takes into account students' academic baseline when assessing their results" (pp. 4). A standardized measurement system that can measure broad areas of academic competence and monitor improvements in educational performance would likely improve the accountability of colleges and universities for their students' success (U.S. Department of Education, 2006). Curriculum-based measurement is a model used to assess individual achievement and monitor growth in performance among school-age students. It is feasible that the CBM model could have similar utility for students and educators at colleges and universities as well (Hosp et al., 2014). Further, CBMs could provide a way to keep colleges and universities accountable for the success of their students, especially those with disabilities (Hosp et al. 2016).

The present literature review will outline research related to curriculum-based measurement, a form of curriculum-based assessment. The characteristics, benefits, and uses of general CBM will be summarized, followed by a review of CBM in written expression. General writing development and assessment will be discussed, as well as the research surrounding CBM-WE by student level (i.e., elementary, middle school, high school, and post-secondary). The literature review will then conclude with the purpose and rationale of the current study.

Curriculum-Based Assessment

Curriculum-based assessment is a dynamic, ongoing method of assessment used to collect information about students who are struggling academically and make instructional decisions based on direct observations of student performance (Gickling & Rosenfield, 1995; Hintze et al., 2006; Shapiro, 2011; Shinn, 1989). The term "curriculum-based" indicates that assessment materials are created directly from a student's local instructional curriculum (Deno, 2003; Elliott & Fuchs, 1997; Shinn, 1989). Although CBA has a variety of specific uses within special education (e.g., referral, screening, classification, instructional planning, progress monitoring, etc.), its most important function, in general, is providing data to enhance instruction and improve student learning (Gickling & Rosenfield, 1995).

CBA can be segmented into two broad categories: a) specific subskill mastery measurement (i.e., measures mastery of individual subskills) and b) general outcome measurements (i.e., measures proficiency of multiple skills within a curriculum area; Hintze et al., 2006). One disadvantage of subskill mastery measurement is that some students will demonstrate adequate performance on a single subskill, but they will be underachieving on a broader level (Espin et al., 2012). General outcome measures, on the other hand, more accurately reflect broad academic competence (Espin et al., 2012). The current paper will focus on a widely used general outcome measurement model known as curriculum-based measurement. Although CBM has many uses, each of which will be discussed subsequently, the current study will focus on its use as a measure of students' current academic proficiency, as well as its use in the development of local norms for students at the college level.

Curriculum-Based Measurement

Curriculum-based measurement stems from data-based program modification (DBPM), an older special education intervention model that used repeated measurement of student performance to assess and improve the effectiveness of instruction and interventions (Deno, 2003; Espin, 2012). In order to evaluate the effectiveness of DBPM, Stanley Deno, along with his colleagues at the University of Minnesota's Institute for Research on Learning Disabilities, developed and tested a set of progress monitoring procedures in the areas of reading, spelling, and written expression from 1977 to 1983 (Deno, 1985; Deno, 2003; Hintze et al., 2006). These progress monitoring procedures were developed into the current conception of curriculum-based measurement (Espin et al., 2012).

Characteristics of CBM

CBM is a reliable and valid tool routinely used by special education and general education practitioners to: a) evaluate and monitor individual student academic competence and growth within the school curriculum and b) make decisions about whether to modify a given student's current instruction or educational program based on the amount of progress (Deno, 1985; Espin, 2012; Fuchs, Fuchs, & Hamlett, 2015; Marston & Magnusson, 1985; McMaster & Campbell, 2008; Shinn, 1989; Wright, 1992). CBMs are considered a form of direct behavioral observation that involves single case design procedures to monitor and respond to individual student progress in the areas of reading, math, written expression, spelling, and vocabulary (Deno, 2003; Hosp et al., 2016; Shinn, 1998). As mentioned, CBMs are a form of general outcome measurement because in most tasks, students are required to demonstrate and integrate a variety of different skills (Elliott & Fuchs, 1997; Fuchs, 2004; Stecker et al., 2005). CBMs are intended to be aligned closely with students' curriculum and instruction (Elliott & Fuchs, 1997; Hosp et al, 2016; Shinn, 1989; Wright, 1992).

CBMs have several distinguishing characteristics. They must be technically adequate; have standardized measurement tasks, materials, directions, and scoring procedures; have decision rules for performance criteria; sample performance through correct and incorrect responses; use a timing device; have multiple probes for a given skill that are equivalent in difficulty and content; be recorded through charts or forms; be time efficient; and be easy to teach and learn (Deno, 2003; Espin, 2012; Hosp et al, 2016; Shinn, 1998; Wright, 1992). Because measures are similar in content, difficulty, and format, a student's performance can be compared from one point in time to another (Stecker, 2005). The measures are scored based on the student's speed, fluency, or accuracy, as these metrics help educators determine a student's overall proficiency and whether the student has mastered a given skill (Larson & Ward, 2006; Wright, 1992).

Alternative to Traditional Assessment

CBM has now been established as a feasible alternative method to assessing special and general education students' overall academic competence or proficiency at a given point in time as compared to the widely accepted standardized achievement tests (Deno, 1985; Elliott & Fuchs, 1997; Espin et al., 2012; Fuchs & Fuchs, 2002; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Hosp et al., 2016; Marston & Magnusson, 1985). Traditional achievement tests have been criticized over time for a variety of reasons. For instance, critics assert that standardized achievement tests are unrelated to content from students' curriculum (Deno, 1985; Wright, 1992). As a result, educators may conclude, based on achievement testing, that a student is not making progress in a given academic area, but the test may not be measuring the skills that the student has been learning in class (Shinn, 1989; Wright, 1992). In contrast, CBMs are directly tied to the curriculum, making them more instructionally sensitive and more efficient in improving student learning and instruction than traditional achievement measures (Elliott & Fuchs, 1997; Gansle et al., 2002; Hosp et al., 2016; Shinn, 1989; Shinn & Marston, 1985).

Standardized achievement tests also tend to be norm-referenced, and as such, cannot provide local norm data (Wright, 1992). As a result, standardized tests can be biased for certain groups of students (e.g., minorities) that do not closely match the demographic of the national norm sample for the test (Elliott & Fuchs, 1997; Shinn & Marston, 1985; Wright, 1992). Comparatively, CBMs are curriculum-referenced and allow for the development of local norms at the class, grade, school, or district level (Deno 1985, 2003; Shinn, 1989; Wright, 1992). Local norms are often useful in school districts where the student demographic may not match that of the overall student population in the U.S. (e.g., urban school districts; Bracken, 2007; Deno, 2003; Elliott & Bretzing, 1980; Kamphaus, 1984). Therefore, CBM local norms may help decrease the likelihood of bias in assessment (Habadank, 1995).

Due to the high cost, long administration time, lack of multiple forms, and thus, infrequent administration, traditional standardized achievement tests are additionally criticized because they do not measure learning rate and short-term academic growth (Elliott & Fuchs, 1997; Shinn, 1989; Wright, 1992). Consequently, critics assert that the tests are technically inadequate for making educational decisions about individual students (Deno, 1985). On the other hand, the simplicity, short administration time, and inexpensive materials associated with CBMs allow educators to use them more frequently to monitor student progress compared to more traditional assessment methods (Deno, 1985; Fuchs & Fuchs, 1997; Gansle et al., 2002; Marston, Tindal, & Deno, 1982; Shinn, 1989; Shinn & Marston, 1985). Teachers are able to use the measures with little to no training, and minimal instructional time is taken away from both students and teachers (Espin et al., 2012).

Repeated and frequent measurement makes CBM more sensitive to short-term growth in student academic performance compared to standardized achievement tests (Deno, 1985; Elliott & Fuchs, 1997; Gansle et al., 2002; Hosp et al. 2016; Wright, 1992). As a result, CBMs have been used to establish standards for academic growth among students in both general and special education (Deno, Fuchs, Marston, & Shinn, 2001). Further, because the probes in a given academic area are similar in content and difficulty, rate of learning can be calculated (Fuchs, 2004; Hosp et al., 2016). In addition, scores can be easily communicated to and understood by parents, teachers, and students because very little interpretation is needed (Deno, 1985, 2003; Espin et al., 2012; Hosp et al., 2016). This also makes consensus among these individuals more likely in regard to scoring and decision-making (Espin et al., 2012).

Current Uses of CBM

In general, CBM is a reliable and valid method of measurement that provides data that can be used within the problem-solving model to guide various special education decisions, including decisions about appropriate planning, instruction, interventions, eligibility, and placement (Deno & Fuchs, 1987; Deno, Martson, & Tindal, 1986; Fuchs, et al., 2015; Malecki & Jewell, 2003; Shinn, 2002; Shinn & Hubbard, 1992). Decisions using CBM data can be made using three different approaches. If the goal of assessment is to compare an individual's performance to previous performance or expected levels of performance, an *individually-referenced approach* can be used. If the goal is instead to compare student performance to an agreed upon standard, a *criterion-referenced approach* can be used. Lastly, a *normative or peer-referenced approach* can be used if the goal of assessment is to compare a student's current performance to the performance of peers (Hosp et al., 2016; Shinn, 2002).

Progress monitoring. The main way that CBM data are collected for decision-making purposes is through monitoring student progress toward long-term goals within the curriculum (Fuchs et al., 2015; Shinn, 1989). According to Hosp et al. (2016), "CBM is the best measurement system currently available to monitor students' progress toward long-term goals" (pp. 30), and it is also the progress monitoring approach with the strongest research base (McMaster & Espin, 2007). Progress monitoring involves: establishing a student's present or baseline performance, setting an academic goal, and keeping track of the student's progress toward that goal on a daily to monthly basis through repeated administration of CBM probes (Deno, 2003; McMaster & Campbell, 2008; Stecker, 2005). Repeated, continuous measurement data allows educators to determine whether students are demonstrating growth in their academic skills, remaining static, or even regressing (Deno, 1985). The data can then be used to formatively evaluate whether a student's current instruction is effective or whether changes or modifications need to be made to the student's educational program so that adequate progress toward goals can be made (Deno, 2003; Espin, 2012; Hosp et al., 2016; Shinn, 1998; Stecker et al., 2005). Further, the various instructional strategies and interventions implemented with a given student can be compared to determine which options have resulted in the greatest academic growth for that student (Fuchs & Fuchs, 2002; Fuchs et al., 2015).

CBMs are especially useful for progress monitoring because they measure student academic behavior using simple numbers, which allows the information to be charted, graphed, or put into a computerized data-management system (Espin, 2012; Hosp et al., 2016; Wright, 1992). As a result, parents, students, and teachers are able to easily visualize and interpret a student's rate of academic progress and level of responsiveness to the current educational program (Espin, 2012; Fuchs, 2004; Hosp et al., 2016). In addition, because CBMs measure clearly defined behaviors, the resulting data can easily be used to write outcome-based goals and evaluate progress toward goals, especially those included in the Individualized Education Programs of special education students (Espin, 2012; Hosp et al., 2016; Shinn, 1989; Wright, 1992).

Effects on instruction and achievement. Progress monitoring has been shown to improve teaching practices, as well as student outcomes (Shinn, 1989). Educators can use student progress monitoring data to improve the effectiveness of instructional planning and decision-making by identifying more accurate and frequent student goals and by individualizing instruction (Deno, 2003; Fuchs & Fuchs, 2002; Fuchs et al., 2015; Hosp et al, 2016; Marston & Magnusson, 1985; Stecker, 2005). Further, progress monitoring data helps teachers decide what skills they need to teach or review and the instructional methods they should use to obtain the best student outcomes (Hosp et al, 2016). Fuchs, Deno, and Mirkin (1984) compared special education teachers using CBM to evaluate student learning and special education teachers using more traditional methods such as teacher made tests and informal observations. Teachers who used CBM were more realistic about whether students would meet their goals, and they more frequently revised student goals based on progress monitoring data. Teachers using CBM were found to increase the structure within their instruction as well. In a similar study, Fuchs, Fuchs, and Stecker (1989) found that teachers using CBM used more specific and acceptable goals with their students, used data more frequently to determine if students were making progress and if instruction needed to be modified, and subsequently modified student programs more often compared to a control group of teachers.

Teachers have been able to use CBM to improve instructional planning and educational programming through the use of progress monitoring, which has then also resulted in significant improvements to student learning and achievement outcomes (Deno, 2003; Elliott & Fuchs, 1997; Fuchs & Fuchs, 2002; McMaster & Campbell, 2008; Shinn, 2002; Stecker & Fuchs, 2000; Stecker et al., 2005). However, Stecker and Fuchs (2000) have demonstrated that it is important that instructional modifications are specific to a given student's performance data. Simply measuring the performance of students with disabilities did not improve learning; in contrast, when instruction was adjusted based on individual CBM data, the students benefited from improved academic performance.

In a review of research related to using CBM data to improve student achievement, Stecker et al. (2005) concluded that teachers who used individual student CBM progress monitoring data to guide instruction had greater improvements in student academic performance compared to teachers who used their own method to monitor progress. Fuchs et al. (1984) also looked at the effect of repeated CBM on student achievement, as well as the effects on student self-awareness. Results indicated that students in the CBM classes had higher reading achievement compared to the control classrooms. Students also had more awareness regarding their learning (i.e., they reported that they knew their goals more often, stated their goals more often, had more accurate ideas about whether or not they would meet their goals, and more frequently stated that they used data to decide whether they believed they would meet their goals).

Special education identification. In addition to providing an alternative method of assessment, CBM has provided a viable alternative to the traditional identification process, including screening, referrals, placement, and program planning since the 1980s (Fewster & Macmillan, 2002; Fuchs & Fuchs, 1997; Marston & Magnusson, 1985; Marston, Mirkin, & Deno, 1984; Marston et al., 1982; Shinn, 1989). Similar to the evaluation of general classroom instruction, CBMs can evaluate pre-referral interventions to determine whether they are effective for given students (Deno, 2003). Elliott and Fuchs (1997) posit that CBM data can be used to strengthen the overall pre-referral assessment process by helping educators determine whether a student can benefit from instructional adaptations within general education or if special education services may be necessary.

For example, if progress monitoring data is showing that a given student is not responding to pre-referral interventions, it may be evident that the student may need supports outside of the general education curriculum. Once the student is referred for assessment, CBM progress monitoring data can also be used to determine whether a learning disability exists. For example, using the dual-discrepancy model, students may meet criteria for a learning disability if progress monitoring data demonstrates both low performance and inadequate academic growth in response to interventions (Fuchs & Fuchs, 2002; Stecker et al., 2005).

Research has shown that the CBM method identifies students that are very similar to the demographic of students diagnosed by traditional learning disability identification models (Fuchs & Fuchs, 1997). In a study by Marston et al. (1984), the traditional teacher referral procedure was compared to the CBM method, in which referrals are based on reading, spelling, and written expression progress monitoring data. The two referral methods resulted in similar numbers of referred students. Further, the students from both groups were equally likely to display an aptitude-achievement discrepancy, and the students from both groups demonstrated equally low performance on cognitive and achievement measures.

Further, CBM literature indicates that the measures are sensitive to differences within individuals, as well as between groups of students (Shinn, 1998). For instance, Shinn and Marston (1985) gave mildly handicapped (i.e., students with learning and intellectual disabilities who received less than half of their school day in special education), low-achieving (i.e., students receiving Title I services), and general education students CBMs to determine whether these groups performed significantly different from one another. The students completed CBMs in reading, spelling, math, and written expression. The authors determined that handicapped students performed significantly lower than low-achieving students, and low achieving-students performed significantly lower than general education students in all four academic areas.

Universal screening. CBMs have additionally been used within multi-tiered systems of support to provide the data needed to make special education eligibility decisions (Elliott & Fuchs, 1997; Hosp et al., 2016; Merrigan, 2012; Stecker et al., 2005). For example, CBMs can serve as a universal screener to identify students who are at-risk for future academic difficulty and are in need of additional support (Deno, 2003; Espin et al., 2012; Hosp et al, 2016). Typically, educators will assess all students within a classroom, grade, school, or district; choose a designated cut score; identify those students who scored below the cut score; and progress monitor these "at-risk" students (Deno, 2003). Universal screening through the use of CBMs can also help schools formatively evaluate whether their general education curriculum is meeting the needs of a majority of the students (Hosp et al, 2016).

Development of norms. In addition to allowing for comparisons of a student's current performance to individual past performance or to a specific criterion, CBM also allows for norm, or peer, referencing (i.e., an individual's performance is compared to same-age or grade peers; Deno, 1985, 2003; Shinn, 1988). Norms are standards (i.e., percentile ranks or standard scores) calculated based a distribution of raw scores among a given group (Elliott & Bretzing, 1980; Ricks, 1971). The scores can be used to describe a population's performance on a given task (e.g., CBM-WE) or compare an individual's performance to the group from which the norms came (Elliott & Bretzing, 1980; Habadank, 1995; Ricks, 1971). Students can be compared to local class, grade, school, or district norms, as well as national or research norms for a given stimulus material, depending on the purpose of assessment as well as the availability of such norms (Deno 1985, 2003; Wright, 1992).

Typically, local norms are developed to determine the relative standing of an individual at a class, grade, school, or district level, as these norms often differ substantially from national norms (Kemphaus, 1984; Ricks, 1971). Shinn (1989) and Habadank (1995) suggest that local norms may actually be more appropriate than national norms for some assessment purposes (e.g., measuring individual progress over time or comparing a student's achievement in different academic areas) because local norms are curriculum-referenced, meaning they allow for measurement of student proficiency within a specific class, school, or district curriculum (Shinn, 1988). Similar to national CBM norms, local norms can also be used to make important special education decisions regarding screening, eligibility, and IEP planning (Shinn, 1988, 1989). Further, local norms can help evaluate the effects of a given program, curriculum, or instructional strategy on a student's performance; define academic expectations for students; and set goals for progress (Habadank, 1995; Hosp et al., 2016). For example, scores at the 50th percentile for a student's grade are often used as year-end goals (Hosp et al., 2016).

In order to develop local norms, a specific measure (e.g., a CBM-WE probe) is administered to a sample of students from a given group (Bracken, 2007). According to Parker et al. (1991a), large samples of 100 to 150 students should be obtained if norms are to have stability. Kamphaus (1984), Shinn (1988), and Habadank (1995) similarly assert that a sample of 100 or more students per group (e.g., grade) is sufficient, as this would allow for stable, whole number percentile ranks. After administering the measure, each student's raw score is established, and the raw scores are used to create a frequency distribution. Means, standard deviations, percentile ranks, standard scores, and interquartile ranges can then be calculated using the scores (Bracken, 2007;

Kamphaus, 1984; Shinn, 1988). If the students are randomly sampled from the group, the scores are understood to represent the larger group from which the sample came (Hosp et al, 2016).

Technical Adequacy

Overall, research on CBM procedures has produced moderate technical adequacy (McMaster & Campbell, 2008). CBM reliability is most often demonstrated through correlations between two different raters' scores for a given probe (i.e., interrater reliability), correlations between a given student's scores on the same measure at two different time periods (i.e., test-retest reliability), and correlations between a given student's scores on two different versions of the same measure (i.e., alternate forms reliability; Espin et al., 2000; Merrigan, 2012). Validity research, alternatively, often focuses on the relationship between CBM scores and the scores of other measures of achievement (i.e., criterion validity; Espin, Scierka, Skare, & Halverson, 1999; Espin et al., 2000) and the ability of CBM scores to predict student outcomes (i.e., predictive validity). An example of predictive validity is the ability of CBM scores to predict success on high-stakes assessments, as well as success in early childhood (Deno, 2003).

Written Expression

As mentioned, written expression is an area of achievement that can be measured using CBM. It is essential that educators assess and monitor students' written communication skills, as they are strongly linked to success in both education and life outside of academic settings (e.g., psychosocial, adaptive, vocational, and economic success; Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Hosp et al., 2016; Troia & Olinghouse, 2013). For instance, written skills are included on important assessments such as graduation tests, standardized state tests of achievement, and college entrance evaluations (Diercks-Gransee, Weissenburger, Johnson, &

Christensen, 2009; O'Neill, Moore, & Huot, 2009; Troia & Olinghouse, 2013). Unfortunately, however, many students are not meeting basic proficiency levels on standardized writing assessments. For example, the National Center for Education Statistics (2003) reported that approximately three-quarters of Grade 12 students were below "proficient" in writing in 2002. Thus, finding ways to help educators identify, monitor, and support students who are struggling in writing is imperative if students are to meet national standards (McMaster & Espin, 2007).

Development of Writing Skills

Learning to write is a complex process, as writing proficiency is made up of a variety of skills including, but not limited to: spelling, grammar, punctuation, handwriting, typing, mechanics, motor skills, sentence construction, planning, revising, creativity, organization, and expressiveness (Baker & Hubbard, 1995; Berninger, et al., 2006; Graham, McKeown, Kihuara, & Harris; 2012; Olinghouse & Leaird, 2009; Shapiro, 2011). Students tend to demonstrate difficulty with writing at an early age, and they often continue to struggle with written expression skills throughout their educational careers (Baker & Hubbard, 1995). Writing is a crucial component of students' literacy, and early written expression skills are able to predict general academic success in school (Baker & Hubbard, 1995; Gansle et al., 2006).

Elementary students begin by learning transcription skills (i.e., handwriting fluency, spelling, and grammar), as these skills are able to predict composition length and quality in later elementary school through college (Berninger et al., 2006; Olinghouse, 2008). Transcription skills must become automatic if other important, higher-level aspects of writing (e.g., organization, planning, revising, expression, etc.) are to be subsequently mastered (Berninger et al., 2006; Olinghouse, 2008). In addition, children in early elementary school typically read and write narrative style texts, as students have more success comprehending narrative texts, rather than expository, at this age (Best, Floyd, & McNamara, 2008; Kulikowich, Mason, & Brown, 2008). Although narrative writing may be less demanding than expository writing, Olinghouse and Leaird (2009) posit that narrative writing is still a complex task that requires a variety of writing skills such as language, vocabulary, and mechanics, as well as an understanding of written elements such as story components, conventions, audience perspective, and abstract concepts.

In upper elementary and secondary school, in contrast, the focus of writing shifts to understanding, learning new information, and solving problems within various subject areas (Best et al., 2008; Kulikowich et al., 2008). Thus, students at this level tend to be required to read more difficult, expository texts (Best et al., 2008). Expository texts require more complex skills than narrative texts, including prior knowledge of a subject area, metacognitive skills, and the ability to self-regulate one's own writing (Kulikowich et al., 2008).

Several meta-analyses have examined writing strategies and instructional practices to determine which are most effective in improving school-aged students' writing (Graham et al., 2012; Rogers & Graham, 2008). In general, Olinghouse (2008) asserts that instruction should incorporate both basic and high level writing skills. One important evidence-based practice is explicit or direct instruction in writing (Graham et al., 2012; Rogers & Graham, 2008). This includes teaching students about planning, drafting, and revising different types of text; regulating writing strategies; creativity; structure and format of various types of text; spelling; handwriting; grammar; word usage; constructing complex sentences; and keyboarding. Another evidence-based practice related to writing is scaffolding students' learning (Graham et al., 2012). Teachers can do this by having students work together to plan, draft, revise, and edit written work; setting clear, specific goals; helping students learn to organize thoughts before writing; and assessing and monitoring learning and performance. Additional best practices include allowing students to use word processors, increasing the amount that students write in order to increase productivity, reinforcing students' writing productivity, and ensuring that writing instruction is comprehensive (Graham et al., 2012; Rogers & Graham, 2008).

As mentioned, it is common for students with learning disabilities to have difficulty with writing (Shapiro, 2011). They often do not use effective strategies when completing tasks related to written expression and when subsequently revising their written work. Similar to instructional strategies for general education students, strategies for school-aged students with learning disabilities tend to focus on self-monitoring of writing productivity, prewriting skills (i.e., planning), writing sentences, generating content for various types of writing, and editing previous work. Furthermore, monitoring the progress of these struggling writers is imperative.

Current options for college students that are in need of writing improvement include mainstreaming students with additional support (e.g., writing centers, tutoring/tutorials, consultation, one-on-one work with professors), developmental writing programs, and additional remediation courses (Hassel & Giordano, 2015; O'Neill et al., 2009). The majority of colleges and universities provide remedial education courses for students who are struggling with college-level academic skills (e.g., writing). However, the courses may also be termed developmental education, skills courses, and college preparation courses (Attewell, Lavin, Domina, & Levey, 2006).

Relles and Tierney (2013) explain that the focus of remedial courses and instruction at the college level depends on the given instructor's or institution's theoretical orientation. In a current-traditional orientation (i.e., concerned with skills such as argumentation, organization, sentence structure, grammar, usage, and mechanics), instruction focuses on skill repetition through instructor comments regarding student mistakes. In a cognitive orientation (i.e., concerned with overall writing process), instruction focuses on students' understanding of the planning, drafting, and revising processes. In an expressionist orientation (i.e., concerned with the writer's personal experience), instruction focuses on developing students' voice and self-expression. Lastly, in an epistemic orientation (i.e., concerned with discourse norms), instruction focuses on helping understand the discourse norms within academia, as well as those outside of educational settings.

Assessment of Writing Skills

Although more research is needed in the area of assessment for writing skills, general best practices are to develop local measures with a specific context and purpose (Huot, 2002), collect a variety of information about student skills through different types of writing measures (Baker & Hubbard, 1995; Runge et al., 2017; Troia & Olinghouse, 2013), and continually monitor student progress so that instructional programs and student learning can be improved (Calfee & Miller, 2013; Conference on College Composition & Communication Committee on Assessment, 2014; O'Neill et al., 2009). In addition, direct methods of writing (i.e., those that require students to produce a written piece) are viewed as more informative in the educational assessment and college composition community, compared to indirect methods (e.g., a multiple choice exam; Baker & Hubbard, 1995; O'Neill et al., 2009). Students should also be involved in the assessment of their written work if their writing proficiency is to improve (Huot, 2002). Further, when choosing an assessment method, practitioners should always consider the decisions that will be made as a result of the assessment (Baker & Hubbard, 1995).

Writing assessment can help schools ensure that all students are being provided with educational experiences related to writing that result in positive outcomes (Huot, 2002). Several important uses of writing assessment at the college level include assessing student proficiency to make placement decisions; evaluating the effectiveness of an institution, department, or program; and enhancing the teaching and learning of writing skills (CCCC Committee on Assessment, 2014; Huot, 2002; O'Neill et al., 2009). In fact, some theorists assert that a writing assessment can only be valid if it is used to make decisions that inform teaching, improve learning, and thus, positively affect the educational environment (Huot, 2002; O'Neill, 2009).

CBM for Written Expression

Similar to the general uses of CBM, CBM-WE is a direct measure of written expression that can be used to make special education referral, screening, placement, and planning decisions, especially in regard to whether a student meets criteria for a learning disability (Fewster & Macmillan, 2002; McMaster & Espin, 2007; Runge et al., 2017; Watkinson & Lee, 1992). CBM-WE is also an efficient method of assessing and monitoring writing instruction and performance, especially for low achieving students and students in special education (Espin et al., 2008; Hosp et al., 2016; McMaster & Espin, 2007; Watkinson & Lee, 1992). Although CBM-WE is a feasible method for measuring student writing achievement and progress, research is severely lacking compared to the available research in reading and math (Fewster & Macmillan, 2002; Merrigan, 2012; Runge et al., 2017). For example, Gansle et al. (2002) searched the PsychINFO database for all CBM research. There were 88 studies available for reading, 35 available for math, and only 8 available for writing. Previous research also demonstrates lower technical adequacy for CBM-WE compared to reading; however, the measures still have definite utility (Fewster & Macmillan, 2002).

One reason for the discrepancy in CBM research may be that CBM-WE is difficult to measure and assess (Espin et al., 2004). Writing does not have the same clear-cut right and wrong answers that reading and math often do (Gansle et al., 2002). In addition, writing proficiency is defined by different skills depending on a student's age and grade (Merrigan, 2012). Therefore, the features of CBM-WE measures (i.e., length, format, scores) need to increase in complexity as students' writing proficiency increases if the measures are to remain useful for older students (Espin et al., 2000; Weissenburger & Espin, 2005). Further, scoring CBM-WE takes more training and time than other CBM measures (Runge et al., 2017). Despite the measurement and scoring challenges, there is a need for research related to CBM-WE and its technical adequacy (Jewell & Malecki, 2005), especially for older students.

Tasks. In general, CBM-WE tasks are short and simple measures of overall skill in written expression, and the measures can easily be administered in both individual and group settings (Hosp et al., 2016; Runge et al., 2017). Research has shown that the technical adequacy of various CBM-WE features (e.g., type of task, response length, scores) varies depending on a given student's developmental level, gender, and age. For instance, students are usually given 3 to 5 min to respond to CBM prompts, as that time range has generally produced the strongest validity (McMaster & Espin, 2007). However, in some situations (e.g., for older students), a 3-min response period does not give sufficient information; thus, students can be given longer response periods of 5 to 10 min (Hosp et al., 2016). In general, when choosing what type of CBM-WE features to use with a student, educators should consider which are most valid and reliable for the intended purpose (e.g., progress monitoring, measuring proficiency, predicting performance) and for the student's individual characteristics (Jewell & Malecki, 2005).

The most typical form of CBM-WE is a story-starter, in which students are given a short sentence to help begin the writing process (Hosp et al., 2016). Other CBM-WE tasks include topic sentences; copying letters, words, and sentences; writing words that begin with a given letter; and picture-word, picture-theme, and photo prompts (Hosp et al., 2016; Merrigan, 2012). Each of these tasks requires students to write letters, words, sentences, or multiple sentences (Hosp et al., 2016). In prompts requiring multiple written sentences, narrative prompts have been used most commonly, although research is demonstrating that expository prompts have more technical adequacy with students at the secondary level (McMaster & Campbell, 2008; Merrigan, 2012).

Scoring. Research has shown that specific skills such as words written or words spelled correctly, for example, are related to students' overall skills in written expression (Hosp et al., 2016). In early research, the most commonly used types of scores for CBM-WE included total words written (TWW), words spelled correct (WSC), and correct word sequences (CWS; Hosp et al., 2016). Additional scores that are now used include correct sentences written (CSW), incorrect word sequences (IWS), correct minus incorrect word sequences (CIWS), total correct punctuation, and various qualitative metrics (Hosp et al., 2016). In general, scoring time is brief; however, it usually increases with grade level, as students are able to write longer passages (Hosp et al., 2016). Interscorer reliability tends to be strong for CBM-WE, test-retest reliability tends to be moderate to relatively strong, and internal consistency and alternate form reliability are usually moderate (McMaster & Espin, 2007).

Based on a review of literature, Runge et al. (2017) posit that TWW and WSC demonstrate high reliability across grades, but they are most valid for elementary students. In contrast, CWS, IWS, and CIWS demonstrate moderate reliability across grades, but they are most valid for middle and high school students. CIWS also has the highest validity with criterion measures such as the Woodcock Johnson Tests of Achievement and standardized state assessments. According to Espin et al. (2000), the CWS, IWS, and CIWS indices are better and more appropriate indicators of written performance for older students because they take into account more complex skills such as semantics and syntax, in addition to early writing skills such as spelling and punctuation. However, Hosp et al. (2016) assert that each of the CBM-WE metrics assesses written expression in a different way, and they are most useful when evaluated together.

Despite the variety of indices available, CBM-WE tasks have been criticized for focusing on word- and sentence-level microskills and for failing to account for more global macroskills (Espin et al., 1999; Merrigan, 2012; Runge, 2017). For instance, typical scores such as TWW, WSC, CWS, IWS, and CIWS are based solely on spelling, grammar, and mechanics, rather than incorporating elements such as planning, content, coherence, and organization. While the most common CBM-WE indices do focus on microskills, Shapiro (2011) argues that it is also possible to examine broader skills such as the creativity and story structure of a student's response through CBM. It is likely that these types of scores are not often used because they are more subjective, which can lead to lower reliability. Further, CBM-WE norms for these areas of writing are limited. However, Shapiro notes that they can still be used to make suggestions and set goals for students' writing skills.

Although the limited scope of CBM-WE scores is a valid concern, research has shown that the strengths of CBM outweigh the current limitations. CBM-WE has successfully been used to measure achievement, monitor progress, and make important instructional decisions for a variety of age groups, from elementary to secondary students, as well as students both with and without disabilities. In addition to measuring current writing proficiency, CBM-WE measures are able to predict overall achievement in writing, as well as course grades and placement decisions (Fewster & Macmillan, 2002; Merrigan, 2012). CBM-WE has also been able to differentiate between students at different levels of proficiency in writing (Fewster & Macmillan, 2002), and it is sensitive to growth within and across grade levels (Parker, Tindal, & Hasbrouck, 1991a). Lastly, it is not essential for CBM-WE to measure every area of writing proficiency, as long as relationships between these brief measures and broader writing skills are demonstrated (Espin et al., 1999; Shinn, 1998). Research has in fact demonstrated that CBM-WE is related to more generalized measures of writing (e.g., holistic writing scores, standardized achievement tests in writing, etc.) through evidence of criterion validity.

Research at elementary level. Research at the elementary level for CBM-WE has supported the notion that simple, direct written expression measures can be repeatedly administered in order to reliably and validly evaluate student performance and educational programs and improve students' writing skills (Deno, Marston, & Mirkin, 1982; Espin et al., 2004). CBM-WE data, in addition to traditional standardized tests, can be used to make screening and identification decisions for elementary students (Marston et al., 1982). Various studies have examined the multiple facets of CBM-WE (i.e., type of task, duration, type of scores) for elementary school students in order to determine which features are the most appropriate at each grade level.

For instance, Lembke, Deno, & Hall (2003) and McMaster et al. (2009) have examined which CBM tasks are the most technically adequate for beginning writers in early elementary school. Lembke and colleagues (2003) gave several CBM-WE measures to 15 (8 females and 7 males) second-grade students in a summer school program for students at-risk for academic difficulty. The school was located in the suburbs of a large Midwestern metropolitan area of the U.S. Half of the students were White, and 52% received free or reduced lunch. In addition, six students received special education services (i.e., one learning disability, three mild/moderate mental impairments, and two speech/language difficulties). Students completed word copying, sentence copying, word dictation, and sentence dictation tasks at the beginning and end of the summer session. The tasks were scored for TWW, WSC, CLS, CWS, and CIWS. Results indicated that the diction tasks had the majority of the strong, statistically significant correlations with criterion measures (i.e., 14 out of 24 correlations were significant for word dictation and 17 out of

32 were significant for sentence dictation). Thus, based on correlations between CBM-WE indices and various atomistic and holistic criterion variables, the study indicated that word and sentence diction measures had the most promise as indicators of early writing proficiency.

McMaster et al. (2009) also examined the use of CBM-WE with beginning writers. The authors gave word copying, sentence copying, story prompts, letter prompts, picture-word prompts, picture-theme prompts, and photo prompts to 100 students in Grade 1. Participants came from four elementary schools in a large metropolitan school district in the Midwestern U.S. In terms of demographic sample characteristics, 25.8% of the students were from culturally or linguistically diverse backgrounds, 17.6% were receiving free or reduced lunch, 5.1% were involved in special education, and 7.8% were receiving ELL services. Student responses were scored for TWW, WSC, CWS, CIWS, and CLS. Based on test-retest and alternate-form reliability coefficients, as well as criterion-related validity coefficients (teacher ratings, a district rubric, and the Test of Written Language), the authors concluded that sentence copying, story prompts, picture-word, and photo prompts are additional promising CBM-WE measures for beginning-level writers. Reliable and valid scores were more consistently found by McMaster and colleagues when students were given a 5-minute response period compared to a 3-min period. Further, WSC, CWS, and CLS produced the most reliability across the different tasks. Lastly, student scores on the measures improved over a three-month period, demonstrating CBM's sensitivity to short-term growth for this age group.

A variety of studies have demonstrated the technical adequacy of TWW, WSC, CWS, CLS, and CIWS for students across the elementary grades. In three research studies by Deno, Mirkin, and Marston (1980), Learning Disabled (LD) and general education students in third to sixth grades from elementary schools in Minnesota were given 5 min to respond to story starters, picture stimuli, and topic sentences. The students were scored based on T-unit (minimal terminable unit) length, TWW, mature words, WSC, and correct letter sequences (CLS). The relationships between these scores and criterion measures (i.e., Test of Written Language, Word Usage Subtest on Stanford Achievement Test, Developmental Scoring System, and program placement) were examined as well. TWW, mature words, WSC, and CLS were the most valid measures of written expression, as measured by correlations with criterion measures.

Marston and Deno (1981) also established the accuracy and stability of TWW, WSC, CLS, and mature words for students with learning disabilities through strong test-retest (*r* = .50-.92), parallel-form (*r* = .74-.96), split-half (*r* = .70 to .99), and interscorer (*r* = .90 or greater) reliability coefficients. The participants in Marston and Deno's study were elementary students from schools in Pennsylvania and the Midwestern U.S. The students were given a story starter with 5 min to respond, but scores were given at 1, 2, 3, and 4 min as well. The students were then given another story starter 1 day, as well as 3 weeks later.

Deno et al. (1982), further examined the relationship between these CBM-WE scores and scores on criterion measures, such as scores from the Test of Written Language, Stanford Achievement Test, and Developmental Sentence Scoring System. The sample was composed of students (44 learning disabled and 86 general education) in Grades 3 to 6 from eight elementary schools, and the students responded to a story starter. TWW (*r* = .58-.84), WSC (*r* = .57-.80), CLS (*r* = .57-.86), and mature words (*r* = .61-.83) had the strongest relationships with the criterion variables and thus, the highest criterion validity for this age group.

Similarly, Gansle et al. (2006) examined the technical adequacy of several CBM scores for students in elementary school (i.e., Grades 1 to 5) through interrater agreement percentages, test-retest reliability, and correlations with criterion measures. The sample in the study consisted of

538 students from a suburban elementary school in the Southwest U.S. In terms of demographics, 69% were Caucasian, 18.5% Hispanic, 6.9% African American, 5.1% Asian American, and 0.5% Native American; 44% were male; 16% received free or reduced lunch; and 6.5% were receiving special education services. The participants in the study completed two story starters, one week apart, and were given 3 min to respond. Test-retest reliability correlation coefficients indicated that TWW (.80), WSC (.82), and CWS (.78) were the most reliable for this age group compared to other scores such as correct capitals (.44), correct punctuation, (.64), complete sentences, (.65), and words in complete sentences (.61). Further, the TWW, WSC, and CWS demonstrated sufficient interrater agreement (93.5-87.7%) and moderate correlations (r = .34-.43) with total language scores on the Stanford-9 achievement test.

Other studies have found that CWS and CIWS scores were technically adequate for students at this age as well (Jewell & Malecki, 2005; Videen, Marston, & Deno, 1982; Weissenburger & Espin, 2005). For example, Weissenburger and Espin (2005) gave 484 students (255 male and 229 female) in Grades 4, 8, and 10 a story starter. The students were given 10 min to write, but the samples were also marked at the end of 3 and 5 min. The samples were scored for TWW, CWS, and CIWS. The sample was a mix of rural and suburban students who came from three school districts in central Wisconsin. Further, 58 students in the sample were special education students, and 5 were receiving English as a Second Language services. The majority of the students were Caucasian, with 10 classifying as Pacific Islander, 3 as Native American, 2 as Asian American, and 2 as Black/African American. The authors concluded that CWS and CIWS scores were valid and reliable for Grade 4 students, as measured by alternate-form correlations (CWS = .79-.84 and CIWS = .73-.80), as well as correlations with the Wisconsin Knowledge and

Concepts Exam language arts test (CWS = .56-.62 and CIWS = .67-.68) and writing assessment (CWS = .58-.60 and CIWS = .58-.65).

In a study by McMaster and Campbell (2008), the authors examined both type of prompt and response time to determine which scores are most appropriate for students in Grades 3 (n = 46), 5 (n = 75), and 7 (n = 96). The students completed passage copying prompts, picture prompts, expository prompts, and narrative prompts. They had 1.5 min to complete the passage copying task and 5 min to complete the other tasks. All students made a slash after 3 min as well. Students were scored based on TWW, WSC, CWS, and CIWS. The authors concluded that CWS scores for 3- to 5-min picture prompts had the most reliability and validity for students in Grade 3, and CWS and CIWS scores for 3- to 5-min narrative and expository prompts were most appropriate for Grade 5 students. Runge et al. (2017) similarly asserted that narrative prompts are more appropriate for lower elementary students, while expository prompts are more appropriate for upper elementary and secondary students. This is not surprising, as students more typically complete narrative writing tasks in elementary school and expository writing tasks in secondary school (Baker & Hubbard, 1995; Kulikowich et al., 2008; McMaster & Campbell, 2008).

Research has also focused on the ability of CBM-WE to discriminate between groups of elementary school students based on gender, disability status, and age. For example, Malecki and Jewell (2003) gave 946 students (48% males and 51% females) in first through eighth grades 3 min to respond to a story starter, and the responses were scored based on TWW, WSC, CWS, percentage of WSC, percentage of CWS, and CIWS. The sample of students came from five different schools in both suburban and rural areas of Northern Illinois. Analyses of the scores indicated a main effect for gender, with girls scoring significantly higher on all scoring indices. However, the effect size was very small.

In a similar study, Jewell and Malecki (2005) gave 203 students (44% male and 56% female) in second, fourth, and sixth grades 3 min to respond to a story starter, and the responses were scored based on the same indices (i.e., TWW, WSC, CWS, percentage of WSC, percentage of CWS, and CIWS). The sample of students came from a school district in a rural area of Northern Illinois. Again, results indicated a main effect for gender, with girls significantly outperforming boys on TWW, WSC, and CWS. Thus, these two studies indicate that practitioners may consider taking gender differences into account when comparing student performance on CBM-WE measures to normative data and when making special education decisions using data from CBM-WE (Malecki & Jewell, 2003).

Tindal and Parker (1991) found that students receiving special education services scored lower than their peers on CBM-WE writing measures. Tindal and Parker administered 10-min CBM-WE story starters to students in third to fifth grade, both with and without learning disabilities. Students completed the probes in the fall and spring, and the responses were scored based on TWW, WSC, CWS, IWS, and total word sequences. Using an analytic scoring system, the students also received scores for story idea, organization-cohesion, and convention-mechanics. Students with learning disabilities performed lower on all writing indices. For example, they did not write as much as their non-disabled peers (e.g., Grade 5 students receiving services had a mean of 44.2 TWW in the spring, while their general education peers had a mean of 54 TWW), and what they did write had more errors than their peers (e.g., Grade 4 students receiving services had a mean of 29.8 CWS in the spring, while their general education peers had a mean of 40.9 CWS). Further, students with learning disabilities received poorer analytic judgments (e.g., Grade 3 students receiving services had a mean score of 2.2 out of 5 for story idea in the spring, while their general education peers had a mean of 3.2). Thus, this study provides evidence that CBM-WE can be useful in differentiating between students with and without learning disabilities at the elementary level.

Similarly, Deno et al. (1980) concluded that LD students performed significantly (i.e., by factors ranging from 1.5 to 2.0 times) lower than their non-LD peers for TWW, mature words, WSC, and CLS. In a related study, Marston et al. (1982) found that LD students performed lower than non-LD students on writing measures. The sample was 82 students (42 males and 40 females) in third to sixth grade. Out of the 82 students, 31 had learning disabilities. Participants responded to story starters for 5 min, and their responses were scored for TWW, WSC, and CLS. Mean scores on each of the indices were lower for LD students. Further, the CBM-WE indices were able to predict students' group membership (i.e., LD or non-LD) in 73% of cases. The authors concluded that CBM-WE was just as effective in predicting group membership as other traditional methods. For example, the Test of Oral Written Language correctly identified 80% of cases, and an aptitude-achievement discrepancy from the Woodcock-Johnson tests correctly identified 74% of cases.

In addition to gender and disability status, CBM-WE has been able to differentiate between grade levels, with middle school students outperforming elementary students and older elementary students outperforming younger elementary students on TWW, mature words, WSC, CLS, CWS, percentage of WSC, percentage of CWS, and CIWS (Deno et al., 1980; Malecki & Jewel, 2003; Mirkin, & Marston, 1980; Videen et al., 1982). For example, Videen et al. (1982) examined the CWS scores of a sample of 50 student responses from the Deno et al. (1980) study. The authors concluded that CWS increased by about 10 with each successive grade level. For example, Grade 6 students (average of 58.8 CWS) wrote over two times the number of CWS written by Grade 3 students (average of 27.3 CWS).

The body of research surrounding the use of CBM-WE with elementary students has demonstrated that in general, the most technically adequate tasks for this age group are 3- to 5-min story starters, topic sentences, and picture stimuli (Espin et al., 2004; McMaster & Campbell, 2008; McMaster et al., 2009). In terms of scores, TWW, WSC, CLS, and CWS produce the most reliable and valid results for elementary students (Deno et al., 1982; Deno et al., 1980; Espin, Weissenburger, & Benson, 2004; Gansle et al., 2006; Marston & Deno, 1981; McMaster & Campbell, 2008; McMaster et al., 2009; Weissenburger & Espin, 2005). Studies have also demonstrated the ability of CBM-WE to differentiate between gender, grade level, and special education status for elementary students (Deno et al., 1980; Jewell & Malecki, 2005; Malecki & Jewell, 2003; Marston et al., 1982; Tindal & Parker, 1991; Videen et al., 1982).

Research at middle school level. Studies at the middle school level have also provided evidence that brief and simple CBM-WE measures can serve as overall indicators of written expression for both general and special education students (Espin et al., 2005; Merrigan, 2012; Tindal & Parker, 1989). CBM-WE scores additionally have the potential to be used for progress monitoring, as well as screening and eligibility decisions, for middle school students with disabilities (Parker, Tindal, & Hasbrouck, 1991b; Tindal & Parker, 1989). Other CBM-WE studies at the middle school level have examined the various features of measures to determine which are most reliable and valid for this age group, as those used at the elementary level are not necessarily adequate (Espin, De La Paz, Scierka, & Roelofs, 2005).

Merrigan (2012) examined various methods of narrative CBM-WE for their reliability and validity. The participants in the study were 422 students (104 sixth graders, 103 seventh graders, and 89 eighth graders) from a rural middle school in Central Pennsylvania. Students were given narrative and expository story starters, and students responded to 3-, 5-, and 7-min probes.

Responses were scored for TWW, WSC, percentage of WSC, CWS, CIWS, and percentage of CWS. Results of the study indicated that CIWS scores from a 3-min narrative task had the most technical adequacy for middle school students, as evidenced by interrater reliability, alternate forms reliability, and criterion validity (i.e., correlations between CBM indices and scores on the Test of Written Language). In contrast, McMaster and Campbell (2008) found that 5- to 7-min expository prompts scored for CIWS were the most reliable and valid for progress monitoring Grade 7 students. In a review of research, however, McMaster and Espin (2007) did not find a significant difference in validity between narrative and expository prompts or 3- and 5-min response periods for middle school students. Espin (2005) and Runge et al. (2017) also posit that both narrative and expository probes are appropriate for this age group.

Other studies have also provided support for CIWS as the most valid and reliable measure of overall writing proficiency for middle school students (Espin et al., 2000; Jewell & Malecki, 2005), followed by CWS (Espin et al., 2000; Videen et al., 1982). For example, Espin et al. (2005) gave a sample of students 35 min to respond to expository writing prompts, and the responses were scored for CWS and CIWS. The students then went through 4 weeks of writing instruction, and they were given the same task. The student sample consisted of 22 seventh- and eighth-grade students (11 males and 11 females) from a suburban middle school in the Southeast U.S. Six of the students were diagnosed with LD, six were considered low-achieving writers, six were considered average-achieving writers, and four were considered high-achieving writers. Espin and colleagues found that the writing criterion measures in the study (i.e., number of functional elements in student essays and quality ratings of student essays) were strongly related to the students' CWS (r = .68-.83) and CIWS (r = .66-.82) scores.

Middle school research has also shown that CWS and CIWS scores are sensitive to student growth over time (Espin et al., 2005; Videen et al., 1982). For instance, in the study by Espin et al. (2005), significant differences were found between CWS and CIWS pre- and post-test scores for all students. When looking only at the first 50 words written in the essays, however, students with LD demonstrated more substantial growth than their peers. Post-test scores for LD students grew an average of 9 CWS and 14 CIWS, while post-test scores for low, middle, and high achieving writers had an average growth of 1.5 CWS and 7 CIWS. Espin and colleagues concluded that longer writing samples are needed to show growth for higher performing students. Thus, response periods may need to differ depending on a student's proficiency in written expression.

CBM-WE measures have additionally been able to differentiate between ability levels for middle school students. In a study by Watkinson and Lee (1992), students in sixth to eighth grades were given 6 min to respond to story starters, and their responses were scored based on TWW, legible words, CSW, CWS, IWS, percentage of legible words, percentage of CSW, and percentage of CWS. The sample of 52 students (36 males and 16 females) came from a suburban middle school near Kansas City. Half of the students had learning disabilities in written expression, and the other half were randomly selected general education students. Results of the study indicated that students with learning disabilities scored significantly lower than their peers on CWS, percentage of legible words, percentage of CSW, and percentage of CWS, and significantly higher on IWS. Further, as with elementary students, CWS and CIWS scores have been shown to differentiate between male and female middle school students, with female students performing higher than males (Jewell & Malecki, 2005; Malecki & Jewell, 2003).

Amato and Watkins (2011) provided evidence that CBM-WE is able to predict scores on other achievement measures, such as the Test of Written Language – Third Edition (TOWL-3) for students in Grade 8. Participants in the study included 447 students from a New Jersey school district. The sample was 48% female, 60% Caucasian, 14% Black, 14% Hispanic, and 12% Asian/Pacific Islander. Further, 15% of the sample received special education services, 9% received free lunch, and 15% received reduced lunch. Students were given story starters, and the responses were scored based on TWW, percentage of WSC, percentage of CWS, CIWS, number of sentences, number of correct capitalizations, and number of correct punctuation marks. The students subsequently completed the TOWL-3 as a criterion measure. The seven CBM-WE indices accounted for a total of 44% of the variance in Overall Writing Quotient scores on the TOWL-3. The only significant predictors were percentage of CWS, correct punctuation marks, and correct capitalizations. Thus, it appeared that more complex CBM-WE scores were the most adequate for predicting middle school students' writing achievement.

Tindal and Parker (1989) also found that CBM-WE scores were able to predict scores on criterion measures. Participants in the study included 172 students (54 females) in sixth to eighth grades from four low to middle socioeconomic status, suburban middle schools in California. All students were either in special education or remedial programs. Students were given a story starter and 6 min to respond. Responses were marked at the end of 3 min as well. The responses were scored based on the following CBM-WE indices: TWW, CSW, CWS, Legible Words, Mean length of correct word sequence strings, percentage of CSW, percentage of CWS, and percentage of legible words. The same writing samples were also scored holistically on a scale of 1 (very poor) to 7 (very effective) for communicative effectiveness. After regressing the students' holistic judgments on the CBM-WE indices, the authors found that percentage of CWS and percentage of WSC were the best predictors. The scores produced moderately strong coefficients (r = .75 and

.73, respectively) and explained more than half of the variance in holistic judgments of special education and remedial middle school students' writing samples.

Research at the middle school level has shown that the most valid and reliable tasks for elementary students are not necessarily the most technically adequate for this age group. Response periods of 3 to 5 min still appear technically sound; however, narrative and expository writing prompts tend to be the most appropriate for middle school students (McMaster & Campbell, 2008; McMaster & Espin, 2007; Merrigan, 2012). In addition, more complex scores such as CWS and CIWS tend to be more appropriate for middle school students, compared to simpler scores such as TWW and CLS (Espin et al., 2005; Espin et al., 2000; McMaster & Campbell, 2008; Merrigan, 2012). Further, CBM-WE is sensitive to individual student growth (Jewell & Malecki, 2005; Videen et al., 1982), and it is also able to predict scores on other measures of writing achievement for middle school students (Amato & Watkins, 2011; Tindal & Parker, 1989). Finally, CBM-WE measures at the middle school level can differentiate between students of different genders (Jewell & Malecki, 2005; Malecki & Jewell, 2005) and ability levels (Watkinson & Lee, 1992).

Research at high school level. CBMs were not developed and used with secondary students until the 1990s, when there was a change in policy and practice toward increasing the frequency of statewide testing for reading and writing (Espin, 2012). However, research regarding CBMs at the secondary level, especially in the area of written expression, is still lacking as most of the available research is conducted with elementary and middle school students (Espin, 2012; Espin et al., 1999; Espin et al., 2008). Further, national norms are only available for students in kindergarten through Grade 8 (Runge et al., 2017). A potential reason for the paucity of research at the high school level is that measurement and scoring tend to be more complex for this age group compared to younger students (Espin et al., 1999). There is more disagreement about the curriculum that should be taught in secondary school, compared to primary school, which also makes it more difficult to measure progress within the curriculum (Shinn, 1998). As a result, the technical adequacy of CBM-WE for high school students is not as strong as for younger students (Espin et al., 2004; Weissenburger & Espin, 2005). Thus, future CBM-WE research should focus on bridging this gap in the literature.

In general, the measures used with primary students are not appropriate for secondary students (Shinn, 1998). For instance, expository writing prompts are more valid, reliable, and sensitive to growth than other writing measures for secondary students (Espin, 2012; Espin et al., 2000; Runge et al., 2017). In addition, research at the secondary level has shown that longer writing samples of 5 to 10 min increase the technical adequacy of CBM-WE measures (Espin, 2012; Espin et al., 2004; Weissenburger & Espin, 2005). For example, Espin et al. (2008) looked at the impact of response duration, among other factors, on CBM-WE technical adequacy. The sample included 183 Grade 10 students (57% female and 43% male). Of the total sample, 5% received special education services, 54% were students of color, and 21% were English Learners. Participants were given a narrative writing prompt to which they had 10 min to respond. Responses were also marked at 3, 5, and 7 min, and they were scored based on TWW, WSC, CWS, and CIWS. In addition, students completed the Minnesota Basic Standards Test/Minnesota Comprehensive Assessments in written expression as a criterion variable.

Espin and colleagues found that alternate-form reliability increased for Grade 10 students when given a 5-min response (r = .74-.77) period compared to a 3-min period (r = .64-.66), and it increased even further for a 7-min period (r = .80-.82). Giving students a 10-min response period (r = .82-.85) did not increase reliability significantly further, however. In terms of CBM indices, CWS and CIWS had the strongest alternate-form reliability (CWS = .66-.85; CIWS = .66-.84) and criterion validity (CWS = .43-.48; CIWS = .56-.60). According to the authors, a limitation of the study was that only one criterion variable was used. Further, the sample of participants had higher overall performance than the average performance of students from their district, but lower overall performance than the average performance of students in the state. Thus, the validity obtained in the study may underestimate the general performance of lower-performing districts and overestimate the general performance of higher-performing districts. Lastly, student performance was measured at a single point in time, rather than showing growth across multiple time points.

Based on a literature review, McMaster and Espin (2007) also asserted that similar to middle school students, CIWS is the most technically adequate score for high school students, followed by CWS (Espin et al., 2008). Although production-independent measures (e.g., percentage of CWS, percentage of CIWS) have been shown to be able to predict academic performance for secondary students (Jewell & Malecki, 2005), they are not always as useful as production-dependent measures (e.g., CWS, CIWS), especially for progress monitoring purposes, because they are less sensitive to growth and differences among students (Espin et al., 2000; Shinn, 1998). McMaster and Espin also note that CWS is able to discriminate among high school students much better than elementary or secondary students.

Similar to CBM-WE with other age groups, the measures are able to discriminate between high school students with and without disabilities as well. For example, Espin et al. (1999) gave 147 (76 male and 71 female) Grade 10 students 3 min to respond to narrative story starters, and the writing samples were scored for TWW, WSC, CWS, characters written, characters per word, sentences written, and mean length of correct word sequences strings. The sample included students with LD (*n* = 9), students in a basic English course (*n* = 39), students in a regular English course (*n* = 50), and students in an enriched English course (*n* = 49). All participants came from a large, middle SES high school in the Midwestern U.S. Espin and colleagues found that across all CBM-WE indices, students with learning disabilities scored the lowest. Their peers in basic English scored second lowest, followed by students in regular English. The Enriched group scored the highest. Scores with significant differences included characters per word, sentences, and mean length of correct sequences strings. The authors posited that because students with learning disabilities perform differently than their non-disabled peers, their academic growth in writing may also look different. Therefore, different measures may need to be used to monitor the progress of students with and without disabilities.

In a study with general and special education students, Diercks-Gransee et al. (2009) similarly found that students with disabilities performed differently than their general education peers. Participants in the study included 82 Grade 10 students from two rural public school districts in Wisconsin. Of the students, 74 were general education students and 8 (6 males and 2 females) were diagnosed with learning disabilities. Students completed two narrative story starters, and they were given 10 min to write. Responses were scored for IWS, correct punctuation, number of adverbs, and number of adjectives. Students with disabilities used significantly fewer punctuation marks and made significantly more errors in spelling, syntax, and grammar when compared to students without disabilities.

Diercks-Gransee and colleagues were also able to use cut scores for incorrect word sequences and correct punctuation to correctly identify those students that had learning disabilities. Results indicated that using a 20% cut score accurately identified 7 out of 8 students with learning disabilities for IWS and 6 out of 8 students for correct punctuation marks. However, there were also 9 misidentifications for IWS and 10 for correct punctuation marks. Despite the misidentifications, Diercks-Gransee and colleagues asserted that educators may use these scores to screen high school students for learning disabilities. A limitation of this study, however, is the generalizability to other populations. All participants in the sample were White or Caucasian and in 10th grade. Thus, the results may not be applicable to more diverse populations or other grade levels.

Compared to the body of research at the elementary and middle school level for CBM-WE, research at the high school level is lacking; however, multiple studies have still shown that the measures have utility with this age group. Similar to middle school students, CIWS scores are the most technically adequate for high school students, followed by CWS (Espin et al., 2008; McMaster & Espin, 2007). As students get older, it also appears that expository prompts with longer response periods of 5 to 10 min are necessary to obtain adequate reliability and validity (Espin, 2012; Espin et al., 2000; Espin et al., 2008; Espin et al., 2004; Weissenburger & Espin, 2005). In general though, a combination of CBM-WE measures and scores, along with obtaining multiple samples, will likely produce the most technically adequate results (Espin et al., 1999; McMaster & Espin, 2007; Shinn, 1998). Lastly, CBM-WE measures can differentiate between high school students with and without disabilities (Diercks-Gransee et al., 2009; Espin et al., 1999).

CBM Research at Postsecondary Level

Espin and colleagues (2012) argue that CBMs could potentially be used in future research to determine the level of performance needed to succeed in postsecondary and employment settings, as well as to monitor individual progress toward that level of performance. Several studies have examined the use of CBMs with postsecondary students, but none have looked at the utility of CBM-WE with the general university student population. For example, Bean and Lane (1990) administered reading CBMs to 57 adults in a basic literacy program in Pittsburgh, Pennsylvania. Participants of the program were unemployed adults who were reading below an eighth-grade level. Participants of the study included 38 females and 52 males. Of the participants, 52 were Black and 5 were White. The participants were required to read both narrative and expository passages, varying from fourth to eighth-grade reading levels, for 1 min. The number of correct words read per minute was then calculated.

Results of the study indicated that the CBMs reliably measured student progress in reading, were sensitive to student growth, gave educators a method to monitor and make adjustments to instruction, and provided a way to regularly give students feedback on their performance. Further, according to responses from a questionnaire that was given upon completion of the study, students and teachers had positive perceptions regarding the use of CBM to assess students' skills and monitor their progress. Despite the overall positive conclusions, the authors noted that the study had some limitations (i.e., small sample size and low variability in scores) that may limit the validity of the findings.

Lewandowski et al. (2003) also administered reading CBMs to university students, but for a different purpose, which was to develop norms for the expected reading rate of college-age students. The sample from this study consisted of 90 students, aged 18 to 26, from a psychology course at a private university in the Northeastern U.S. The demographic composition of the sample was 66.7% female, 53.5% freshmen, 82.2% Caucasian, 6.7% African American, 3.3% Asian American, and 5.6% Latin American. Participants were given three oral reading fluency CBM probes. The passages were at an eighth-grade reading level or below, and the median number of words read correctly per minute (WRCM) was calculated for each participant. Participants also completed several subtests from the Nelson Denny Reading Test (NDRT) and the Woodcock Johnson Tests of Achievement – Third Edition (WJ-III).

Based on the oral reading fluency CBM probes, students read an average of 189 WRCM. In addition, WRCM had stronger correlations with subtest scores on the WJ-III than the NDRT scores did. Further, WRCM was a stronger predictor of reading comprehension, as measured by the NDRT, compared to the reading rate score on the NDRT. In general, the authors concluded that using CBMs to measure reading fluency has utility with the adult population, and CBMs may even be a more technically adequate measure than traditionally used standardized reading assessments such as the NDRT.

In another study, Larson and Ward (2006) administered CBM vocabulary probes to university students to determine whether course grades or general reading ability, as measured by college entrance exam scores, predicted CBM trend lines. Participants included 69 undergraduate students (22 males and 46 females) enrolled in introductory psychology courses at a historically Black college in the Mid-Atlantic U.S. Of the entire sample, 72% were Black, 9% were White, and 7% identified as “Other plus African American.” Further, 72% were freshmen, 19% were sophomores, 9% were juniors, and 1% were seniors. Participants were asked to complete vocabulary CBM probes once per week for nine weeks, and the scores were subsequently graphed. The probes required the students to match vocabulary words with their definitions. Results indicated that the direction of CBM vocabulary trend lines did not have a significant relationship with reading ability or course grades.

According to the authors, a limitation of the study was that student grades were not normally distributed across the classes (i.e., there were very few Ds and Fs). Despite this limitation and the insignificant results, important implications for universities were discussed. For instance, if CBM measures were shown to be valid and reliable for use with postsecondary students, university instructors would be provided with a simple method for assessing and monitoring student progress during a course. Similar to use with school-age students, CBM data could be used to identify students who are struggling within a given course and who may need additional supports to succeed. Providing interventions and supports early on to struggling students could help prevent long-term negative outcomes, such as student drop out (Larson & Ward, 2006; McMaster et al., 2009).

Lastly, Hosp et al. (2014, 2018) administered reading, math, and writing CBMs to college students; however, all participants in the study had documented disabilities. The purpose of the two studies was to examine criterion validity when using CBMs with students with intellectual and developmental disabilities. The 2014 study included 41 postsecondary students (ages 19 to 23) from a two-year certificate program in the Midwestern U.S. for individuals with cognitive and intellectual disabilities. Of the total sample, 19 participants had cognitive/intellectual disabilities, 9 had Autism Spectrum Disorder, and 13 had multiple disabilities. Further, 34% of the participants were female and 90% were White. Participants in the study completed the following CBM tasks: oral passage reading (i.e., students read out loud for 1 min and were scored on words read correctly and errors), maze (i.e., students silently read a passage with missing words, decided from several options which words fit, and were scored for correct restorations), math concepts and application (i.e., students completed as many math application problems as they could in 8 min and were scored for correct answers), and written expression (i.e., students were given 1 min to think, 3 min to respond to a story starter, and were scored for TWW, WSC, CWS, and CIWS). Lastly, students were given the Woodcock-Johnson III Tests of Achievement (WJ-ACH) as a criterion measure.

Analyses from the 2014 study indicated moderate to strong correlations between CBM measures and broad achievement scores on the WJ-ACH. A finding specific to written expression was that CWS correlated more strongly with broad written language scores on the WJ-ACH compared to TWW and WSC. This was not surprising since CWS is a broader writing index that involves more complex skills than TWW and WSC. Limitations of the study were a small sample size, non-random selection of participants, and lack of a nationally representative sample. An additional limitation was that although the CBM tasks were technically adequate, they did not necessarily align with the students' instructional level. Further, only one probe was administered per measure and only one aspect of technical adequacy was examined (i.e., criterion validity measured by moderate to strong correlations between CBM and WJ-ACH scores). Regardless of the limitations, the authors concluded that CBMs have utility as a measure of academic performance among college students with intellectual and developmental disabilities.

The 2018 replication study included 45 postsecondary students with developmental disabilities from a two-year certificate program in the Midwestern U.S. Students completed the same oral passage reading, maze, math computation, and math concepts and applications CBM tasks, along with the WJ-ACH. Again, there were moderate to strong correlations between the CBM and WJ-ACH scores. The authors identified similar limitations to the 2014 study (i.e., small sample size, non-representative sample, non-random sample, and examination of only one type of technical adequacy). As a result of the 2014 and 2018 studies, the authors posited that CBMs have the potential to be used to make screening decisions with college students with intellectual and developmental disabilities. Further, through their use as progress monitoring measures, CBMs could improve both academic instruction and student performance over time.

Although research related to CBM is lacking at the high school level, CBM literature for college students is even more sparse. Five studies to date have examined the utility of CBM for this population of students. Results indicated that CBMs: a) can be used to measure academic performance and student progress, b) provide a way to monitor course instruction, c) are sensitive to student academic growth, and d) are helpful in giving feedback to students (Bean & Lane, 1990; Hosp et al., 2014, 2018). The studies also provided evidence of reliability, predictive validity, and criterion validity of CMB for postsecondary students, as well as evidence for the use of CBM to develop local norms for this age group (Bean & Lane, 1990; Hosp et al., 2014, 2018; Lewandowski, 2003). Of the five studies reviewed, only one examined the use of CBM-WE with college students (Hosp et al., 2014); however, all students in the study had intellectual or developmental disabilities. The study suggested that as with middle and high school students, broad CBM-WE indices involving more complex writing skills (i.e., CWS) are more technically adequate for college students compared to scores such as TWW and WSC.

It is important to note that CBMs may not be the most authentic or natural form of writing assessment for college students. The tasks are not particularly contextualized, which is an important component of college writing assessment, according to the position statement by the Conference on College Composition and Communication (CCCC) Committee on Assessment (2014). The small bursts of writing produced by CBM-WE are quite different from the curriculum of college writing courses, which includes components such as discourse and deliberation. Despite its limitations, CBM-WE has been shown to predict students' performance on other writing assessments, as well as overall writing achievement (Fewster & MacMillan, 2002; Merrigan, 2012). However, more research is needed on the predictive validity of CBM-WE with the college population. Although the use of CBM-WE with college students may be fairly narrow in scope, the tasks can be used with a combination of other measures to assess college students' general proficiency in written expression and to develop local norms for use with a specific group of students (Fewster & Macmillan, 2002). CBMs have previously been limited in their use with postsecondary students, but it is possible that they could be useful with this demographic of students in addition to the school-age population.

Summary

CBM is an alternative to traditional standardized achievement assessments that can be used for a variety of purposes such as measuring academic proficiency, monitoring progress, making special education decisions, and developing local norms. CBM has been developed for the academic areas of reading, math, vocabulary, and written expression. However, compared to other areas of CBM literature, there is a paucity of research related to CBM in written expression. Despite this gap in the CBM literature, available research on CBM-WE indicates that similar to other types of CBM, CBM-WE can validly and reliably evaluate both general and special education students' academic performance and educational programs. CBM-WE can also be used to monitor academic progress, make screening and eligibility decisions, improve writing skills, measure student growth, predict scores on other achievement measures, and differentiate between groups of students.

Within the CBM-WE literature, most studies have examined the technical adequacy of CBM-WE for elementary students, followed by middle school students. Research for elementary students has shown that the most technically adequate tasks for this age group are 3- to 5-min story starters, topic sentences, and picture stimuli scored for TWW, WSC, CLS, and CWS. Middle school research demonstrates that 3- to 5-min narrative and expository writing prompts scored for CWS and CIWS tend to be the most appropriate for these students. Although less research is available for older students, CBM-WE studies involving high school students indicate that 5- to 10-min expository prompts scored for CWS and CIWS result in the highest reliability and validity. Only one study (Hosp et al., 2014) has examined CBM-WE for college students, but the results similarly indicated that more complex scores such as CWS are the most appropriate. After reviewing the literature related to CBM-WE, it is clear that more research is currently needed on the utility of CBM-WE for older students, especially those in college.

The Current Study: Research Aims

According to Lewandowski et al. (2003), students with disabilities, especially learning disabilities, have increasingly been attending college and requesting accommodations for their disabilities. As a result, services for students with learning disabilities have also increased among colleges and universities. For example, as of June 2018, the Pennsylvania State University (Penn State) had 1,577 students actively being serviced by Student Disability Resources (K. Jervis, personal communication, June 27, 2018). Of these students, 670 (approximately 43%) were students with a learning disability. Typical accommodations for students with learning disabilities at Penn State include additional time on tests, use of reading software, and note-taking assistance.

Despite the availability of services and accommodations, students with disabilities do not benefit from attending college as much as the general student population (Hosp et al., 2014). For instance, students with disabilities are less likely to enroll, less likely to be employed, and they typically do not hold jobs as long as individuals without disabilities. A possible reason for some of the negative outcomes associated with students with disabilities is that students with learning disabilities often experience difficulty with written expression skills (Diercks-Gransee et al., 2009; McMaster, Du, & Petursdottir, 2009).

One way to potentially diminish this discrepancy between disabled and non-disabled students is through the use of curriculum-based measurement in written expression. Although CBMs have historically been limited in their use with postsecondary students, it is feasible that their utility could be extended to include this age group. Currently, however, CBM-WE norms only exist up through Grade 8. If CBMs are to be used by postsecondary institutions and students, development of norms at the college age is necessary, especially because the college student population is likely to differ considerably from the general student population (Kamphaus, 1984). Therefore, the purpose of the current study was to create CBM-WE local norms for freshmen college students at Penn State.

In order to develop these norms, a sample of freshmen from First-Year Seminar in Education courses at Penn State responded to a single CBM-WE probe. Their scores were then converted into percentile ranks. The resulting norms have several potential uses for secondary and postsecondary students. First, the norms provide CBM-WE scores and corresponding percentile ranks for Penn State freshmen. Prospective students can compare their current writing performance (measured by CBM-WE) to the typical performance of freshmen at Penn State or other similar universities. Prospective students can also use the norms to set a goal for their writing proficiency, and they can monitor their progress toward the goal through continued use of CBM-WE. Second, the norms can be used to measure Penn State students' proficiency in written expression to aid in the identification of students with learning disabilities.

CHAPTER 3

Method

This section describes the participant sample and recruitment methods, training of research assistants, data collection, measures and scoring, and analyses. The Institutional Review Board at the Pennsylvania State University approved all procedures.

Participants

Participants included a sample of 116 freshmen from six Fall 2018 College of Education freshmen seminar courses at The Pennsylvania State University's main campus, University Park. Kamphaus (1984) asserts that 100 participants per group is desirable when developing local norms, since each participant's score can be set to a percentile rank; thus, the sample size was sufficient for the purposes of the current study. The majority of the sample was female (82.8%), and students were ages 18 (79.3%), 19 (19.8%), and 21 (.9%). The majority of students were also White (89.7%), followed by Biracial (i.e., the students identified as more than one ethnicity; 4.3%), Black (2.6%), Hispanic (2.6%), and Asian (.9%). Of the entire sample, 12.9% reported receiving services from Student Disability Resources, 6.9% were first generation college students, and less than 1% reported that English was their second language.

Students in the sample were asked to report their freshman English course. The majority of students (94.8%) were enrolled in ENGL 15 (Rhetoric and Composition), 1.7% were enrolled in ENGL 30 (Honors Composition), and less than 1% were enrolled in ENGL 4 (Basic Writing Skills), CMLIT 10 (World Literatures), and American Literature (specific literature course was not reported). Students were also asked to report their current major. Students reported a variety of education majors, and these were collapsed into four categories of majors (i.e., Education, Early Education, Secondary Education, and Other). The majors were collapsed into four larger groups so that the groups would be similar enough in size to run subsequent analyses examining the effect of major on CBM-WE scores. Out of the total sample, 30.2% were classified as Early Education (included Early Childhood Education and Elementary majors and Early Childhood Education majors), 23.3% as Secondary Education (included Secondary Education and Middle Level Education majors), 16.4% as Education (included Education, Special Education, Science Education, Social Studies Education, and World Languages Education majors), and 30.2% as Other (included Rehabilitation and Human Services, History, Hospitality Management, Labor and Employment Relations, Nutrition, Psychology, Undecided, and Dual majors).

Procedure

Training of research assistants. Two research assistants were recruited from the Penn State school psychology doctoral program. Students in the program had adequate experience administering and scoring CBM-WE measures from doctoral-level coursework in school psychology and clinical experiences. The two graduate students who agreed to assist in scoring the CBM-WE probes attended an hour-long training on Penn State's campus before scoring began. The training was conducted by the principal investigator. Each research assistant was fiscally compensated for time spent at the training, as well as time spent scoring responses.

Prior to the training session, research assistants were given instructions/guidelines on how to score the CBM-WE probes. At the beginning of the training session, the research assistants were given time to ask questions about the scoring guidelines. The research assistants were then required to demonstrate proficiency with scoring procedures by scoring two CBM-WE probes. Similar to the training procedure used by McMaster and Campbell (2008), the research assistants' CWS, IWS, and CIWS scores for each probe were compared to a master copy, scored by both the principal investigator and a professor in Penn State's school psychology program.

The percentage of agreement for each score was calculated by dividing the smaller score by the larger score and multiplying the result by 100. Point-by-point agreement was also calculated by dividing the overall number of agreements between the two raters by the overall number of agreements plus disagreements. Average percent agreement for the training probes was within acceptable limits, with 99% agreement for CWS, 93.75% agreement for IWS, and 98% agreement for CIWS. Point-by-point agreement (99.25%) was also within acceptable limits. For scores that did not agree with the master copy, the research assistant and principal investigator discussed and reconciled the discrepancies until the research assistant's final agreement with the mastery copy was 100% for each type of score.

Participant recruitment. After the current study was approved by the Penn State Institutional Review Board, professors teaching College of Education freshmen writing seminar courses were contacted and given information about the purpose of the study. The principal investigator requested permission from the professors to come into their freshmen courses, during the last 10 min of a single class period, to administer CBM-WE writing probes to willing student participants. Once a professor agreed, a specific date and time was scheduled for the data collection. On the administration date, all students present were invited to participate in data collection. Students received a brief overview of the purpose of the study and what their participation would entail. As an incentive for students to participate, instructors were asked if they were willing to offer their students extra credit for participating. All professors agreed to offer one point of extra credit for participation. Students who did not want to participate could still earn the extra credit by completing the CBM exercise and turning their materials in to their professor to be shredded, instead of turning their materials into the principal investigator. All participants who supplied their email addresses were also placed into a drawing for five $10 Amazon gift cards.

Data collection. Students within a given class who agreed to participate were assessed as a group. All participants were given a short demographic questionnaire and a paper with the CBM-WE probe. The questionnaire was completed first, and participants were asked to identify their gender, ethnicity, age, major of study, and first year English composition placement. They were also asked to indicate whether: a) they received services from Student Disability Resources, b) they were a first generation college student), and c) their first language was English. Students were given 1 min to read their probe and think about what they were going to write. Subsequently, they were given 3 min to respond to the prompt. After the 3 min passed, students were told to stop writing and return their materials. Standardized directions were read aloud to the group of students. These directions were adapted from Powell-Smith and Shinn (2004) and Merrigan (2012), and they can be found in Appendix A.

Scoring. Probes were divided among the principal researcher and the two research assistants for scoring. Each response received a CWS, IWS, and CIWS score. See Appendix B for scoring guidelines. Guidelines were taken from a combination of CBM-WE scoring procedures outlined by Hosp et al. (2016), Pearson (2009), Powell-Smith and Shinn (2004), and Wright (2013).

Reliability. Reliability of scores was demonstrated through the calculation of percent agreement and point-by-point agreement for a subset of 24 (20%) of the total number of probes, as suggested by Runge et al. (2017). Thus, each probe in the subset was scored a second time by a different rater. The percentage of agreement between the two raters for each type of score (i.e., CWS, IWS, CIWS) was calculated by dividing the smaller score by the larger score and multiplying the result by 100. The point-by-point agreement between the two raters was calculated for each of the 24 probes by dividing the total number of agreements by the total number of agreements plus disagreements.

After calculating the original percent agreement and point-by-point agreement between raters for the subset of 24 probes, any discrepancies between the two raters were discussed and corrected. For each individual word sequence upon which the raters originally disagreed, the two raters discussed why they each marked the word sequence as correct or incorrect. The raters then referenced the scoring guidelines in order to come to an agreement on the individual word sequence rating. They also changed their ratings on the probes to match one another. For difficult or unclear word sequences, the third rater was consulted in order to come to a decision. Thus, final percent agreement and point-by-point agreement scores were all 100%.

Original average percent agreement scores for CWS (96.5%) and CIWS (92.42%) were within acceptable limits (i.e., above the desired amount of 80%). The average percent agreement of IWS scores (65.38%) was below the expected level. As noted, all differences in scores between raters for the sample of 24 probes were reconciled so that final percent agreement for CWS, CIWS, and IWS were all 100%. Reliability of the IWS scores tended to be lower due to the generally small number of IWS scores for each probe (i.e., 0 to 28). For instance, some responses only had 1 IWS. If one rater scored 1 IWS and the other scored 0 IWS, the percent agreement would drop down to 0%, even though the two raters' scores only differed by one word sequence.

It is important to note that raters could have arrived at the same score, even if they did not agree on every word sequence. For instance, one rater might have found a spelling error, while one rater may have instead found a capitalization error. Although this would result in the same CWS, IWS, and CIWS scores, and thus 100% agreement across scores, the raters would have actually differed on two different word sequences. Thus, point by-point-agreement was calculated, as this score takes into account all agreements and disagreements between two raters on a given probe. Point-by-point agreement for the 24 probes was within acceptable limits, with an average of 94.71%, a minimum of 80%, and a maximum of 100%. Again, any discrepancies between raters for the subset of 24 probes were discussed and reconciled so that final point-by-point agreement was 100% for each response.

During scoring sessions in which the research assistants were present, a probe was consensus scored by all three raters once every hour to ensure consistency of scoring procedures. Percent agreement and point-by-point agreement were calculated, and disagreements were discussed and reconciled in the same manner as the 24 reliability probes. For the 5 probes that were consensus scored, average percent agreement for these 7 probes was within acceptable limits for CWS (96.9%) and CIWS (93.2%), as was point-by-point agreement (94.7%). However, average percent agreement for IWS (72.4%) was again below the expected level. Discrepancies for the consensus probes were also discussed until raters reached final agreement scores of 100%.

Measures

CBM-WE probe stimuli. The expository probe stimulus was chosen from a list of writing prompts published by Learning Express (2003). The probe option that was used in the current study was: Describe your favorite game and explain why it is your favorite. Given that CWS, IWS, and CIWS scoring procedures produce the most reliable and valid scores for older students (Espin et al., 2012; McMaster & Campbell, 2008; McMaster & Espin, 2007; Runge et al., 2017), student responses were scored using these three methods.

Existing research using CBM-WE varies widely in terms of the reliability and validity of scores, with the majority of research using CWS, followed by CIWS and IWS. According to McMaster and Espin (2007), CBM reliability coefficients below .60 are “weak,” coefficients between .60 and .70 are “moderate,” coefficients between .70 and .80 are “moderately strong”, and coefficients greater than .80 are “relatively strong.” The procedure for deriving each score, as well as a description of each score’s technical adequacy follows.

Correct Word Sequences. CWS are any two adjacent writing units (i.e., words, capitalization, and essential punctuation) that are acceptable to a native English language speaker, within the context of what is written (Breaux & Frey, n.d.; Hosp et al., 2016; Pearson, 2009; Powell-Smith & Shinn, 2004; Wright, 2013). CWS is one of the most useful types of scores because it considers multiple writing factors, including meaning, punctuation, grammar, syntax, spelling, and punctuation (Hosp et al., 2016). Other advantages include that it can provide information about patterns of errors, and it is more sensitive to instruction than other scores, making it especially useful for progress monitoring purposes (Hosp et al., 2016).

Prior research has shown that interscorer agreement for CWS has ranged from 86% to 98% (Espin et al., 1999; Espin et al., 2008; Gansle, et al., 2002; Gansle et al., 2006; Videen et al., 1982), and inter-rater reliability coefficients have ranged from .95 to .99, falling in the relatively strong range (Malecki & Jewell, 2003; Watkinson & Lee, 1992). Alternate form reliability for CWS has ranged from weak to relatively strong, with correlations of .46 to .93 (Espin et al., 2000; Espin et al., 2008; Gansle et al., 2002; McMaster & Campbell, 2008; Weissenburger & Espin, 2005). More specifically, Espin et al. (2008) found that alternate form reliability for Grade

10 students given 7 min to respond to a narrative prompt was .82. McMaster and Campbell (2008) gave students in Grades 3, 5, and 7 expository probes with a 7-min response period as well. Alternate form reliability for CWS ranged from .76 to .82. Gansle et al. (2006) also examined test-retest reliability of CWS for elementary students who completed two CBM story starters a week apart, and the authors found a moderately strong correlation of .78. Lastly, correlations with criterion measures have ranged from .18 to .84, again falling in the weak range up to the relatively strong range (Espin et al., 2005; Espin et al., 1999; Espin et al., 2000; Espin et al., 2008: Gansle et al., 2002; Jewell & Malecki, 2005; Weissenburger & Espin, 2005).

Incorrect Word Sequences. IWS are any two adjacent writing units that are not considered a CWS (Breaux & Frey, n.d.; Pearson, 2009). Watkinson and Lee (1992) calculated interrater reliability for scores obtained from students in sixth to eighth grade. The students were given 6 min to respond to a story starter. The interrater reliability coefficient for IWS was relatively strong at .87. Espin et al. (2000) had alternate form reliability coefficients in the moderate range for IWS, ranging ranged from .60 to .67. However, correlations with criterion measures were weak to moderate, ranging from .30 to .63.

Correct minus Incorrect Word Sequences. CIWS is the difference between the CWS and IWS scores for a given response. Interscorer agreement for CIWS has ranged from 88.32% to 92.49% (Espin et al., 2000; Espin et al., 2008). Throughout existing research, alternate form reliability for CIWS has ranged from .61 to .91, falling in the moderately to relatively strong range (Espin et al., 2000; Espin et al., 2008: McMaster & Campbell, 2008; Weissenburger & Espin, 2005). More specifically, alternate form reliability for a 7-min response period was .80 for students in Grade 10, (Espin et al., 2008), and it ranged from .78 to .87 for students in Grades 3, 5, and 7 (McMaster & Campbell, 2008). Lastly, CIWS correlations with criterion measures have been weak to relatively strong, with coefficients ranging from .56 to .82 (Espin et al., 2005; Espin et al., 2000; Espin et al., 2008).

Design and Data Analyses

Descriptive statistics and norms. Descriptive statistics (i.e., means and standard deviations) for each type of score were calculated using SPSS statistical software. Norms were computed by converting raw scores into percentile ranks using SPSS. To interpret a given student's performance on the CBM-WE measure, the individual raw score was compared to the percentile rank for that score (easyCBM, 2016). For the purposes of the current study, percentile ranks describe where a given student's score falls in relation to his or her peers. For example, a student who receives a score at the 50^{th} percentile will have performed the same or better than 50 percent of peers sampled. A percentile rank of 50 thus reflects the Average performance of a freshman in college at that point in time (easyCBM, 2016). Although percentile ranks for each score were reported, percentile ranks of 16, 50, and 84 are highlighted. Based on a normal curve, the score at the 50^{th} percentile would be both the mean and median among a set of scores, with 50% of scores falling above and 50% falling below this score (Brock, n.d.). Percentiles of 16 and 84 are one standard deviation away from the mean, and scores within this range are typically considered Average (Brock, n.d.) Thus, scores below the 16^{th} percentile are considered to fall below the Average range, while scores above the 84^{th} percentile are considered to fall above the Average range.

Group differences. In order to determine whether there were group differences based on student performance on measures of CBM-WE, analysis of variance (ANOVA) was performed using SPSS. The following assumptions of ANOVAs were evaluated: a) the dependent variable is continuous, b) the independent variable consists of at least two categorical, independent groups, c) observations should be independent of one another, d) there are no significant outliers, e) the dependent variable is approximately normally distributed for each category of the independent variable, and f) homogeneity of variances is present (Laerd, 2018a).

The original variables of interest included gender, ethnicity, age, major of study, first year English composition placement, disability status, first generation college student status, and English learner status. English composition placement was meant to serve as a measure of criterion validity, as it was expected that students' CBM-WE scores would be related to their placement (i.e., students with higher CBM-WE scores are more likely to have been placed in a more difficult first year placement). However, major was the only variable that was analyzed due to large differences in group sizes among the other variables.

CHAPTER 4

Results

Descriptive Statistics and Norms

All analyses for the 116 participants were conducted using the IBM SPSS software. Descriptive statistics for the CWS, IWS, and CIWS scores are reported in Table 1. In three minutes, students produced as many as 120 CWS and as few as 36. Some students did not have any IWS, while others had as many as 28. After accounting for errors, students' CIWS ranged from 7 to 113.

Table 1

Descriptive Statistics for CBM-WE Scores

Score	M	SD	Range
CWS	70.16	15.48	36-120
IWS	6.08	4.57	0-28
CIWS	63.52	17.62	7-113

Percentile ranks for the three types of scores are reported in Table 2. The scores at the 50^{th} percentile for students were 69 CWS, 5 IWS, and 63 CIWS. Scores at the 16^{th} percentile were 54 CWS, 9 IWS, and 45.72 CIWS, while scores at the 84^{th} percentile were 85.28 CWS, 2 IWS, and 80.28 CIWS. Therefore, Average scores ranged from 54 to 85.28 for CWS, 2 to 9 for IWS, and 45.72 to 80.28 for CIWS. Another interpretation is that Average students were able to write between 18 and 28.4 CWS per minute.

Table 2

Percentile Ranks for CBM-WE Scores

Percentiles (1-50)	CWS	IWS	CIWS	Percentiles (51-99)	CWS	IWS	CIWS
1	36.17	27.49	9.55	51	69.00	5.00	63.00

2	37.34	22.96	24.72	52	69.84	5.00	63.84
3	39.53	19.00	30.51	53	70.00	5.00	65.00
4	43.04	16.28	32.36	54	70.00	5.00	65.00
5	44.00	14.15	34.70	55	70.70	5.00	65.00
6	48.00	12.98	36.04	56	72.00	5.00	65.00
7	48.19	12.00	38.00	57	72.69	4.31	65.69
8	49.36	12.00	38.72	58	73.00	4.00	68.58
9	50.00	11.47	41.06	59	73.03	4.00	69.00
10	51.40	11.00	42.00	60	74.00	4.00	69.00
11	52.00	11.00	42.87	61	74.37	4.00	69.00
12	53.00	10.00	43.04	62	75.54	4.00	69.54
13	53.21	10.00	44.00	63	76.00	4.00	70.00
14	54.00	9.62	44.38	64	76.00	4.00	70.88
15	54.00	9.00	45.00	65	76.05	4.00	72.00
16	54.00	9.00	45.72	66	77.00	4.00	72.00
17	54.89	9.00	46.00	67	77.00	4.00	72.39
18	55.00	9.00	47.00	68	77.00	4.00	73.00
19	55.23	8.77	47.23	69	77.73	4.00	73.73
20	56.00	8.00	48.40	70	78.00	4.00	74.00
21	56.57	8.00	49.00	71	79.00	3.00	75.00
22	57.00	8.00	49.00	72	79.00	3.00	75.00
23	57.00	8.00	49.91	73	79.00	3.00	75.00
24	58.08	8.00	51.00	74	79.58	3.00	75.00
25	59.00	8.00	51.25	75	80.00	3.00	75.00
26	59.00	8.00	52.42	76	80.00	3.00	75.92
27	59.59	8.00	53.00	77	80.00	3.00	76.00
28	60.00	8.00	53.00	78	80.00	3.00	76.26
29	60.93	8.00	53.93	79	80.86	3.00	77.00
30	63.00	8.00	54.00	80	82.00	3.00	77.60
31	63.27	7.73	54.00	81	82.00	2.23	78.00
32	64.00	7.00	54.44	82	82.94	2.00	78.94
33	64.00	7.00	55.61	83	83.22	2.00	79.11
34	64.00	7.00	56.78	84	85.28	2.00	80.28
35	64.95	7.00	57.00	85	86.00	2.00	81.45
36	65.12	7.00	58.12	86	86.62	2.00	82.00
37	66.00	7.00	59.00	87	87.00	2.00	82.00
38	66.00	6.54	59.46	88	87.00	2.00	82.00
39	66.63	6.00	60.00	89	88.39	2.00	84.13
40	67.00	6.00	60.00	90	91.00	2.00	85.00
41	67.00	6.00	60.00	91	92.88	1.53	86.41
42	67.00	6.00	61.00	92	95.64	1.00	88.64
43	67.00	6.00	61.00	93	96.00	1.00	89.00
44	67.48	6.00	61.48	94	96.00	1.00	92.92
45	68.00	6.00	62.00	95	98.15	0.85	94.30
46	68.00	6.00	62.00	96	99.32	0.00	96.00
47	68.00	6.00	62.00	97	100.00	0.00	96.49

48	68.16	5.84	63.00	98	103.30	0.00	97.66
49	69.00	5.00	63.00	99	117.45	0.00	110.45
50	69.00	5.00	63.00				

Group Differences

Before running a one-way ANOVA to look at group differences in CBM-WE scores by major, the assumptions of ANOVAs were evaluated. The three dependent variables (i.e., CWS, IWS, and CIWS) were on a continuous scale. The independent variable, student major, consisted of four independent groups (i.e., Early Education, Secondary Education, Education, and Other), and each observation in the sample was independent of one another. Skewness and kurtosis values were within the acceptable range (i.e., -1 to 1) for CWS (skewness = .266; kurtosis = .232) and CIWS (skewness = -.134; kurtosis = .373), but not for IWS (skewness = 1.989; kurtosis = 6.334).

Histograms demonstrated approximately normal distributions for CWS and CIWS. However, the histogram for IWS exhibited a slight positive skew. Further, the Kolmogorov-Smirnov test indicated that CWS [F (116) = .056, p = .200] and CIWS [F (116) = .045, p = .200] did not deviate from normality. The Shapiro-Wilk test similarly demonstrated evidence of normality for CWS [F (116) = .990, p = .571] and CIWS [F (116) = .995, p = .942], but IWS continued to deviate from normality according to the Kolmogorov-Smirnov [F (116) = .147, p = .000] and Shapiro-Wilk tests [F (116) = .844, p = .000]. Although the assumption of normality was violated for IWS, Cohen (2008) asserted that, “the F test for ANOVA is not very sensitive to departures from the normal distribution” because it is considered a “robust” test (p. 360).

According to the Levene’s tests for CWS [F (3, 112) = 2.852, p = .041] and CIWS [F (3, 112) = .620, p = .604], the variances between the four independent variable groups (i.e., Early

Education, Secondary Education, Education, and Other) did not significantly differ from one another. Therefore, the assumption of homogeneity of variances was met for those two independent variables. The assumption of homogeneity of variances was not met for IWS, as the Levene's test for IWS [F (3, 112) = .483, p = .695] indicated that the variances between the four independent variable groups did significantly differ from one another.

As discussed, a one-way ANOVA was performed to determine if students' majors had a significant impact on their CBM-WE scores. Thus, major was included in the ANOVA as the independent variable, while CWS, IWS, and CIWS scores served as the three dependent variables. The results of the one-way ANOVA indicated that CWS [F (3, 112) = .863, p = .463], IWS [F (3, 112) = 5.734, p = .848], and CIWS [F (3, 112) = 1.249, p = .295] scores did not differ significantly between the groups (i.e., Early Education majors, Secondary Education majors, Education majors, and Other majors). Thus, students' major of study did not have a significant impact on their CBM-WE scores.

Chapter 5

Discussion

The purpose of the current study was to develop local CBM-WE norms based on a sample of college freshmen at Penn State. A literature review surrounding the use of CBM indicated that there is a gap in current research related to CBM-WE, and more specifically, the use of CBM-WE with secondary and postsecondary students. Several studies have examined the utility of CBMs for college-age students (Bean & Lane, 1990; Hosp et al., 2014; Hosp et al., 2017; Larson & Ward, 2006; Lewandowski et al., 2003), but none of these studies analyzed the use of CBM-WE for a general, non-disabled population of college students. Based on available research, it was concluded that the most technically adequate CBM-WE scores for the college age group are CWS and CIWS based on a response to an expository prompt. Thus, students in the current study were given 1 min to think and 3 min to respond to the CBM-WE prompt, "Describe your favorite game and why it is your favorite." Their responses were then scored for CWS, IWS, and CIWS.

Subsequently, descriptive statistics and percentile ranks were obtained for the current sample of students' three types of scores. Results indicated that students had a wide variety of writing skills. In the 3 min response period, students wrote between 36 and 120 CWS, 0 to 28 IWS, and 7 to 113 CIWS. Students in the Average range (i.e., 16^{th} to 84^{th} percentiles) wrote approximately 54 to 85 CWS, and they had between 2 and 9 IWS. When accounting for both CWS and IWS, students produced approximately 45 to 80 CIWS. Thus, Average college freshmen in the College of Education at Penn State were able to write roughly 18 to 28 CWS per minute.

In regard to age/grade level, the most relevant norms available for comparison to those obtained from the current study are for students in Grade 8. Hosp et al. (2016) provided 2015 CWS norms from AIMSweb® for students in kindergarten through Grade 8. The median score (i.e., score at the 50th percentile) for students in the fall of Grade 8 was 49. In contrast, the current study indicated that the median score for students in the fall of their freshmen year was 69 CWS. Thus, the median CWS score appears to grow by approximately 20 CWS over the course of the 5 years, or 4 CWS per year, from students' eighth-grade year to their freshman year of college. Because IWS and CIWS scores were not available for the Grade 8 sample, it is unknown whether: a) college freshman simply wrote more than eighth graders or b) college freshman and eighth graders produced responses similar in length, but the college students had fewer errors.

When making this comparison between Grade 8 students and college freshmen, it is important to keep in mind that a sample of eighth graders consists of a broader sample of students than a college freshmen sample. For example, a sample of eighth-grade students includes both college bound and non-college bound students. Thus, the variability in writing achievement is likely much wider, and the median CBM-WE scores may be lower overall, than if the sample had only included college-bound eighth graders.

In addition to examining descriptive statistics and percentile ranks as part of the current study, the reliability of the students' CWS, IWS, and CIWS scores was also assessed for 20% of the responses. One way that reliability was measured in the study was through the calculation of point-by-point agreement, which accounts for all agreements and disagreements between two raters for a given probe. Point-by-point agreement fell in the acceptable range for the entire subset of 24 probes that were scored by two raters. Thus, this provides evidence that the scoring guidelines used in the current study were effective in producing reliable CBM-WE scores among the three raters.

Another way that reliability was measured in the current study was through the calculation of interscorer agreement. Average percent agreement for the CWS and CIWS scores was in the acceptable range (i.e., above 80%). Previous research has similarly shown that interscorer agreement for CWS and CIWS has been high, ranging from 86% to 98% for CWS (Espin et al., 1999; Espin et al., 2008; Gansle, et al., 2002; Gansle et al., 2006; Videen et al., 1982) and 88.32% to 92.49% for CIWS (Espin et al., 2000; Espin et al., 2008). Although the reliability of CWS and CIWS scores in the current study was acceptable, average percent agreement for IWS was below the acceptable range, likely because most responses had few IWS. No previous studies examining interscorer agreement for IWS were found, but other measures of reliability (i.e., interrater reliability, alternate form reliability, and correlations with criterion measures) were similarly lower for IWS, ranging from weak (.30) to relatively strong (.87; Espin et al., 2000; Watkinson & Lee, 1992).

A final goal of the current study was to examine various demographic variables to determine whether there were significant group differences in writing performance, as measured by scores on the CBM-WE task. Although information was collected on students' gender, ethnicity, age, major of study, first year English composition placement, disability status, first generation college student status, and English learner status, major was the only variable of interest that was examined due to large differences in group sizes among the remainder of the variables. A one-way ANOVA indicated that CBM-WE scores did not differ significantly based on students' major of study (i.e., Early Education, Secondary Education, Education, and Other). A possible explanation for the lack of findings is that there was little variability in student majors since all students in the sample were enrolled in the College of Education. Limitations and directions for future research related to group differences in CBM-WE performance will be discussed following the possible implications from the current paper.

Potential Implications

As noted previously, the primary goal of the current study was to develop CBM-WE norms using scores from a population of college students. The purpose of developing these norms was twofold: 1) to aid in the identification of learning disabilities in written expression among college students and 2) to provide a benchmark of the writing skills of college students for current high school students who wish to attend Penn State or other similar universities. These two implications will subsequently be discussed in greater detail.

Although research has shown that students with disabilities are generally less likely to enroll in college than their non-disabled peers (Hosp et al., 2014), over time it has become more common for students with disabilities, especially learning disabilities, to attend colleges and universities (Lewandowski et al., 2003). This has resulted in more students requesting and receiving accommodations (e.g., additional time on tests, use of reading software, note-taking assistance, etc.) in and out of the classroom. In most cases, colleges and universities require documentation of a student's disability in order for the student to receive services.

For instance, in order to be eligible for disability services, the Penn State Student Disability Resources (SDR) requires a student to provide documentation (e.g., neuropsychology evaluation, psychoeducational evaluation, medical evaluation, etc.) that he or she has a disorder or impairment that meets the definition of a disability according to the definitions set forth by Title II of the Americans with Disabilities Act, Amendments Act (ADAAA) of 2008 and Section

504 of the Vocational Rehabilitation Act of 1974. According to Section 504 of the Vocational Rehabilitation Act, a disability is "a physical or mental impairment which substantially limits one or more major life activities." Title II of ADAAA prohibits discrimination against individuals with disabilities, and it requires public and private institutions to make reasonable accommodations to allow students with disabilities full participation in the same programs and activities available to students without disabilities.

In regard to learning disabilities, Penn State SDR requires that a student submit a comprehensive neuropsychological or psychoeducational evaluation completed by a psychologist or neuropsychologist. The specific guidelines Penn State has for learning disabilities are adapted from the Diagnostic and Statistical Manual – Fifth Edition (American Psychiatric Association, 2013). According to these guidelines, "Individuals with learning disorders have persistent difficulty learning or performing academic skills at a level commensurate with their intelligence and age."

The guidelines further require the following elements: a) Evidence of persistent learning difficulties causing academic performance below expectations despite targeted intervention, b) Diagnosis of learning disorder occurring after onset of school years even if learning difficulties were apparent prior to school-age, c) Documented areas of academic skills difficulties as measured by objective and statistically sound aptitude and achievement assessments that are reported in terms of specific subtests and standard scores, d) Learning difficulties must interfere with or reduce the quality of functioning, whether academically, socially, occupationally, or other area of functional impairment, e) Learning difficulties cannot be attributed to or better explained by another diagnosis or environmental factor, including but not limited to intellectual disability, mental disorder, sensory impairment, neurological disorder, psycho-social difficulty, language difference, or lack of access to adequate instruction, and f) Learning difficulties and associated functional limitations in the academic environment and possibly other settings should warrant reasonable accommodations, which are presented in terms of a summary and recommendations (i.e., learning difficulty and suggested reasonable accommodation to mitigate learning difficulty).

These guidelines differ considerably from those used to identify school-age students with learning disabilities. Learning disabilities among K-12 students are instead defined by the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004. The IDEIA defines a specific learning disability as:

> a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia…Specific learning disability does not include learning problems that are primarily the result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage" (p. 46757).

Whereas public K-12 schools have the responsibility of evaluating students and identifying educational disabilities according to IDEIA federal definitions, college students themselves have the responsibility of providing documentation of their disabilities in order to receive services. However, if a student does not have documentation because they have not previously been evaluated for a disability, he or she may have the option of seeking out an evaluation directly from his or her college or university. For example, the Herr Clinic at Penn State provides school psychological services not only to school age children, but also to college-age students. College students may refer themselves for psychoeducational evaluations due to academic (i.e., reading, writing, or math) difficulties, with the purpose of determining whether they have a learning disability. It is common for students with learning disabilities to have skill deficits in written expression (Diercks-Gransee et al., 2009; McMaster et al., 2009), and it is important to identify these deficits, as writing skills are related to success in higher education and the workplace (Huot, 2002; Kellogg & Raulerson, 2007; Troia & Olinghouse, 2013).

Written expression skills are most commonly assessed through standardized achievement tests. However, it is best practice to assess a student's current academic skills using multiple sources of information (McMaster & Espin, 2007; Runge et al., 2017; Troia & Olinghouse, 2013). One way to corroborate evidence of academic skill deficits obtained through a standardized achievement test is to gather additional data through CBMs. As mentioned, this is one of the main potential uses of the CBM-WE norms developed as part of the current study; the norms can be used to aid in measuring freshmen Penn State students' proficiency in written expression to determine whether a learning disability exists.

For example, if a student is evaluated at the Penn State Herr clinic to determine whether he or she has a learning disability related to writing skills, the student would be administered a standardized achievement test that includes assessment of writing skills since the SDR guidelines require that academic skill difficulties are "measured by objective and statistically sound aptitude and achievement assessments that are reported in terms of specific subtests and standard scores." The student could also complete a CBM-WE probe to provide additional evidence that the student is demonstrating "academic performance below expectations," another component of the SDR guidelines. The student's response would then be scored for CWS, IWS, and CIWS.

Consequently, the student's scores would be compared to the norms from the current study to determine the percentile ranks at which his or her scores fall. The percentile rank would provide information regarding how the student compares to peers of the same academic standing (i.e., other freshmen at Penn State). Typically, scores at or below the 10^{th} percentile are considered to be evidence for underperformance in Pennsylvania. Once skill deficits are identified, the deficits can then be addressed through additional supports and accommodations, which may help prevent negative student outcomes (e.g., drop out; Larson & Ward, 2006; McMaster et al., 2009).

In addition to aiding in the identification of learning disabilities at the college age, the CBM-WE norms from the current study could be useful for secondary educators and their students who are looking to attend Penn State or a similar university. Espin et al. (2012) asserted that CBMs could be used to: a) determine the level of academic performance needed to succeed in college and b) monitor progress toward that level of performance. Thus, the norms from the current study could serve as a benchmark, providing educators and prospective students with information on how the average Penn State freshman performs on a CBM writing task.

For instance, a high school teacher could administer a CBM-WE task, similar to the one completed in current study, to a student. The teacher could then score the response and compare the student's CWS, IWS, and CIWS scores to the present norms to determine the student's percentile rank for each type of score. This would allow the teacher to determine whether the student's writing skills are similar to, above, or below the average freshman at Penn State. If the student's performance is below that of the average freshman, growth may be necessary in order to be admitted to such a university since writing skills are often considered as part of the admission process.

If the student receives special education services, the CBM norms from the current study could additionally be used by educators to write IEP goals related to writing for the student. If the student is old enough (e.g., age 14 or above in Pennsylvania), the norms could also be used to help develop a transition plan. A transition plan is a portion of the IEP that focuses on the student's plans for after graduation, such as attending a specific university (e.g., Penn State). Several important components for addressing the secondary transition process include: describing the student's present levels of academic performance, determining measurable annual goals that address academic skill deficits and lead to postsecondary goals, and monitoring progress based on data. (PaTTAN, 2018).

As noted above, an educator would first administer a CBM-WE task to a student to help measure the student's current academic performance in the area of written expression, as compared to a sample of Penn State freshmen. The student's relative performance would then help guide the student's writing goal. For example, a potential goal might be for the student to produce scores at or above the 50^{th} percentile from the sample. In this case, the student's goal would be to produce 69 or more CWS, 5 or fewer IWS, and 63 or more CIWS when given 1 min to think and 3 min to respond to an expository CBM-WE probe.

Lastly, CBM-WE probes could be administered to the student throughout the school year on a weekly or bi-weekly basis to determine whether the student is making progress toward the writing goal. If the student is ultimately unable to meet the goal, this may be a method by which to document that despite targeted intervention, writing continues to be a significant academic deficit, and the student will be in need of supports and accommodations within the college setting. In contrast, if the goal is met, the student can conclude that he or she is performing the same or better than 50% of a sample of freshmen at Penn State. Thus, the student is performing in the Average range compared to Penn State freshmen. As noted subsequently in the Future Research section, more research is needed related to the concurrent and predictive validity of CBM-WE for older students; however, a high school student that performs in the Average range compared to college freshmen may be likely to demonstrate Average overall writing skills during his or her freshman year.

An important consideration in using the current CBM-WE norms to measure high school student writing performance, set writing goals for students with disabilities, and monitor students' progress toward their goals is that postsecondary educators must first be aware of and have access to these norms. The current norms and their implications could be communicated to educators directly or the information could be communicated to school psychologists, who could then bring the information back to their individual schools or districts. One potential way to disseminate the results of the current research is through publishing in newsletters such as the Communique, which is circulated by the National Association of School Psychologists (NASP); journals such as the Phi Delta Kappan, which serves professional educators; or social networking outlets such as Schoology, which is a learning management system for teachers that has various resources and articles available. The results could also be presented at conferences such as the NASP annual conference or the annual meeting of the Association of Teacher Educators.

Limitations

While the current study fulfilled the purpose of developing college freshmen CBM-WE norms for use with both secondary and post-secondary students, there are several limitations to consider related to the choice of the CBM-WE task, the assumptions and reliability of the IWS scores, and the variability of the sample. When considering which CBM-WE task to administer in the current study, the decision to use an expository probe scored for CWS, IWS, and CIWS was based on research with secondary students, as there is not currently research available related to the use of CBM-WE with a general population of college students. Larson and Ward (2006) posited that a literature review of CBM at the secondary level provides:

> a partial background for studying CBM measures for university students. This use of the CBM research should, however, be used with caution because the two populations may be sufficiently different so that what is known about secondary students may not hold true with university students (p. 48).

Thus, in regard to age/grade, CBM research with high school students is the most relevant literature available to base decisions for research with the college age group. However, as Larson and Ward stated, results should be interpreted with caution because more research is needed to determine whether the same tasks and scores used with high school students are also technically adequate for college students.

Although CBMs were originally intended to be directly tied to course curriculum, a potential criticism of current CBMs are that they are not always related to the curriculum within a specific course. For instance, when CBMS are being used for progress monitoring purposes or to compare student performance to a national sample of same-age peers, the probes are not often created from specific course content, as course curriculum differs considerably in schools across the nation. Instead, programs such as AIMSweb®, created by test developers such as Pearson, provide both CBM probes and corresponding norms. Similarly, within the college setting, it is difficult to tie CBM-WE probes directly to the curriculum, as content areas vary widely. For instance, an expository prompt related to the content within an early childhood education course would be vastly different from a prompt related to the content within a molecular biology course. Therefore, a limitation of the current study is that the specific prompt used (i.e., Describe your favorite game and explain why it is your favorite) was not tied to the freshmen seminar course content.

In addition to determining the type of prompt (i.e., narrative vs. expository) and the type of scores to use in the current study, the amount of time given to students to complete their written responses was also a consideration prior to data collection. As noted in the CBM-WE high school research section of the current literature review, there is evidence that longer writing samples of 5 to 10 min increases the technical adequacy of CBM-WE measures for secondary students (Espin, 2012; Espin et al., 2004; Weissenburger & Espin, 2005). Therefore, a limitation of the current study was that the response interval given was only 4 min (i.e., 1 min to think and 3 min to write). However, this response period was ultimately chosen because national CBM-WE norms for school-age children are based off of a 3 min response period. Using a 3 min response period in the current study allows for more accurate comparisons of scores between college and school-age samples.

When examining the ANOVA assumptions and the reliability of scores as part of statistical analyses for the current study, a few issues with the IWS variable were noted. Normality was violated for the IWS variable, as indicated by skewness and kurtosis values, a histogram, and Shapiro-Wilk and Kolmogorov-Smirnov tests. However, as noted, Cohen (2013) asserted that F-tests are robust, meaning that they are not very sensitive to departures from normality. The assumption of homogeneity of variances was also violated for the IWS variable, as indicated by the Levene's test. Further, the original average percent agreement was lower for

IWS scores compared to CWS and CIWS scores due to the small number of IWS present for most responses.

Finally, the population in the current study was generally homogenous. It was limited to a convenience sample of students in the college of education. Therefore, the sample was not nationally representative, and participants were not randomly selected. The vast majority of students were female, White, 18 years old, and their first language was English. Further, most students were enrolled in the ENGL 15 course (Rhetoric and Composition), they did not have a disability, and they were not first generation college students. Thus, due to the lack of variability among these variables (i.e., gender, ethnicity, age, first language, English course, disability status, and first generation college student status), they were unable to be analyzed as part of the current study. In addition, because the variability of the population was limited, it is possible that the results would not generalize to students outside of these demographics.

Directions for Future Research

Several ideas for future research were generated as a result of the current study, including expanding the variability of the sample, examining the technical adequacy of CBM-WE for college students, and creating additional CBM norms based on high school and college populations. Due to the lack of diversity in the sample from the current study, a goal for future research would be to diversify the sample in order to examine additional variables of interest. This could be done by recruiting students of varying demographics (e.g., gender, ethnicity, age, major, freshman English course, English Learner status, disability status, first generation college student status, etc.). The sample could also be expanded to include students from other state universities.

As noted, the CBM prompt from the current study was not particularly contextualized within the college writing curriculum. However, if the sample were expanded to include students from various majors and courses, an idea for future research could be to give students writing prompts that are in fact curriculum-based, or related to the content of the specific course in which they are being assessed. For instance, perhaps if students in an English course were reading a novel, they could be prompted to write about their favorite character from the novel and why that character is their favorite.

In regard to disability status, students from the current study were asked if they currently received services through Student Disability Resources. However, future research might ask participants to additionally indicate if they received accommodations and/or special education services prior to attending college. Given that the criteria to qualify for a disability differ considerably between K-12 and college age students, there may be students who have a disability and received school-age services, but who do not qualify for or require additional supports in college.

Expanding the sample would result in greater external validity and generalizability to the larger college freshmen population. In addition, if group sizes among the demographic variables are sufficient in size, research could analyze group differences to determine whether any of these factors have an effect on students' written expression skills, as measured by CBM-WE scores. For example, previous research has shown that gender is one of the predictors of students' quality of writing (Olinghouse, 2008). A study examining the writing skills of college students concluded that both gender and major had an impact on writing skills, with women outperforming men and humanities/social sciences majors outperforming natural sciences/engineering majors (Oppenheimer, Zaromb, Pomerantz, Williams, and Park, 2017).

Another direction for future research would be to further examine the reliability and validity of CBM-WE scores for college students. More specifically, future research should examine whether the CBM scores from the current study (i.e., 3 min expository CBM-WE prompts scored for CWS, IWS, and CIWS) are valid and reliable measures of overall writing proficiency for students at the post-secondary level. Although the current study utilized one measure of reliability (i.e., interrater reliability), other types such as alternate forms and test-retest should be examined in the future, as prior CBM research has focused on these three forms of reliability.

As noted previously, a potential criticism of CBM-WE is that it likely is not the most authentic method by which to assess college students' writing. For instance, the tasks are not contextualized within a particular course or content area. In addition, the scoring criteria vary greatly from those typically used within the college writing curriculum, where components such as discourse and deliberation are often integrated. However, research with school-age students has provided evidence that CBM-WE scores are correlated with the scores of other, more comprehensive writing assessments. The scores are also able to predict students' general writing proficiency. Thus, if CBM-WE scores were correlated with scores on other writing measures and able to predict performance on other writing measures for college students, this would provide evidence of concurrent and predictive validity, as well as additional support for the use of CBM-WE with this age group.

For example, one of the original goals of the current study was to determine whether students' English course had an effect on their CBM-WE scores since placement in Penn State freshman English courses is based on individual performance on an English proficiency exam. However, the English course variable could not be analyzed, as 94.8% of the sample of students was enrolled in the same course (ENGL 15). Future research might be able to examine the ability of CBM-WE scores to predict English course placement if the sample contains a more equal number of students in each English course. Other ways to assess the predictive validity of CBM-WE scores could be to examine variables such as students' final freshman English course grades, SAT writing scores, or scores on an individual standardized writing assessment (e.g., the Written Expression composite on the Wechsler Individual Achievement Test – Third Edition).

An additional idea for future research would be to develop additional CBM norms for both high school and college populations. As noted previously, national CBM norms currently exist up through Grade 8. The development of national reading, writing, and math CBMs for the general population of 9^{th} to 12^{th} graders would be useful for measuring overall academic performance, identifying learning disabilities, and progress monitoring among this age group. Future research could additionally develop CBM norms for both the college-bound and non-college-bound populations of high school students to determine if scores differ between these two groups. General and college-bound high school norms would also allow for more useful comparisons to the scores of college freshmen, as the students would be closer in age and grade than the Grade 8 sample currently available.

In terms of the college population, CBM-WE norms could also be created for sophomores, juniors, and seniors. In addition, CBM math and reading norms could be developed for each cohort of students. Similar to the norms from the current study, these additional norms could then also be used to: a) provide a benchmark of the reading and math skills of college freshmen for college-bound high school students and b) determine whether college students of all academic standings have learning disabilities in the areas of reading, math, and writing, should they request an evaluation. Further, future research could examine college student CBM-

WE scores at the beginning, middle, and end of each year to determine the typical rate of growth throughout a given year, from one year to another, or across a student's college career.

A final idea for future research would be to administer CBM-WE probes using a word processor to determine whether there are significant differences between scores obtained through paper and pencil responses and scores obtain through word-processed responses. Lovett, Lewandowski, Berger, and Gathje (2010) examined whether response mode (i.e., handwritten vs. word-processed) had an impact on college students' writing ability. The study concluded that students who used a word processor wrote significantly more than their peers who handwrote their essays. However, the quality of the writing was not significantly different between the two groups.

Similar to the word processing program used by Lovett et al. (2010), future research would need to utilize a program that does not have spell-check or other grammatical aids for CBM-WE scores to still be technically adequate. Use of word processing programs may provide a more accurate measure of college students' writing skills, as previous research has shown that word processors are associated with better writing quality and quantity (Goldberg, Russell, & Cook, 2003). Further, if there are differences in word-processed versus hand-written scores, professors and universities might consider offering word-processors to students with learning disabilities as an accommodation for writing assignments.

Concluding Remarks

The current paper has extensively examined the use of curriculum-based measurement in written expression for college-age students. As previously reviewed, curriculum-based measurement is a type of assessment commonly used among school-age students to measure academic proficiency related to reading, writing, and math skills; monitor progress within these academic areas; and aid in the identification of specific learning disabilities. Currently, the research surrounding CBM is lacking in several areas, including written expression and the utility of CBM-WE scores with secondary and postsecondary students. The current study aimed to narrow this research gap by developing CBM-WE norms based on a sample of freshmen college students at The Pennsylvania State University.

With the development of these norms, there are two main potential implications for practice within the field of education. The first implication is that the norms can be used to provide a benchmark of the writing skills of typical college freshmen at universities similar to Penn State. Using the norms, students can compare their current writing proficiency to that of typical Penn State freshmen and can set goals for monitoring their progress. The second implication is that the norms can be used as a method of assessing college students' proficiency in written expression to help identify whether a learning disability in written expression exists. Although this paper contributes to the gap in the literature related to CBM-WE for college students, future research is needed to further assess the technical adequacy of CBM-WE for this age group; examine group differences based on student demographic information; and develop additional norms at the secondary and postsecondary level.

References

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

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Americans with Disabilities Act Amendments Act of 2008, Pub. L. No. 110-325.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education, 77*, 887-924.

Baker, S., & Hubbard, D. (1995). Best Practices in the assessment of written expression. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology III* (pp. 717–728). Bethesda, MD: National Association of School Psychologists.

Bean, R. M., & Lane, S. (1990). Implementing curriculum-based measures of reading in an adult literacy program. *Remedial and Special Education, 11*, 39-46.

Berninger, V. W., Rutberg, J. E., Abbott, R. D., Garcia, N., Anderson-Youngstrom, M., Brooks, A., & Fulton, C. (2006). Tier 1 and tier 2 early intervention for handwriting and composing. *Journal of School Psychology, 44*, 3-30.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29,* 137-164.

Bracken, B. A. (2007). *Bracken School Readiness Assessment: Examiner's manual.* (3rd ed.). San Antonio, TX: NCS Pearson, Inc.

Breaux, K. C., & Frey, F. E. (n.d.). *Assessing writing skills using correct-incorrect word sequences: A national study.* Retrieved from https://images.pearsonclinical.com/images/products/wiat-iii/WIAT-III_NASP_Poster.pdf

Brock, S. E. (n.d.). *Descriptive statistics in psychological testing.* Retrieved from http://www.csus.edu/indiv/b/brocks/courses/eds%20250/eds%20250/handouts/11/descrptive%20statistics%20and%20the%20normal%20curve.pdf

Calfee, R. C., & Miller, R. G. (2013). Best practices in writing assessment for instruction. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction, 2nd Edition* (pp. 351-377). New York: Guilford Press.

Campbell, H., Espin, C. A., & McMaster, K. (2012). The technical adequacy of curriculum-based writing measures with English learners. *Reading and Writing, 26*, 431-452.

Cohen, B. H. (2008). *Explaining psychological statistics* (3rd ed.). John Wiley & Sons, Inc.: Hoboken, NJ.

The College Board. (2018). *SAT essay scoring.* Retrieved from https://collegereadiness.collegeboard.org/sat/scores/understanding-scores/essay

Conference on College Composition & Communication Committee on Assessment. (2014). *Writing assessment: A position statement.* Retrieved from http://www.ncte.org/cccc/resources/positions/writingassessment

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

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

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19,* 1-16.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48,* 368-37.

Deno, S. L., Marston, D., & Tindal, G. (1986). Direct and frequent curriculum-based measurement: An alternative for educational decision-making. *Special Services in the Schools, 2,* 5-27.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). *Relationships among simple measures of written expression and performance on standardized achievement tests* (Vol. IRLD-RR-22). University of Minnesota, Institute for Research on Learning Disabilities.

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

easyCBM. (2016). *Interpreting easyCBM Test Results.* Retrieved from https://easycbm.com

Elliott, S. N., & Bretzing, B. H. (1980). Using and updating local norms. *Psychology in the Schools, 17,* 196-201.

Elliott, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review, 26*, 224-234.

Espin, C. A., De La Paz, S., Scierka, B. J., & Roelofs, L. (2005). The relationship between curriculum-based measures in written expression and quality and completeness of expository writing for middle school students. *The Journal of Special Education, 38*, 208-217.

Espin, C. A., McMaster, K. L., Rose, S., & Wayman, M. (Eds.). (2012). *A measure of success: The influence of curriculum-based measurement on education.* University of Minnesota Press: Minneapolis, MN.

Espin, C. A., Scierka, N. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 15*, 5-27.

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

Espin, C., Wallace, T., Campbell, H., Lembke, E. S., Long, J. D., & Ticha, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Council for Exceptional Children, 74*, 174-193.

Espin, C. A., Weissenburger, J. W., & Benson, B. J. (2004). Assessing the writing performance of students in special education. *Exceptionality, 12*, 55-66.

Fewster, S., & Macmillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23,* 149-156.

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

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy. *American Educational Research Journal, 21,* 449-460.

Fuchs, L. S., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30, 1-16.*

Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment responders. *Peabody Journal of Education, 77,* 64-84.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (2015). Republication of "Curriculum-based measurement: A standardized, long-term goal approach to monitoring student progress." *Intervention in School and Clinic, 50*, 185-192.

Fuchs, L. S., Fuchs, D., & Stecker, P. M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities, 22,* 51-59.

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

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,* 435-460.

Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment, 8*, 1-26.

Gickling, E. E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology III* (pp. 587-595). Bethesda, MD: National Association of School Psychologists.

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment, 2,* 1-51.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879-896.

Habadank, L. (1995). Developing local norms for problem solving in the schools. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology III* (pp. 701–716). Bethesda, MD: National Association of School Psychologists.

Hassel, H., & Giordano, J. B. (2015). The blurry borders of college writing: Remediation and the assessment of student readiness. *College English*, *78*, 56–81.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*, 45-56.

Hosp, J. L., Ford, J. W., Huddle, S. M., & Hensley, K. K. (2018). The importance of replication in measurement research: Using curriculum-based measures with postsecondary students with developmental disabilities. *Assessment for Effective Intervention, 43*, 96-109.

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

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

Huot, B. (2002). *(Re)Articulating Writing Assessment for Teaching and Learning.* Logan, UT: Utah State University Press.

Individuals with Disabilities Education Improvement Act, H.R. 1350, 108th Congress (2004).

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.

Kamphaus, R. W. (1984). Developing local norms for individually administered tests. *School Psychology Review*, *13*, 491-498.

Kellogg, R. T., & Raulerson, B.A., III. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review, 14*, 237-242.

Kulikowich, J. M., Mason, L. H., & Brown, S. W. (2008). Evaluating fifth- and sixth-grade students' expository writing: Task development, scoring, and psychometric issues. *Reading and Writing, 21*, 153-175.

Laerd Statistics. (2018a). *One-way ANOVA in SPSS statistics.* Retrieved from https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

Laerd Statistics. (2018b). *Pearson's product-moment correlation using SPSS statistics*. Retrieved from https://statistics.laerd.com/spss-tutorials/pearsons-product-moment-correlation-using-spss-statistics.php.

Larson, W. C., & Ward, C. (2006). Developing a predictor of university students' course grades using curriculum-based measurement: An initial investigation. *Journal of Reading and Learning, 37*, 43-57.

Learning Express. (2003). *501 Writing Prompts.* LearningExpress, LLC: New York, NY.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary schools students. *Assessment for Effective Intervention, 28,* 23-35.

Lewandowski, L. J., Codding, R. S., Kleinmann, A. E., & Tucker, K. L. (2003). Assessment of reading rate in postsecondary students. *Journal of Psychoeducational Assessment, 21,* 134-144.

Lovett, B. J., Lewandowski, L. J., Berger, C., & Gathje, R. A. (2010). Effects of response mode and time allotment on college students' writing. *Journal of College Reading and Learning, 40*, 64-79.

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40,* 379-390.

Marston, D., & Deno, S. (1981). *The reliability of simple, direct measures of written expression* (Tech. Rep. No. 50). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*, 266-276.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18,* 109-117.

Marston, D., Tindal, G., & Deno, S. L. (1982). *Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification.* University of Minnesota, Institute for Research on Learning Disabilities.

McMaster, K. L., & Campbell, H. (2008). New and existing curriculum-based writing measures: Technical features within and across grades. *School Psychology Review, 37*, 550-566.

McMaster, K. L., Du, X., & Petursdottir, A. L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities, 42*, 41-60.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*, 66-84.

Merrigan, T. E. (2012). *Alternative methods to curriculum-based measurement for written expression: Implications for reliability and validity of scores.* Retrieved from https://search.proquest.com/docview/1348169148?pq-origsite=gscholar

National Center for Education Statistics. (2003). *The nation's report card: Writing 2002.* Retrieved from https://files.eric.ed.gov/fulltext/ED476189.pdf

National Commission on Writing for America's Families, Schools, and Colleges. (2004). *Writing: A ticket to work or a ticket out. A survey of business leaders.* Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing, 21*, 3-26.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545-565.

O'Neill, P., Moore, C., & Huot, B. (2009). *A Guide to College Writing Assessment.* Logan, UT: Utah State University Press.

Oppenheimer, D., Zaromb, F., Pomerantz, J. R., Williams, J. C., & Park, Y. S. (2017). Improvement of writing skills during college: A multi-year cross-sectional and longitudinal study of undergraduate writing performance. *Assessing Writing, 32,* 12-27.

Parker, R., Tindal, G., & Hasbrouck, J. (1991a). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality, 2*, 1-17.

Parker, R. I., Tindal, G., & Hasbrouck, J. (1991b). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children, 58,* 61-73.

Pennsylvania Training and Technical Assistance Network (2018). *What is secondary transition?* Retrieved from https://www.pattan.net/Graduation-Post-Secondary-Outcomes/Educational-Initiatives/What-is-Secondary-Transition

Pearson (2009). *Wechsler Individual Achievement Test – Third Edition (WIAT-III): Examiner's manual.* San Antonio, TX: NCS Pearson, Inc.

Pearson (2014). *AIMSweb*. Retrieved from www.aimsweb.com

Powell-Smith, K. A., & Shinn, M. R. (2004). *Administration and scoring of written expression curriculum-based measurement (WE-CBM) for use in general outcome measurement.* Retrieved from https://www.aimsweb.com/wp-content/uploads/written-expression-cbm-manual.pdf

Relles, S. R., & Tierney, W. G. (2013). The challenge of writing remediation: Can composition research inform higher education policy? *Teachers College Record*, *115*, 1–45.

Ricks, J. H. (1971). *Local norms - When and why.* The Psychological Corporation: New York, NY.

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100*, 879-906.

Runge, T., Menko, J., Young, A., Paulson, S., McCracken, D., & Meiss, K. (2017, October). *MTSS for writing: Pennsylvania's model illustrated in a case study.* Presentation at the annual conference of the Association of School Psychologists of Pennsylvania, State College, PA.

Section 504 of the Rehabilitation Act of 1973, 34 C.F.R. Part 104.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention* (4th ed.). New York, NY: Guilford.

Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education decision making. *School Psychology Review, 17*, 61-80.

Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *The Guilford School Practitioner Series. Curriculum-Based Measurement: Assessing Special Children* (pp. 90-129). New York, NY, US: Guilford Press.

Shinn, M. R. (1998). *Advanced applications of Curriculum-Based Measurement*. New York, NY, US: Guilford Press.

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (4th ed., pp. 671-698). Bethesda, MD: National Association of School Psychologists.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. *Focus on Exceptional Children, 24*, 1-20.

Shinn, M., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6,* 31-38.

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

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

Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *The Journal of Special Education, 23*, 169-183.

Tindal G., & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research & Practice, 6,* 211-218.

Troia, G. A., & Olinghouse, N. G. (2013). The common core state standards and evidence-based educational practices: The case of writing. *School Psychology Review*, *42,* 343-357.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education.* Retrieved from https://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

Videen, J., Marston, D., & Deno, S. L. (1982). *Correct Word Sequences: A Valid Indicator of Proficiency in Written Expression* (Vol. IRLD-RR-84, pp. 61). Minnesota University, Minneapolis Institute for Research on Learning Disabilities.

Watkinson, J. T., & Lee, S. W. (1992). Curriculum-based measures of written expression for learning-disabled and nondisabled students. *Psychology in the Schools, 29*, 184-191.

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 from http://www.jimwrightonline.com/pdfdocs/cbaManual.pdf

Wright, J. (2013). *Curriculum-based measurement-written expression: Guidelines for use.* Retrieved from http://www.interventioncentral.org/sites/default/files/pdfs/pdfs_blog/wright_Learning_Spark_Blog_29_March_2013_Grs_1_5_Identfy_Writing_Difficulties_Instructions.pdf.

Appendix A

Institutional Review Board (IRB) Approval Letter

PennState

Office for Research Protections
Vice President for Research
The Pennsylvania State University
205 The 330 Building
University Park, PA 16802

814-865-1775
Fax: 814-865-8699
orp@psu.edu
research.psu.edu/orp

EXEMPTION DETERMINATION

Date: November 19, 2018

From: Courtney Whetzel, IRB Analyst

To: Victoria Buser

Type of Submission:	Initial Study
Title of Study:	Expanding the use of Curriculum-Based Measurement for written expression: A college freshmen norming study
Principal Investigator:	Victoria Buser
Study ID:	STUDY00010827
Submission ID:	STUDY00010827
Funding:	Association of School Psychologists of Pennsylvania
Documents Approved:	• ASPP Grant Application (1), Category: Sponsor Attachment • CBM-WE Probe (1), Category: Data Collection Instrument • Demographics Survey (1), Category: Data Collection Instrument • HRP-591 - Protocol for Human Subject Research (2), Category: IRB Protocol

The Office for Research Protections determined that the proposed activity, as described in the above-referenced submission, does not require formal IRB review because the research met the criteria for exempt research according to the policies of this institution and the provisions of applicable federal regulations.

Continuing Progress Reports are **not** required for exempt research. Record of this research determined to be exempt will be maintained for five years from the date of this notification. If your research will continue beyond five years, please contact the Office for Research Protections closer to the determination end date.

Changes to exempt research only need to be submitted to the Office for Research Protections in limited circumstances described in the below-referenced Investigator Manual. If changes are being considered and there are questions about whether IRB review is needed, please contact the Office for Research Protections.

Penn State researchers are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within CATS IRB (http://irb.psu.edu).

This correspondence should be maintained with your records.

Appendix B

Data Collection Script

1. Introduce yourself and the study with the following script: ***"Hello, my name is Tori Buser, and I am a student in the school psychology graduate program here at Penn State. For my dissertation, I will be measuring the writing skills of college freshmen at Penn State through a brief, 4-minute writing exercise. Participation in this exercise is completely voluntary, and whether you decide to participate will in no way affect your grade in this class. If you choose to participate in the writing exercise, you will receive 1 point of extra credit and you will be placed into a drawing for five different $10 Amazon gift cards, as long as you write your email address on this piece of paper after turning in your materials to me. If you do not want to participate, but still want the point of extra credit, you can complete the same task, but turn the papers in to your professor to be shredded. If you would like to participate and/or earn the extra credit, please stay seated and take out a pen or pencil."***
2. Say these specific directions to the students: ***"Thank you for agreeing to participate in the study. Once I pass out the materials, please fill out the demographic questionnaire on page 1. When you are finished, put your pen/pencil down. Please don't flip ahead to the next page."***
3. Provide each participant with a packet including a demographic questionnaire and response sheets.
4. Once all students have finished filling out the demographic section, say these specific directions to the students: ***"Please turn to page 2. In a moment, you are going to respond to the prompt on your paper. You will have 1 minute to think about what you***

will write and 3 minutes to write your response. Please be sure to write legibly so that your responses can be read at a later time. Are there any questions before we begin? For the next minute, think about what you will write. [Allow 1 min]***. Now begin writing. Please remember to write as neatly as possible.*** [Allow 3 min]***. Stop writing and put your pen or pencil down. We are finished, so please give me your packet before you leave the room if you have agreed to participate in the study. If not, please give the packet to your professor. Also, if you turned your paper in to me and wish to be placed into the drawing for an Amazon gift card, please write down your email address on this paper before leaving. Thank you all again for your participation."***

Appendix C

Scoring Guidelines

1. Circle words that are spelled incorrectly to help determine pairs of correct adjacent words.
2. Place a caret (^) between writing units that are grammatically, mechanically, semantically, and syntactically correct.
3. Place a dot (·) between written units that are incorrect.
4. Sum the number of carets for the CWS score, and sum the number of dots for the IWS score. Subtract your IWS score from the CWS score to obtain the CIWS score.

General Scoring Rules:

- A single caret or dot should be placed between each word sequence. Word sequences can include two words, as well as a word and an ending punctuation mark. A dot or caret should be placed before the first word in the writing sample, but not after the last word or punctuation mark.
- Do not penalize for errors that appear to be due to poor penmanship.
- Apostrophes are required if the word cannot stand alone without it.
- Do not penalize for commonly used abbreviations that are spelled correctly, with or without abbreviation periods (e.g., USA). Abbreviations must be capitalized appropriately. Penalize for abbreviations that would not be acceptable in formal writing (e.g., b/c, w/o, etc.).
- Do not penalize for failing to hyphenate words or using optional hyphenations (e.g., She is a six-year-old), unless a word boundary error is made. Incorrectly placed hyphens

should be penalized (e.g. She is six-years-old). If a word is hyphenated, a dot or caret should be placed between each individual word.

- Do not penalize for use of a slash mark. Include the information before and after the slash(es) as one word.
- Penalize for a symbol used in place of a word. Do not penalize for dollar and percentage symbols used in conjunction with numerals.
- Do not penalize for use, inconsistent use, or omission of colon to introduce bulleted information and a capital letter following a bullet.
- Essay titles or ending phrases (e.g. The End) should be scored, but they must meet the scoring criteria for spelling, punctuation, capitalization, syntax, and semantics.

Specific Scoring Rules:

- ***Words must be spelled correctly.*** Penalize for misspelled words. Do not penalize for use of can not or cannot. Penalize for symbols or numerals used in place of letters. Do not penalize for numerals used in place of spelling number amounts or when referring to ordinal rank. Penalize for word boundary errors. A word boundary error occurs when two words are incorrectly combined into one word or when one word has been separated into two. Do not penalize for spelling of proper names of people/animals, places, or things.
- ***Words must be capitalized correctly.*** Proper names of people/animals and places must be capitalized, unless they can be considered a common noun based on the context of the sentence (i.e., the board game "Life" should be capitalized, whereas games such as "tag" or "softball" should not be). Penalize for words in a sentence that incorrectly begin with a distinctly capital letter. Do not penalize for capital letters in the middle or end of a word. Penalize if the first letter of a sentence incorrectly begins with a lowercase letter.

Students who write using all uppercase letters must differentiate between uppercase and lowercase letters by their relative size in order to receive credit for capitalization. Do not penalize for an entire word that is capitalized for emphasis. Beginning a new sentence with a numeral that is an abbreviation for a word (e.g., 1^{st}) is penalized for failing to use a capitalized word. Use of a numbering system prior to beginning a sentence with a capital letter is not penalized. Beginning a sentence with a numeral in place of spelling a number amount is not penalized.

- ***Words and sentences should be punctuated correctly.*** Do not penalize for the addition or omission of optional commas. A comma is required before a coordinating conjunction (for, and, nor, but, or, yet, so) connecting two independent clauses (group of words that includes verb and subject and expresses complete thought). A comma is also required when a dependent clause precedes an independent clause (e.g., When my team scores, I get excited.). Commas are required around nonessential information (i.e., The house, with the blue shutters, is mine.). When using a numbering/lettering system to make a list, a comma or period should be used (e.g., 1. I like the game because it's fun. OR One, I like the game because it's fun.). Commas should also be used after enumerative words/phrases (e.g., First,...) and after introductory words/phrases (e.g., During that time, ...). Do not penalize for use of multiple punctuation marks at the end of a sentence involving exclamation marks and/or question marks. Penalize for use of multiple ending punctuation marks involving periods or commas.
- ***Words and sentences must be syntactically correct.*** Incorrect grammar is penalized. Do not penalize for awkward or poor sentence construction that is grammatically correct. Penalize for omitted words and word endings. Penalize for repeated words. Penalize each

word that is out of order. Mark one dot in the space where a period or semicolon should have been for run-on sentences. Do not penalize the next word for incorrect capitalization. Penalize a fragmented sentence by penalizing for one or more omitted words or for use of ending punctuation after the fragment. When a sentence includes an awkward shift, penalize the word that shifted. Do not penalize for beginning a complete sentence with a conjunction.

- ***Words must be semantically correct.*** Penalize for misused words. Conjunctions and transitions must express an appropriate relationship between ideas.

VITA

Victoria Lee Buser

toribuser@gmail.com

EDUCATION

May 2020	Ph.D., School Psychology, The Pennsylvania State University
May 2017	M.Ed., School Psychology, The Pennsylvania State University
May 2015	B.S., Psychology, Juniata College

PROFESSIONAL EXPERIENCE

2019-Current	School Psychologist, *Blue Mountain School District,* Orwigsburg, PA
2018-2019	School Psychology Graduate Assistant, *Penn State University*, State College, PA
2015-2018	Outreach Analytics and Reporting Graduate Assistant, *Penn State University*, State College, PA

NEWSLETTER ARTICLES

Buser, V. L. (2017). Free speech and internet disruptions in the school environment. *National Association of School Psychologists Insight Newsletter.*

Dirsmith, J., Hutchinson, M., Runion, M., & **Buser, V. L**. (2017). Unlocking the key to school safety: Home, school, and community partnerships. *National Association of School Psychologists Insight Newsletter.*

CONFERENCE PRESENTATIONS

Buser, V. L., & Nelson, P. M. (2017, February). *Class size and student perceptions of the class environment.* Poster presented at the National Association of School Psychologists Annual Conference, San Antonio, TX.

Buser, V. L., & Gilman, A. (2015, March). *It's more than the reading level: Adapting survey questions for children's comprehension*. Poster presented at the Eastern Psychological Association Annual Conference, Philadelphia, PA.

McKellop, J. M., Westcott, K. M., & **Buser, V. L.** (2014, August). *Team or individually taught introductory psychology? Evaluating student performance satisfaction.* Poster presented at the American Psychological Association Annual Conference, Washington, D.C.

Koestler, A., **Buser, V. L**., Park, S., Peterson, M., & Gilman, A. (2014, March). *Auditory Clipping Complicates Pitch Change Detection.* Poster presented at the Eastern Psychological Association Annual Conference, Boston, MA.