Western Kentucky University
TopSCHOLAR®

Masters Theses & Specialist Projects | Graduate School

8-1-2012

Gender Differences in Written Expression at the Elementary Level

Ashley D. Melloy
Western Kentucky University, ashley.smith2@topper.wku.edu

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Child Psychology Commons, Curriculum and Instruction Commons, Curriculum and Social Inquiry Commons, Developmental Psychology Commons, Educational Psychology Commons, and the School Psychology Commons

Recommended Citation
Melloy, Ashley D., "Gender Differences in Written Expression at the Elementary Level" (2012). *Masters Theses & Specialist Projects.* Paper 1183.
http://digitalcommons.wku.edu/theses/1183

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact connie.foster@wku.edu.

GENDER DIFFERENCES IN WRITTEN EXPRESSION AT THE ELEMENTARY LEVEL

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

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

By
Ashley Melloy

August 2012

GENDER DIFFERENCES IN WRITTEN EXPRESSION
AT THE ELEMENTARY LEVEL

Date Recommended June 19, 2012

Carl Myers

Carl Myers, Director of Specialist Project

Reagan Brown

Cassie J. Zippay

Cassie Zippay

Dean, Graduate Studies and Research

7/17/12

Date

ACKNOWLEDGEMENTS

This project would not have been possible without the constant support of my thesis chair, Dr. Carl Myers. I owe thanks to Dr. Myers for his hard work throughout this long process and for the support throughout the three long years of graduate school. I have been very fortunate to be your student. I would also like to thank my committee members, Dr. Cassie Zippay and Dr. Reagan Brown, for their hard work and for serving on my committee. I am very grateful to each of you.

I also want to thank my family, friends, and peers for their constant support. I am most thankful to my parents for always being there to support my dreams. I would not be the person I am today without them. I thank my husband for providing an escape from the reality of graduate school and making me laugh during the good and bad times. You were always there to encourage and support me. I also want to thank my friends for being there for me throughout my college career. To my school psychology peers, I thank you for the friendship we have made throughout graduate school. I will cherish all of the memories we have made.

TABLE OF CONTENTS

List of Figures…… v

List of Tables……. vi

Abstract…….. vii

Introduction……. 1

Literature Review…… 3

Method……25

Results…… 30

Discussion…….. 38

Appendix A: Story Starters…… 42

Appendix B: Human Subjects Review Board Approval…… 43

References……. 44

LIST OF FIGURES

Figure 1. Mean Scores for Total Words Written Over Time……. 36

Figure 2. Mean Scores for Words Spelled Correctly Over Time……... 36

Figure 3. Mean Scores for Correct Word Sequences Over Time…….. 37

LIST OF TABLES

Table 1. Demographics Information of Participants…….. 26

Table 2. Mean Total Words Written by Grade and Gender-Fall, Winter,

and Spring……....31

Table 3. Mean Words Spelled Correctly by Grade and Gender-Fall,

Winter, and Spring ……..... 32

Table 4. Mean Correct Word Sequences by Grade and Gender-Fall,

Winter, and Spring ……..... 33

GENDER DIFFERENCES IN WRITTEN EXPRESSION
AT THE ELEMENTARY LEVEL

Ashley Melloy August 2012 48 pages

Directed by: Dr. Carl Myers, Dr. Reagan Brown, and Dr. Cassie Zippay

Department of Psychology Western Kentucky University

The use of Curriculum-Based Measurement (CBM) in schools is increasing, as it is a useful indicator of students' basic academic skills. CBM measures are often used for identifying students at-risk, monitoring their progress during interventions, and even making special education eligibility determinations. Much of the research has focused on CBM in the area of reading. Relatively few studies have examined the area of CBM-Written Expression. A couple of studies indicated there are gender differences on CBM-Written Expression measures. This study sought to determine if gender differences exist at the elementary level and, if so, at what grade level such differences appear.

This study investigated the differences between boys and girls on three CBM-Written Expression production-dependent scores (i.e., Total Words Written, Words Spelled Correctly, and Correct Word Sequence) across first through fifth grade levels. Statistically significant differences were found beginning in the first grade. However, effect sizes suggest practical differences do not occur until the second or third grade. The results indicate that schools using CBM-Written Expression data should develop gender-specific norms.

Introduction

Writing has become an essential skill for communicating in school and at work. In order for students to be successful in college and/or career, "the quality of writing must be improved" (National Commission on Writing, 2003, p. 7). According to the National Assessment of Educational Progress (NAEP), only 33% of eighth grade students and 24% of twelfth grade students are writing at or above the proficient level (Salahu-Din, Persky, & Miller, 2008). As students progress through the grade levels, the demands for writing continue to increase in the classroom and on state-mandated assessments. Therefore, identifying and intervening with students who are not writing at the proficient level is essential for student success in the future.

School systems across the country are increasingly using Curriculum-Based Measurement (CBM) procedures to monitor students' academic growth. In addition to monitoring student progress, CBM is also used to make special education eligibility decisions (Jewell & Malecki, 2005). "Curriculum-Based Measurement (CBM) is a set of standard simple, short-duration fluency measures of reading, spelling, written expression, and mathematics computation" (Shinn & Bamonto, 1998, p. 1). Most research has examined CBM procedures with the area of reading. CBM-Written Expression is thought to be a time efficient and cost effective assessment that provides easy and reliable measures that can be given frequently (Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006). However, relatively little attention has been given to the measurement of written expression skills through CBM procedures. Thus, CBM-Written Expression is the focus of the current study.

There are many factors that contribute to the success of writing. It is important for school personnel to know what affects students' learning of writing and how to improve such skills. In addition, if there are gender differences in students' writing skills, the knowledge of how they differ will be important for school personnel in order to appropriately address the needs of boys and girls. While girls have consistently outperformed boys in writing on the NAEP at the high school level (Salahu-Din et al., 2008), it is unclear if the differences exist at the elementary level.

In the current study, CBM-Written Expression probes administered to 1,002 students in grades 1 through 5 in a local school district were obtained. Using those data, Youngman (2010) developed local norms for the school district by developing tables of raw scores that correspond to various percentile ranks for each grade level. The same data set is used in this study. Of particular interest to this study is whether significant gender differences exist between boys and girls and, if so, at what grade level such differences begin.

Literature Review

The literature reviewed focuses on aspects of Curriculum-Based Measurement (CBM) in the area of Written Expression. First, an overview of CBM is presented. Next, all research studies on the topic of CBM-Written Expression are reviewed. Because this thesis focuses on gender differences, an overview of gender differences in academic achievement is presented, and gender differences in writing are emphasized.

Curriculum-Based Measurement

Curriculum-Based Measurement is a standardized tool of short duration that assesses basic academic skills in the areas of reading, mathematics, spelling and written expression (Deno, 2003). CBM consists of standard instructions, materials (reading passage, math worksheet, writing probe, etc.), scoring guidelines, and recording forms to track scores (Hosp, Hosp, & Howell, 2007). The materials and directions are very similar to what a typical student would see in the classroom on a daily basis and usually takes less than five minutes to complete. Because CBM looks like part of the curriculum, it typically does not interrupt instruction. After CBM probes are scored, recording student scores on a spreadsheet is an effective way of tracking individual and class progress.

CBM can be used for a variety of reasons including screenings, evaluating interventions, making placement decisions, developing individualized education program's (IEP), and determining eligibility for special education (Deno, 2003). According to Deno, CBM possesses the following characteristics: (a) it is technically sufficient, (b) it has standardized administration and scoring procedures, (c) it is a valid indicator of academic skills, (d) it is time conscious (1 to 5 minutes), and (e) it is

teachable to a variety of school personnel such as teachers and counselors. Therefore, CBM is an optimal choice for efficiently assessing basic academic skills.

CBM is a time and cost efficient assessment tool, which allows school districts to develop local norms with relative ease in order to obtain data for peer comparisons (Deno, 2003). Establishing local norms provides data to identify at-risk students within the local curriculum. Local norms also make it easier to monitor an individual student's progress. According to Stewart and Silberglitt (2008), developing local norms allows educators to compare individual students to peers in the same classroom, school, and/or district. The data obtained on one individual can be used in identifying problems, making instructional groups, goal setting, determining the focus and intensity of progress monitoring, moving students to different tiers, making unique educational goals and decisions, and developing Individual Education Plans (Hessler & Konrad, 2008; Stewart & Silberglitt, 2008). Fewster and MacMillan (2002) state:

> The ease of CBM's time-efficient administration procedures allows repeated measurements across extended time periods…these features add to the utility of CBM as a tool for the identification of learning problems and can also directly link assessment with the development and evaluation of instructional plans. (p. 149)

Therefore, CBM is an efficient technique useful for a variety of purposes. It is especially useful in identifying at-risk students because all students in a class, grade, or school can be quickly assessed. It is also an especially useful data-driven method to monitor the progress of individuals who are struggling in basic academic skills and are receiving academic interventions to address concerns.

Areas Assessed by Curriculum-Based Measurement

The four academic areas that CBM assesses are reading, mathematics, spelling and written expression. Each area of CBM has its own materials, instructions, and scoring guidelines. CBM-Reading combines two different measures: oral reading fluency and maze passage reading (Hosp et al., 2007). Oral reading fluency measures how many words a student can read correctly in one minute on a given passage. Hosp et al. suggest oral reading fluency passages to be equivalent in reading difficulty for each grade level and have at least 200 words in each passage.

Maze passage reading measures the ability to read a passage silently and to correctly insert a word that is missing in the passage. Maze passages should be equivalent in reading level for each grade, have 300 words, and 42 missing words with three replacement words for the student to choose from for each missing word (Hosp et al., 2007). At the kindergarten or early reading level, CBM uses letter sound fluency and word identification fluency (Hosp et al.). Letter sound fluency measures the number of correct letter sounds a student can pronounce in one minute. Word identification fluency measures the number of words a student can correctly pronounce in one minute. Hosp et al. suggest using multiple lists of words that the student should master by the end of the school year.

CBM-Math can be assessed in three different areas: early numeracy, computation (which is the most widely used), and concepts and application (Hosp et al., 2007). For computation, a student is given a worksheet with basic math computation problems that align with the students' curriculum. The student is given two minutes to complete the worksheet and the score is based on the number of digits the student gets correct. Correct

individual digits are used for scoring, not solely the answer, because it is thought to be a measure more sensitive to growth (Shinn, 1989).

As described by Hosp et al. (2007), CBM-Spelling measures a student's ability to correctly spell a word using correct letter sequences and words spelled correctly in two minutes. Correct letter sequence is scored based on the "total number of pairs of letters that are in the correct sequence" (Hosp et al., p. 76). The method of words spelled correctly is measured by the word as a whole. The student is given up to 17 words in two minutes with the examiner presenting a new word every seven to 10 seconds (depending on grade level).

CBM-Written Expression measures an individual's writing skills (Hosp et al., 2007). Traditionally, CBM-Written Expression is assessed by giving students three minutes to write in response to a grade-appropriate story starter, which is a short phrase about a particular situation (Shinn, 1998). Standardized directions and scoring procedures are used. The written probes are traditionally scored by counting the total number of words written (TWW), the number of words spelled correctly (WSC), and the number of correct word sequences (CWS). CWS consists of "two adjacent writing units (words and punctuation) that are correct within context of what is written" (Powell-Smith & Shinn, 2004, p. 11). Scoring methods will be discussed in detail in the Methods section.

Much of the research on CBM has examined the area of reading and the research has supported its use as a valid indicator of students' reading skills and its usefulness for progress monitoring (Hosp et al., 2007; Shinn, 1998; Shinn, Knutson, Good, Tilly, & Collins, 1992). For example, much of the current emphasis in literacy education reforms

on a national level (e.g., No Child Left Behind Act of 2001; Response To Intervention special education reforms) use CBM reading probes to assess and monitor students' reading abilities (McCook, 2006). Relatively few studies, however, have examined the other academic areas assessed by CBM (Fuchs, 2004; Shinn, 1998). The intent of the next section is to review the research on CBM-Written Expression.

CBM – Written Expression Research

Although the research literature on CBM-Written Expression is relatively sparse, 14 studies were identified that evaluated aspects of the assessment method. One of the earliest studies was by Espin et al. (2000), who conducted research to find valid indicators of performance for CBM-Written Expression at the middle school level. The researchers also examined whether different types of writing samples and the length of time allowed for writing would result in more valid measures. Participants, 112 seventh and eighth grade students, completed four writing pieces (two pieces with story starters and two descriptive writing samples). The students wrote their responses on the computer. The students were given the story starter or topic sentence, had 30 seconds to think about what they wanted to write, and had three minutes to write. The students were asked to type a pound sign at the end of three minutes and then continued writing for an additional two minutes. The methods used to score the writing samples were "number of words written, words spelled correctly, words spelled incorrectly, characters written, sentences written, characters per word, words per sentence, correct word sequences (CWS), correct minus incorrect word sequences (CWS-IWS), and mean length of correct word sequences (MLCWS)" (Espin et al., p. 143). The criterion variables included

teacher ratings on the students' writing proficiency and scores obtained on a district-wide writing assessment.

Results indicate that CWS-IWS was the strongest indicator of writing compared to teacher ratings of writing (r = .65 to r = .70). Results indicate that CWS-IWS, words per sentence, and sentences written correlated the highest with the district writing assessment. There were no differences for the different durations or types of writing samples.

Fewster and MacMillan (2002) conducted research to evaluate the relationship between CBM reading and written expression scores and school awarded grades at later grade levels. The participants of the study consisted of 465 students in grades 6 and 7. The participants were tested using the CBM reading and written expression methods. Teachers scored the writing probes using Total Words Written (TWW) and Words Spelled Correctly (WSC). The students' CBM scores were compared to their year-end English and social studies grades in eighth, ninth and tenth grades. Fewster and MacMillan found "correlations between…WSC scores in Grade 6 and Grade 7 with school-awarded grades in English and social studies in Grades 8, 9, and 10 were significant, $p < .005$" (p. 154). They concluded that CBM measures are "valid indicators of academic achievement" (p. 155).

Gansle, Noell, VanDerHeyden, Naquin, and Slider (2002) discuss alternative scoring measures for CBM-Written Expression. There were anecdotal concerns that the traditional scoring methods of CBM-Written Expression (i.e., TWW, WSC, & CWS) were not "meaningful or instructionally useful" (p. 492). The researchers wanted to go beyond the traditional measures of CBM and test other forms of measurement. The

following measures were tested in this study in addition to TWW, WSC, and CWS: parts of speech, long words, total punctuation marks, correct capitalization, complete sentences, words in complete sentences, sentence fragments, simple sentences, and computer-scored variables. The criterion variables were teacher rankings of students' writing skills and test scores from the Iowa Tests of Basic Skills and Louisiana Educational Assessment Program.

The participants in the Gansle et al. (2002) study consisted of 179 third and fourth graders from one elementary school. The teachers ranked their students according to their writing skill before the students were asked to complete the writing probes. The students had to complete two writing probes within a two-day period. Based on the student rankings, the researchers found "that total words written was not the best choice for predicting skill in written language" (p. 494). Correct word sequence, correct punctuation marks, and words in complete sentences were found to correlate highest with other measures of written expression (i.e., Iowa Tests of Basic Skills and Louisiana Educational Assessment Program). Primary limitations of this study are that there were only 179 participants and all of these students were either third or fourth graders from the same elementary school. Gansle et al. noted that additional research is needed to see if the results generalize to other students and other grade levels.

Malecki and Jewell (2003) examined three different forms of scoring indices for CBM-Written Expression, which are production-dependent, production-independent, and accurate-production. Production-dependent scoring is measured by the traditional methods of Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Word Sequence (CWS). Those measures are called production-dependent because the

more that is "produced" by a student, the higher the score will be. Production-independent scoring results in a percentage of correct responses and is named as such because it does not matter how much the student writes, just how precisely. Examples in the Malecki and Jewell study included Percentage of Words Spelled Correctly (%WSC) and Percentage of Correct Writing Sequences (%CWS). Accurate-production scoring is aptly named because it considers both accuracy and the amount of writing produced. An example of this method used in this study is Correct Minus Incorrect Writing Sequences (CMIWS). The participants in the study were 946 students ranging from first grade to eighth grade. These students were tested twice in the school year, once in the fall and once in the spring.

Malecki and Jewell (2003) found that the production-dependent, production-independent, and accurate-production measures were all useful at the elementary levels (first through fifth graders). However, at the middle school level (sixth through eighth graders), the production-dependent measures become less reliable and the production-independent and accurate-productions were considered the best measures to use. Gender differences were also noted in the results. Malecki and Jewell found that girls outperformed boys at all grade levels. In addition, the researchers found "that on CWS, girls' scores are higher than boys' and this gap grows over time" (p. 383). As anticipated, the participants scored higher at the end of the school year than in the beginning of the school year. The authors noted that schools need to "be aware that girls may have an advantage on these indices, or that boys may be over-identified for difficulties in writing if normative data does not account for gender" (p. 388).

Gansle et al. (2004) conducted another study to investigate alternate curriculum-based measures of written expression. The 45 participants were third and fourth grade students, 25 female and 20 male students. Participants were chosen randomly from one school in a Southeastern state. The participants took part in a short intervention in writing, and were assessed with two CBM writing probes and a subtest from an academic achievement test called the Woodcock Johnson-Revised (WJ-R). The students participated in the short intervention between the administration of the two CBM writing probes. Six methods were used to score the CBM written expression probes: total words written, all punctuation marks, correct punctuation marks, words in complete sentences, correct word sequence, and simple sentences.

Gansle et al. (2004) found that "total words written, the current hallmark measure of CBM in writing, did not enter the regression equation that predicted WJ-R writing samples, and only correlated .23 with the subtest" (p. 298). The researchers believe that "the variables that predicted the criterion measure of writing are total punctuation marks, simple sentences, and words in complete sentences" (p. 298). As a result, Gansle et al. concluded that it is still unclear which measure is best for scoring written expression and more research is needed in this area because of inconsistent results between the two studies completed by Gansle et al. (2002) and Gansle et al. (2004).

Espin, De La Paz, Scierka, and Roelofs (2005) examined CBM-Written Expression scoring methods in middle school students. However, rather than obtaining standard three-minute writing samples, Espin et al. used 35-minute expository essays to examine the reliability and validity of CBM scoring methods, specifically CWS and Correct minus Incorrect Word Sequence (CIWS). Participants included 22 seventh and

eighth grade students. The students did not receive any help from the teacher when writing the essays. The students wrote a pretest and a posttest essay over a four week period of time. During the four weeks, the students received instruction on writing strategies. In addition to the CBM scoring methods (i.e., CWS & CIWS), Espin et al. used criterion measures called functional essay elements and quality ratings of the essay to score the essays. Correlations between the measures were examined. Results indicated that CWS and CIWS were strongly correlated with the criterion measures and "both measures were sensitive to change in student performance over time" (p. 213). Espin et al. concluded that CBM measures such as CWS and CIWS are valid indicators for assessing performance in writing at the middle school level.

Jewell and Malecki (2005) conducted another study to research the differences with three types of scoring methods used in CBM-Written Expression with 203 second, fourth, and sixth graders from three different schools. The three different scoring methods included "production-dependent indices (Total Words Written, Words Spelled Correctly, and Correct Writing Sequences), production-independent indices (Percentage of Words Spelled Correctly and Percentage of Correct Writing Sequences), and an accurate-production indicator (Correct Minus Incorrect Writing Sequences)" (p. 27). The students' writing probes were scored using the various CBM methods and those scores were compared to scores from the Tindal and Hasbrouck Analytic Scoring System (THASS), the Scholastic Aptitude Test (SAT), and language arts grades. The THASS system scores students' writing based on story-idea, organization-cohesion, and conventions-mechanics. All students also took the SAT within a month of providing the writing samples.

The researchers found that "as grade level increased, fewer of the CBM scoring indices were significantly correlated with the criterion measures of SAT subtest scores, language grades, and scores on the THASS" (p. 40). The researchers concluded that when students reach the sixth grade, production-dependent fluency measures become less appropriate. However, the production-dependent measures are appropriate at the elementary level (i.e., second and fourth grades). By sixth grade, CWS was the only measure that was significantly related to the THASS scores. Results indicated that girls outperform boys on production-dependent tasks, but boys and girls perform similarly on production-independent and accurate-production measures.

Weissenburger and Espin (2005) conducted a study to determine if CBM-Written Expression scoring methods and the duration of the writing probe needed to be changed as students get older and become better writers. The 484 participants in the study were in fourth, eighth, and tenth grades from three different school districts. Participants wrote for ten minutes each time in response to two different story starters (Form A and B). The participants marked on the paper after the 3-, 5-, and 10-minute segments. Scores were obtained using total words written, correct word sequences, and the correct minus the incorrect word sequences. Scores from the Wisconsin Knowledge and Concepts Examinations (WKCE), which is a statewide test, were used as criterion measures.

Results indicated that alternate-form reliability ranged from .55 to .84 across grade levels, each being significant at the $p < .001$ level. The correlation coefficients increased with longer writing samples and decreased with grade level. Correlation coefficients between the CBM and the WKCE Language Arts scores ranged from "moderate to large, with 18 out of the 27 grade level correlations reaching significance at

the $p < .001$ level" (p. 161). The authors concluded the technical adequacy of the CBM measures decreases with age. However, it was found that the strongest curriculum-based measures are correct word sequence and correct minus incorrect word sequence at the fourth and eighth grade levels. There was no support for the use of curriculum-based measures at the tenth grade level.

Gansle et al. (2006) conducted research "to examine the technical characteristics of curriculum-based measures (CBMs) for the assessment of writing" (p. 435). They also evaluated the Six Trait model, which is another assessment for writing abilities. The 538 participants in their study were students in first through fifth grade. The students completed two CBM writing probes, one week apart, and one story for the Six Trait model. The CBM writing probes were scored with the following measures: total words written, correct word sequences, words spelled correctly, correct punctuation marks, correct capitalization, complete sentences, and words in complete sentences. The teacher ratings for the Six Trait model were based on a number system where scores ranged from one to five, one being weaker and five being stronger. Teachers evaluated the following areas of writing: ideas, organization, voice, word choice, sentence fluency, and conventions.

When looking at the data for the CBM scoring methods, Gansle et al. (2006) found that "total words written, words spelled correctly, and correct word sequences all correlate very highly with one another" and correct word sequences resulted in "the highest validity coefficient" when compared to scores on the Stanford-9, a standardized assessment of written language (p. 445). However, regarding the results with the Six Trait model, the researchers believe it "to be technically inadequate" indicating that it is

neither a good instrument to use independently or congruently with CBM (p. 447). The authors believe that more research is needed in the area of CBM-Written Expression because there are multiple dimensions in writing, making it difficult to assess.

McMaster and Espin (2007) conducted a literature review to summarize the technical qualities of CBM-Written Expression. The researchers used electronic databases to find reports and articles on CBM-Written Expression. The search resulted in nine journal articles and 17 reports from the Institute for Research on Learning Disabilities (IRLD). Upon review of the IRLD reports, McMaster and Espin concluded that "valid measures of written expression could be obtained with brief writing samples and relatively efficient, objective scoring procedures" (p. 76). The validity coefficients were strongest between production-dependent measures and the Test of Written Language (TOWL) and the Developmental Scoring System (DSS).

After reviewing CBM scoring methods, McMaster and Espin concluded that production-independent scoring may be more technically adequate at the middle and secondary levels than production-dependent scoring (i.e., TWW, WSC, CWS), which are typically used at the elementary level.

McMaster and Campbell (2008) examined the technical adequacies of CBM-WE measures across writing task, duration, and scoring procedures. Teachers from third, fifth, and seventh grades participated in the study. Participants included 25 third-graders, 43 fifth-graders, and 55 seventh-graders in the fall data collection and 21 third-graders, 32 fifth-graders, and 41 seventh-graders in the spring data collection. The participants were asked to complete two of each of following tasks: passage copying tasks, picture prompts, narrative prompts, and expository prompts. Narrative prompts are the same as

the standard CBM-Written Expression administration method. The students were allowed one and a half minutes on the passage copying tasks. On the other prompts, students were asked to mark a slash at the end of writing three minutes and again at the end of five minutes for grades three and five. For seventh grade, the students wrote a total of seven minutes. To score the writing samples, the researchers used total words written (TWW), words spelled correctly (WSC), correct word sequences (CWS), and correct minus incorrect word sequences (CIWS). McMaster and Campbell used the following as criterion measures: the Test of Written Language, Third Edition (TOWL-3), the Minnesota Comprehensive Assessment, and the students' language arts grade-point average (GPA).

The results indicated that all measures yielded sufficient ($r \geq .70$) alternate-form reliability coefficients. Passage copying produced sufficient criterion validity with the TOWL-3 for WSC and CWS in grade 3. In addition, picture prompts yielded sufficient criterion validity with the TOWL-3 for CWS and CIWS (3-minutes) and WSC and CWS (5-minutes) in grade 3, and CIWS (7-minutes) in grade 7. CWS and CIWS with a duration of 7-minutes produced sufficient criterion validity with language arts GPA. Narrative prompts produced sufficient criterion validity across criterion measures for CWS and CIWS for all durations with the exception of CWS (3-minutes) in grade 5. Expository prompts yielded sufficient criterion validity for CWS and CIWS (5- and 7-minutes) in grades 5 and 7.

Statistically significant fall-to-spring growth was noted in narrative and expository prompts with CWS and CIWS in grades 5 and 7. No growth was noted in grade 3. Overall, McMaster and Campbell support the use of more complex measures

(i.e., CWS and CIWS) instead of using TWW and WSC because the complex measures are more reliable and valid.

McMaster, Du, and Pétursdóttir (2009) examined the technical features of curriculum-based measures for beginning writers. One-hundred first grade students were asked to complete writing samples that included word or sentence copying, story prompts, letter prompts, picture-word prompts, picture-theme prompts, and photo prompts. The probes were scored using total words written (TWW), words spelled correctly (WSC), correct word sequences (CWS), correct minus incorrect word sequences (CIWS), and correct letter sequence (CLS). The criterion measures included classroom writing performance (as assessed by teachers) and the Test of Written Language, Third Edition (TOWL-3).

Results indicated that all scoring measures yielded sufficient reliability and validity on the three and five minute sentence copying task. The five minute story and picture-word prompts produced sufficient reliability and validity for WSC, CWS, and CLS. The five minute story prompts yielded sufficient reliability and validity for CWS and CLS. Significant growth was noted from February to May for the reliable and valid measures mentioned above. McMaster et al. (2009) concluded, "sentence-copying, story, picture-word, and photo prompts emerged as promising beginning-writing measures" (p. 41).

McMaster et al. (2011) conducted another study to examine the technical features of curriculum-based measures of beginning writing. Participants included 85 first grade students. The participants completed the following CBM tasks: sentence copying, picture-word prompts, and story prompts. McMaster et al. used TWW, WSC, and CWS

as scoring measures and the TOWL-3 was the criterion measure. Results indicate that TWW on sentence copying and TWW, WSC, and CWS on picture-word prompts resulted in reliability coefficients greater than $r = .70$, which the authors considered sufficient. These findings are consistent with previous research conducted by McMaster et al. (2009).

Amato and Watkins (2011) examined the predictive validity of 10 CBM-Written Expression scoring methods in a sample of 447 eighth grade students. CBM measures that were used included TWW, WSC, %WSC, CWS, %CWS, CWS-IWS, number of sentences (SEN), number of correct capitalization (CC), number of punctuation marks (TPM), and number of correct punctuation marks (TPM). Amato and Watkins used the TOWL-3 as the criterion variable because the reliability and validity of the instrument was considered sufficient by the authors. Results indicated that %CWS had the strongest correlation ($r = .61$) with the TOWL-3 scores and "contributed the most unique variance in the prediction of TOWL-3 scores ($\beta = .49$)" (p. 201). However, the authors indicated that %CWS should be used with caution when monitoring student progress because the raw scores are transformed into a percentage. A percentage can be a misleading indicator of growth. In addition, results indicated that TPM and %CWS were the greatest predictors of writing proficiency for eighth graders.

Summary of Research on CBM – Written Expression

Most of the current research focuses on the validity of the measures of CBM-Written Expression. After a review of the literature on CBM-Written Expression, it is evident that Correct Word Sequences and Correct minus Incorrect Word Sequences are consistently considered good indicators of writing proficiency across grade levels (Espin

et al., 2000; Espin et al., 2005; Gansle et al., 2002; Gansle et al., 2006; Weissenburger & Espin, 2005). Other measures that are deemed adequate often depend on the grade level of the student. The production-dependent, production-independent, and accurate-production measures were all useful at the elementary level while at the middle school level, the production-independent and accurate-productions are the best measures to use (Jewell & Malecki, 2005; Malecki & Jewell, 2003; McMaster & Espin, 2007). Some authors suggested that TWW is not a good indicator of writing proficiency, at least at the third and fourth grade levels (Gansle et al., 2002; Gansle et al., 2004) and above sixth grade (Malecki & Jewell, 2003). McMaster et al. (2011) did find TWW to be a good measure at the first grade level.

A few studies reported gender differences were found. Results indicated that girls outperformed boys at all grade levels (Jewell & Malecki, 2005; Malecki & Jewell, 2003). Results also indicated that girls outperform boys on production-dependent tasks, but boys and girls perform similarly on production-independent and accurate-production measures (Jewell & Malecki, 2005). One of the findings noted by Jewell and Malecki (2005) was that "girls significantly outperformed boys" and "girls at all grade levels tended to write more and produce more correctly spelled words and correct writing sequences than boys" (p. 39). As a result, Jewell and Malecki suggest that different norms for boys and girls should be considered regarding writing skills.

Limitations of the literature on CBM-WE gender differences include limited sample size and lack of inter-rater reliability. The first limitation, inadequate sample size, was noted in most of the studies. As examples, Espin et al. (2005) had 22 participants and Gansle et al. (2004) had only 45 participants. The number of grade levels in many of

the studies was also limited. Furthermore, none of the studies conducted inter-rater agreement checks for scoring the CBM-Written Expression probes (McMaster & Espin, 2007).

Overview of Gender Differences in Academic Achievement

Research noting gender differences in academic achievement has been discussed in a variety of education and psychology journals. In general, girls mature faster and master verbal and motor skills (e.g., drawing) more rapidly than boys (Eggen & Kauchak, 2013). Research has shown that girls tend to outperform boys in reading throughout the school years (Clark & Kragler, 2005; Gibb, Fergusson, & Horwood, 2008; Limbrick, Wheldall, & Madelaine, 2010; Marks, 2008). As previously mentioned, girls have consistently outperformed boys in writing in grades 8 and 12 (Salahu-Din et al., 2008).

Clark and Kragler (2005) conducted a study examining the effectiveness of a writing intervention in preschool classrooms. Three preschool classrooms consisting of 34 four- and five-year-olds participated in the study. Observations were conducted to collect data on the use of literacy activities and materials in each classroom and teacher interviews recorded the teachers' perception on early literacy learning. The students were also assessed in the fall and in the spring using The Test of Early Reading Ability II (TERA-II), a rhyming activity, and a writing sample. For the rhyming activity, the children were asked to match four pairs of pictures that sound alike. For the writing sample, the children were prompted to write anything on the paper, even if it was only their name. The researchers provided workshops to the teachers to explain how to intervene in the classroom using literacy development and literacy learning. The teachers

were to enhance literacy and writing activities in the classroom to promote early literacy development.

Results of the observations and teacher interviews indicated that as a result of the intervention, the teachers included more literacy activities in their classrooms. The TERA-II, rhyming activity, and writing sample indicated that the students made growth from the fall to spring. According to Clark and Kragler, girls outperformed boys in early literacy skills as early as preschool (four- to five-years-old) as measured by the TERA-II. Girls scored higher on both administrations of the TERA-II (i.e., fall and spring). The researchers also concluded, "girls had more overall literacy concepts than the boys throughout the study" (p. 296).

Although writing was not assessed, Gibb et al. (2008) conducted a study examining gender differences in reading and math that provides information on differences between boys and girls. Participants included individuals who participated in the Christchurch Health and Development Study, which is a longitudinal study. Sample sizes ranged from 775 to 1,110 individuals. Individuals were given assessments at different ages. The Burt Word Reading Test was given at ages 8, 10, 12, and 18; the Progressive Achievement Test of Reading Comprehension was given at ages 10 and 12; the Progressive Achievement Test of Mathematical Reasoning was given at age 11; and the Wechsler Intelligence Scale for Children – Revised (WISC-R) was given at ages 8 and 9. Classroom behavior was monitored at ages 6, 8, 10 and 12.

Results of the standardized educational assessment indicated that girls significantly outscored boys at ages 8, 10, 12, and 18 on a word reading subtest and also at age 10 and 12 on a reading comprehension subtest. There were no statistically

significant gender differences on the performance of a math reasoning subtest or on the WISC-R. Gibb et al. concluded that boys have an academic disadvantage with reading as early as age 8 that continues throughout the school years.

Marks (2008) examined student performance in reading and math in 31 countries. The data used in this study were obtained from the Organisation for Economic Co-operation and Development's (OECD) 2000 Programme for International Student Assessment (PISA) study. The PISA study examined academic achievement in reading, math and science and included 172,000 15-year-old students. Results of the study indicated that in all 31 countries there are gender gaps in reading. Specifically, girls performed higher than boys did in reading. In approximately half of the countries, boys significantly outperformed girls in math (Marks, 2008). The gender gaps for math are not as severe as the gaps are for reading.

Limbrick et al. (2010) used data from the 2008 National Assessment Program-Literacy and Numeracy (NAPLAN) to examine gender differences in academics. The NAPLAN was used to assess all students enrolled in school in Australia, which includes over 1 million students. Students were assessed in grades 3, 5, 7, and 9. The NAPLAN produces five scales: reading, writing, spelling, punctuation and grammar, and numeracy. Results indicated that girls outperformed boys in writing, spelling, grammar and punctuation at every grade level assessed. Math was the only area that boys outperformed girls throughout all levels of school (Limbrick et al., 2010).

Berninger, Nielsen, Abbott, Wijsman and Raskind (2008) conducted a study to examine gender differences in students with writing and reading disabilities. Participants included 122 children (mean age 11.5 years) who displayed characteristics of dyslexia.

The children were assessed in reading, writing, orthographic coding, phonological coding, rapid automatic naming, and executive functions. Results indicated that girls outperformed boys as measured by the Written Expression subtest on the Wechsler Individual Achievement Test, Second Edition. Girls also outperformed boys in orthographic coding, one phonological task, and the inhibition/switching executive function task. However, gender differences were not found in reading, rapid automatic naming, oral- or grapho-motor, or oral language (Berninger et al.).

Purpose of Present Research

Overall, there is evidence that gender differences are apparent in the areas of reading and math. However, there is little research on the gender differences in the area of writing. This study examines gender differences in the area of writing at the elementary level. Based upon a review of the literature, different ways of scoring CBM-Written Expression probes have been evaluated and the standard production-dependent measures (i.e., TWW, WSC, and CWS) are appropriate at the elementary level. TWW is more questionable but was included in the present research to evaluate that method further. In the two studies reporting on gender differences with CBM-Written Expression, girls outperformed boys on every grade level. This study also addresses some of the limitations found in the reviewed studies (e.g., small sample size, all elementary grade levels, inter-scorer agreement).

This study will provide much needed research on gender differences in the written expression area. It is important to know if gender differences exist in order to know if schools should use gender-specific norms. This study investigated the differences in the production-dependent scores (Total Written Words, Words Spelled Correctly, and

Correct Word Sequence) by gender across five grade levels (1st – 5th). Particularly, this study will provide answers to the following research questions:

Research Question 1: Do gender differences exist in written expression at the elementary level as measured by production-dependent curriculum-based measures (i.e., TWW, WSC, CWS)? Given that gender differences have been found with CBM-Written Expression by Jewell and Malecki (2005) and Malecki and Jewell (2003), it is hypothesized that at the elementary level, girls will outperform boys on written expression probes.

Research Question 2: If gender differences do exist, at what grade level do gender differences begin? Given that gender differences in language exist in early childhood (Clark & Kragler, 2005), it is expected that gender differences in writing will be observed in first grade.

Method

Participants

The participants in this study were obtained from a data set established by Youngman (2010). Youngman assisted a school district with the development of CBM-Written Expression norms for its elementary school students (grades 1-5) from five different elementary schools within one district in south central Kentucky. CBM-Written Expression probes were administered on three occasions (fall, winter, and spring) and a total of 1,348 students participated in at least one of those assessments. Youngman determined what raw scores corresponded to various percentile ranks (i.e., 10th, 16th, 25th, 50th, 75th, 84th, 90th) at each grade level for all three assessments.

While 1,348 was the total number of participants, at any one of the administrations, several participants missed the assessment for various reasons (e.g., absence, moved out of the district, moved into the district later in the school year). When the sample of students was restricted to only those students who participated in all three assessments over the course of the school year, a final sample of 1,002 students was obtained. Table 1 displays the demographics of gender, ethnicity, disability, and limited English proficiency (LEP) for the participants by grade level.

Instrument

Fifteen CBM-Written Expression story starters (5 grade levels x 3 administrations) were selected from AIMSweb (2008). The story starters can be found in Appendix A. AIMSweb provides dozens of CBM-Written Expression story starters for the following levels: primary, intermediate, and advanced. For this study, primary level story starters were selected for first, second, and third graders and intermediate level story

Table 1

Demographic Information of Participants

	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Total	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Gender												
Boys	103	(50.0)	100	(48.1)	94	(46.3)	135	(61.4)	79	(47.9)	511	(51.0)
Girls	103	(50.0)	108	(51.9)	109	(53.7)	85	(38.6)	86	(52.1)	491	(49.0)
Ethnicity												
Caucasian	126	(61.2)	122	(58.7)	134	(66.0)	139	(63.2)	118	(71.5)	639	(63.8)
African American	35	(17.0)	42	(20.2)	38	(18.7)	43	(19.5)	24	(14.5)	182	(18.1)
Hispanic	35	(17.0)	28	(13.5)	21	(10.3)	29	(13.2)	11	(6.7)	124	(12.4)
Asian	5	(2.4)	8	(3.8)	5	(2.5)	4	(1.8)	7	(4.2)	29	(2.9)
Other	5	(2.4)	8	(3.8)	5	(2.5)	5	(2.3)	5	(3.0)	28	(2.8)
Disability												
No	173	(84.0)	172	(82.7)	169	(83.3)	187	(85.0)	136	(82.4)	837	(83.5)
Yes	33	(16.0)	36	(17.3)	34	(16.7)	33	(15.0)	29	(17.6)	165	(16.5)
LEP												
No	161	(78.2)	167	(80.3)	167	(82.3)	183	(83.2)	137	(83.0)	815	(81.3)
Yes	45	(21.8)	41	(19.7)	36	(17.7)	37	(16.8)	28	(17.0)	187	(18.7)

starters were selected for fourth and fifth graders. Story starters were selected based on likely interest for the students. Each student received a piece of lined paper with a story starter typed at the top of the paper.

Procedure

After obtaining school board approval and Western Kentucky University (WKU) Human Subjects Review Board approval (see Appendix B), a WKU psychology professor who had previously received extensive training on CBM trained approximately 15 school counselors and school psychologists from the school district on the administration of the CBM-Written Expression probes. The school counselors and school psychologists administered all writing probes within a two-week period in the fall, winter, and spring of the 2008-2009 school year. The school personnel went into the classroom and handed out the writing probes facedown. The children were instructed to turn the paper over, write their name, and turn the paper back over. Then, the school counselor or school psychologist read the standardized directions to the participants. The examiners gave the following standardized directions to the class based on Hosp et al. (2007):

- Say: "Today I want you to write a story. I am going to read a sentence to you first and then I want you to compose a short story about what happens. You will have 1 minute to think about what you will write and 3 minutes to write your story. Remember to do your best work. If you do not know how to spell a word, you should guess. Are there any questions?" (Pause) "Put your pencils down and listen. For the next minute, think about … [insert story starter]."

- After reading the story starter, begin your stopwatch and allow 1 minute for the student(s) to think. (Monitor students so that they do not begin writing.) After 30 seconds say: "You should be thinking about… (insert story starter)." At the end of 1 minute, restart your stopwatch for 3 minutes and say: "Now begin Writing."
- Monitor students' attention to the task. Encourage the students to work if they are not writing.
- After 90 seconds say: "You should be writing about…(insert story starter)."
- At the end of 3 minutes say: "Thank you. Put your pencils down." (p. 88)

After all of the writing probes were administered, they were collected and given to researchers from Western Kentucky University to score. The scorers consisted of three graduate students in the school psychology program who received training in the scoring of the probes. The writing probes were scored using the standard production-dependent procedures of counting Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Word Sequences (CWS). To score TWW, the researcher must count the total number of words in the story. Powell-Smith and Shinn (2004) state that "any letter or group of letters separated by a space is defined as a word, even if the word is misspelled or is a nonsense word" (p. 9). Then, the scorer must count WSC, words that are "spelled correctly in the English language" even "if the word was used incorrectly in context" (Malecki & Jewell, 2003, p. 382). Finally, CWS was scored. According to Powell-Smith and Shinn (2004), to determine CWS the scorer must identify "two adjacent writing units (words and punctuation) that are correct within context of what is written" (p. 11).

After scoring the writing probes, 20% of the probes were selected to be re-scored to determine inter-rater agreement. A minimum inter-rater agreement of 80% is

recommended (Sattler, 2002). Inter-rater agreement for TWW ranged from 81% to 100%, WSC ranged from 58% to 100%, and CWS ranged from 50% to 100%. When agreement was below 80% in a particular classroom, all probes from that classroom were re-scored. No probes were re-scored for TWW; however, probes from 19 of the 66 classrooms for WSC and 36 of the 66 classrooms for CWS were re-scored. When differences in scores occurred, the probes were discussed as a group to reach an agreement and the resulting score was recorded.

Results

The mean scores for boys and girls in the fall, winter, and spring by grade level and CBM measure (i.e., TWW, WSC, CWS) were determined. To address the research questions about the possibility of gender differences, the results were evaluated using one-way analyses of variance (ANOVA) and effect sizes (i.e., Cohen's *d*). Ferguson (2009) recommends that effect sizes should be used to calculate the magnitude of difference between groups for social science data. Cohen (1992) recommends the following operational definitions of Cohen's *d*: small effect $\geq$ .20, medium effect $\geq$ .50, and strong effect $\geq$ .80. The results of the analyses with the one-way ANOVAs and effect sizes are reported in Tables 2, 3, and 4, for TWW, WSC, and CWS, respectively.

The first research question addressed whether or not gender differences on CBM-Written Expression would be present at the elementary level. It was hypothesized that girls would outperform boys on the CBM writing probes on every measure. Mean scores for Total Words Written (TWW) for boys and girls for all grades are presented in Table 2. At every grade level for the fall, winter, and spring administrations, girls had a mean score higher than boys. A series of one-way ANOVAs were used to evaluate the differences between boys and girls. For TWW, fifth grade in the spring was the only administration time that did not show statistically significant different scores between boys and girls. TWW scores for all other times of the year, at every grade level, were significantly different. Effect sizes were calculated to determine the magnitude of these differences and are also presented in Table 2. Of the 15 comparisons, eleven of the comparisons indicated small effect sizes and four comparisons indicated medium effect sizes. In the fall, effect sizes were larger at the fourth and fifth grade levels and smaller

Table 2

Mean Total Words Written by Grade and Gender-Fall, Winter, and Spring

	Boys		Girls		
Grade	*M*	(SD)	*M*	(SD)	Effect Size
			Fall		
1	7.5	(6.1)	9.5*	(6.8)	.30
2	16.8	(9.7)	19.9*	(9.6)	.32
3	29.2	(11.8)	33.5*	(13.4)	.34
4	31.7	(14.1)	39.6***	(16.1)	.52
5	40.5	(17.3)	49.3***	(16.5)	.52
			Winter		
1	16.3	(9.1)	19.8*	(11.0)	.36
2	25.2	(11.4)	30.1**	(14.1)	.38
3	38.7	(15.3)	46.6***	(16.7)	.49
4	41.5	(17.4)	48.8**	(16.2)	.44
5	46.4	(18.9)	54.9**	(16.1)	.49
			Spring		
1	18.8	(11.2)	23.2**	(12.1)	.37
2	26.7	(11.3)	34.0***	(12.0)	.62
3	37.1	(15.8)	44.6***	(16.5)	.46
4	37.8	(15.0)	48.0***	(15.5)	.67
5	42.4	(18.8)	46.9	(15.7)	.26

Note. Cohen's *d* was used for determining the effect sizes. Descriptions of effect sizes were: small effect $\geq$ .20, medium effect $\geq$ .50, and strong effect $\geq$ .80 (Cohen, 1992). *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3

Mean Words Spelled Correctly by Grade and Gender-Fall, Winter, and Spring

	Boys		Girls		
Grade	*M*	(SD)	*M*	(SD)	Effect Size
			Fall		
1	4.4	(5.2)	5.7	(4.7)	.26
2	13.0	(9.3)	15.7*	(9.0)	.30
3	26.2	(11.6)	30.5*	(12.9)	.36
4	28.7	(13.9)	37.7***	(15.5)	.61
5	37.7	(16.7)	47.3***	(16.4)	.58
			Winter		
1	12.6	(8.4)	15.5*	(10.3)	.31
2	21.8	(11.3)	27.0**	(13.6)	.41
3	35.8	(14.8)	43.4***	(16.7)	.48
4	38.2	(16.4)	46.4***	(15.8)	.51
5	42.9	(17.8)	53.2***	(15.7)	.61
			Spring		
1	15.5	(10.4)	19.7**	(11.8)	.38
2	23.7	(11.2)	30.9***	(11.8)	.62
3	34.8	(15.5)	42.2***	(16.8)	.46
4	35.5	(14.7)	46.0***	(15.2)	.71
5	39.6	(17.8)	45.4*	(15.0)	.35

Note. Cohen's *d* was used for determining the effect sizes. Descriptions of effect sizes were: small effect $\geq$ .20, medium effect $\geq$ .50, and strong effect $\geq$ .80 (Cohen, 1992). *$p < .05$. **$p < .01$. ***$p < .001$.

Table 4

Mean Correct Word Sequences by Grade and Gender-Fall, Winter, and Spring

	Boys		Girls		
Grade	*M*	(SD)	*M*	(SD)	Effect Size
			Fall		
1	2.8	(4.2)	2.7	(2.9)	.03
2	9.7	(8.8)	11.4	(8.5)	.19
3	22.1	(12.5)	26.9*	(14.5)	.35
4	25.0	(14.6)	35.2***	(15.6)	.67
5	32.6	(15.9)	45.0***	(17.5)	.75
			Winter		
1	9.1	(7.9)	11.2	(9.3)	.25
2	17.0	(11.6)	21.8**	(12.9)	.39
3	31.6	(14.5)	39.2***	(17.1)	.48
4	32.1	(15.3)	42.7***	(15.8)	.68
5	38.2	(18.3)	51.5***	(16.6)	.76
			Spring		
1	11.0	(10.5)	14.7*	(11.3)	.34
2	19.6	(12.2)	25.9***	(13.5)	.49
3	29.7	(15.7)	38.6***	(18.8)	.51
4	30.1	(15.1)	42.8***	(16.9)	.79
5	36.0	(18.7)	44.5**	(15.4)	.50

Note. Cohen's *d* was used for determining the effect sizes. Descriptions of effect sizes were: small effect $\geq$ .20, medium effect $\geq$ .50, and strong effect $\geq$ .80 (Cohen, 1992). *$p < .05$. **$p < .01$. ***$p < .001$.

at the first, second, and third grade levels. All effect sizes were considered small in the winter. Effect sizes varied inconsistently between small and medium in the spring.

The means for CBM-Written Expression data measured by Words Spelled Correctly (WSC) are presented in Table 3. Again, girls' mean scores were higher than boys' mean scores at every grade level and time of the year. One-way ANOVAs indicated there were statistically significant differences between boys and girls at all grade levels with the exception of one comparison. There was no statistically significant difference between first grade boys and girls in the fall. At the fall and winter assessments, medium effect sizes were present in grades 4 and 5 and small effect sizes in grades 1, 2, and 3. A mixture of small and medium effect sizes were present in the spring.

Table 4 presents the means for CBM-Written Expression data measured by Correct Word Sequences (CWS). Using one-way ANOVAs, statistically significant differences were demonstrated between boys and girls at all grade levels except for three comparisons. There were no statistically significant differences between boys and girls at Grade 1 in the fall and winter and no differences at Grade 2 in the fall. Both groups (i.e., boys and girls) scored very low on the CWS measure in first grade. From the spring of third grade through the spring of fifth grade, all effect sizes were at the medium level. For the fall in first and second grade, effect sizes were < .20 and considered trivial differences (Cohen, 1992). All other effect sizes were considered small.

Since gender differences exist at the elementary level, the second research question, which was when gender differences begin, was also addressed through the results in Tables 2, 3, and 4. It was hypothesized that gender differences would begin in

first grade. If statistically significant differences are used to determine when gender differences occur, a clear pattern emerges. Girls produce more words than boys from the beginning of first grade. By the middle of first grade, girls are also correctly spelling more words than boys. By the end of first grade, girls are using the words in sentences (CWS) better than boys.

If effect sizes are used to determine when gender differences occur, a clear pattern does not emerge for TWW and WSC, but a pattern does emerge for CWS. Trivial differences were found for the fall of first and second grades. Small effect sizes were found for the winter and spring of first and second grades and the fall and winter of third grade. Medium effect sizes were found starting in the spring of third grade and continued through fourth and fifth grades. Thus, the effect sizes indicate substantial gender differences begin in the spring of third grade. Figures 1, 2, and 3 provide a visual representation of gender differences and growth in skills over the grade levels.

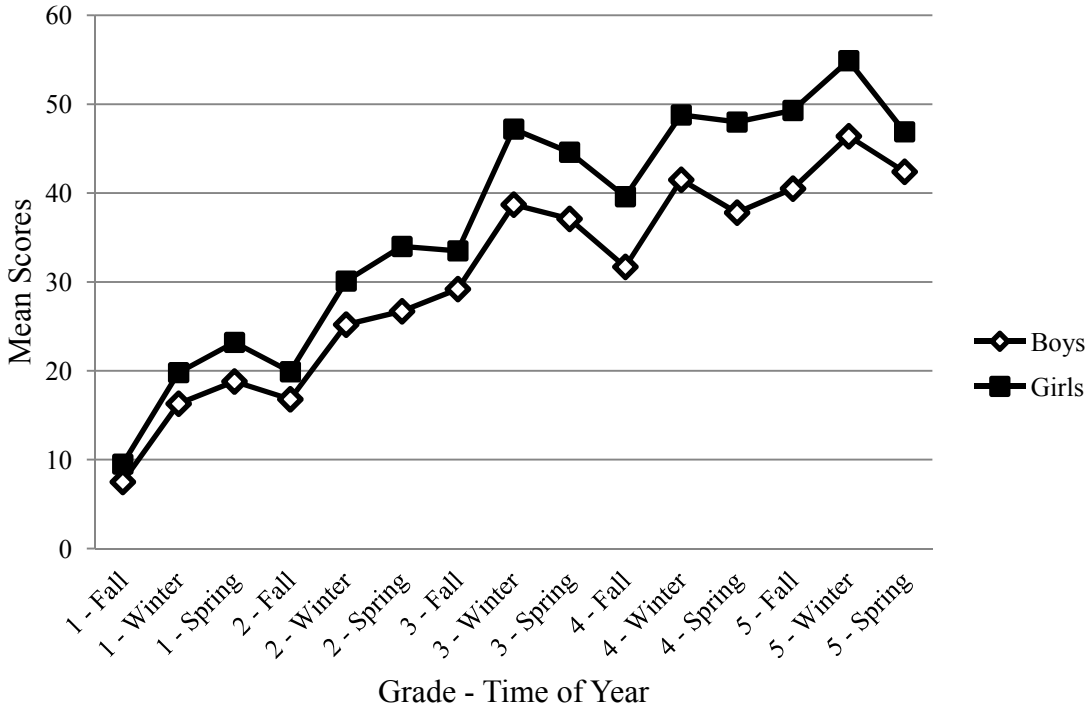

Figure 1. Mean Scores for Total Words Written Over Time

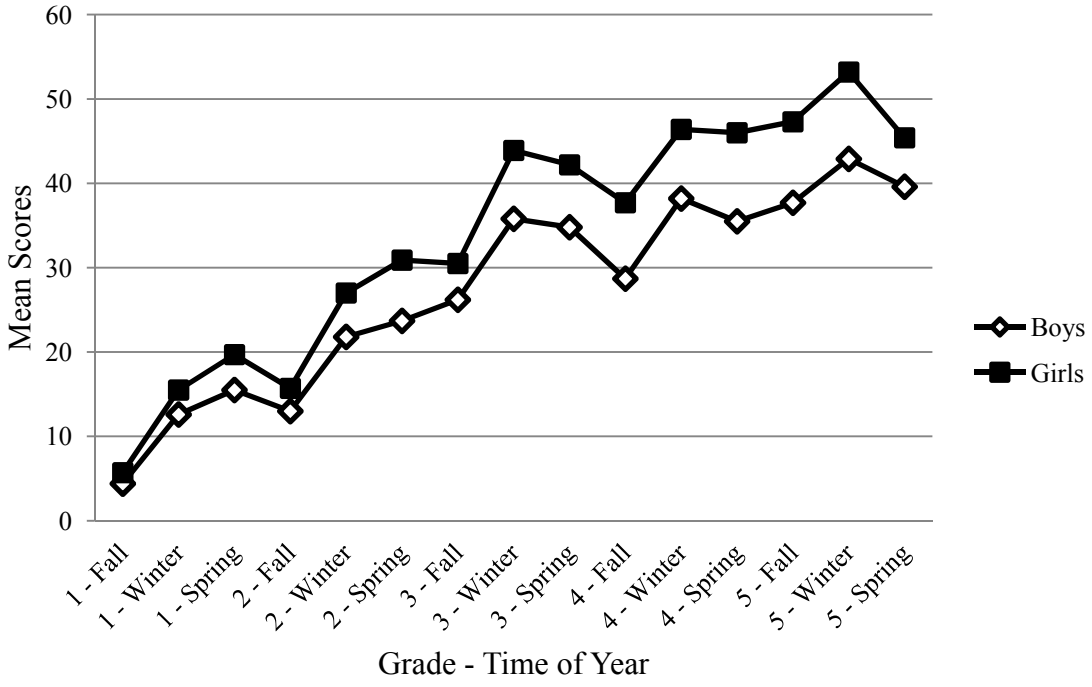

Figure 2. Mean Scores for Words Spelled Correctly Over Time

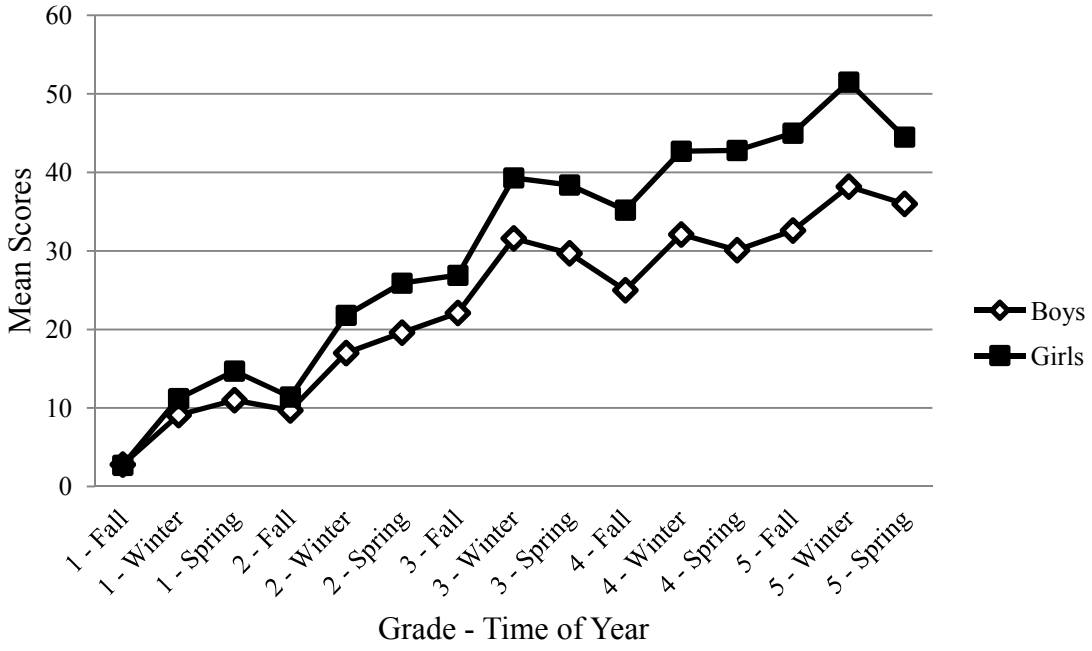

Figure 3. Mean Scores for Correct Word Sequences Over Time

Discussion

Curriculum-Based Measurement (CBM) is an assessment measure that has been used to determine which students are in need of interventions, to make special education eligibility decisions, and to monitor student progress in the areas of reading, mathematics, spelling, and written expression. The purpose of the study was to add to the literature on gender differences in the area of written expression. This study investigated the differences in the production-dependent scores (TWW, WSC, CWS) by gender at the elementary level. CBM-Written Expression probes were administered to students in grades 1-5 on three occasions (fall, winter, spring) for almost all students in one school district. The mean scores and standard deviations of each measure were calculated by gender and grade. Data were analyzed using one-way ANOVAs and Cohen's *d*.

It was hypothesized that girls would outperform boys on the production-dependent scores. In 40 of the 45 comparisons, girls' scores were statistically significantly higher than boys' scores. The other five comparisons were not statistically significant. For most of the comparisons, girls in third grade scored higher than boys in fourth grade, and girls in fourth grade scored higher than boys in fifth grade. The finding that girls outperform boys on various CBM writing scoring indices across grade levels is consistent with Malecki and Jewell (2003) and Jewell and Malecki (2005). Malecki and Jewell (2003) also noted the gender gap for CWS increases over time. In the current study, the gender gap also increases over time (with the exception of spring of fifth grade).

Gender differences with writing skills are present as early as the first grade, which is consistent with the findings of Malecki and Jewell (2003) and general observations of

young children's skill development (Eggen & Kauchak, 2013). If considering medium effect sizes as a criterion, the spring of third grade was the beginning of substantial gender differences in writing between boys and girls. The effect sizes varied between small and medium with little consistent pattern across grade levels for TWW and WSC. However, a clear pattern of effect sizes across grade levels was present for CWS. Other studies have indicated CWS to be the strongest production-dependent measure (Espin et al., 2000; Espin et al., 2005; Gansle et al., 2002; Gansle et al., 2006; Malecki & Jewell, 2003; Weissenburger & Espin, 2005). The consistency of effect size results for CWS may reflect others' findings that CWS is the best indicator of writing proficiency at the elementary level.

The current study is an important addition to the current literature on gender differences in CBM-Written Expression. It is important for school personnel to realize there are differences in writing scores between boys and girls. Malecki and Jewell (2003) warned that if the CBM-Written Expression scores for boys and girls were combined for the purpose of developing norms, then boys could "become over-identified as having difficulty in the area of writing" (p. 390). Additionally, girls in need of assistance with written expression skills might be under-identified. The results of the current study strongly support such recommendations. It is important that schools develop norms based upon gender so that school personnel will be able to assess students more accurately and develop more appropriate writing goals for intervention purposes.

Strengths and Limitations

A strength of the current study is that it provided a larger sample size (n = 1,002) at the elementary level than either of the previous two studies that examined gender

differences in CBM-Written Expression. A second strength is that this study also assessed all students from first through fifth grade. Although Malecki and Jewell (2003) assessed first through eighth graders, Jewell and Malecki (2005) only assessed second and fourth graders. Third, the current study administered CBM-Written Expression probes in the fall, winter, and spring. Malecki and Jewell (2003) administered writing probes in the fall and spring, and Jewell and Malecki (2005) only administered writing probes in the fall. Finally, the current study examined inter-scorer agreement, unlike other studies that have examined CBM-Written Expression.

A limitation of the present study is related to the scoring of CWS. Inter-scorer agreement was low for CWS scoring and required more than half of the classrooms to have all probes re-scored. The scoring guidelines produced by Hosp et al. (2007) were used to score the writing probes; however, several sequences were difficult to score. For example, CWS requires accurate capitalization. Several letters (i.e., c, m, n, o, s, u, v, w, x, z) look the same when printed lower case or capitalized. It was often difficult to tell whether those letters were correctly or incorrectly capitalized. A second limitation was the lack of integrity checks for the actual administration of the CBM writing probes. That is, although all those administering the probes practiced the administration during their training session, no one observed them to make sure they administered the probes during the school year exactly as directed. None of the studies reviewed included administration integrity checks either. A third limitation, common to most studies, is the issue of generalizability. While a diverse sample was used, the results will not necessarily generalize to other school districts.

Future Research

This study only examines gender differences at the elementary level using the production-dependent scoring measures. Additional research beyond grade five would be beneficial to examine gender differences at the middle and secondary levels. In addition to studying the upper grade levels, other curriculum-based measures could be used (i.e., production-independent and accurate-production scoring).

Scoring guidelines for CWS should be expanded upon to provide school personnel more specific rules for scoring questionable writing pieces. Of the three measures used in this study, CWS was the most difficult to score. Therefore, additional scoring guidelines related to grammar, capitalization, and punctuation would help school personnel to accurately score CBM-Written Expression probes.

APPENDIX A

Story Starters

Fall

1st grade: I once had a magic pencil and...

2nd grade: One day, I became invisible and...

3rd grade: One day our teacher disappeared and...

4th grade: I stepped into the time machine and...

5th grade: My friend and I were walking by an old deserted house and...

Winter

1st grade: If I could fly I would...

2nd grade: The dog climbed on the table and...

3rd grade: I looked out the window and to my surprise...

4th grade: The lights went out and...

5th grade: I knew I was in trouble when I couldn't find...

Spring

1st grade: As he opened the door the...

2nd grade: One day my mom surprised me and brought home a...

3rd grade: The police officer stopped the driver for speeding and then...

4th grade: He crossed his fingers and opened the box. Suddenly...

5th grade: I saw colored lights in the sky and...

APPENDIX B

Human Subjects Review Board Approval

WKU

A LEADING AMERICAN UNIVERSITY WITH INTERNATIONAL REACH

HUMAN SUBJECTS REVIEW BOARD

In future correspondence, please refer to HS10-172, February 16, 2010

Dr. Carl Myers
Psychology
WKU

Dr. Carl Myers:

Your research project, *Scoring of Written Expression Curriculum Based Measurement Probes*, was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that you need to orient participants as follows: (1) signed informed consent is not required; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

This project is therefore approved at the Exempt from Full Board Review Level.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project. Also, please use the stamped approval forms to assure participants of compliance with The Office of Human Research Protections regulations.

Sincerely,

Paul J. Mooney, M.S.T.M.
Compliance Coordinator
Office of Sponsored Programs
Western Kentucky University

HUMAN SUBJECTS REVIEW BOARD WKU APPROVED

cc: HS file number Myers HS10-172

The Spirit Makes the Master

Office of Sponsored Programs | Western Kentucky University | 1906 College Heights Blvd. #11026 | Bowling Green, KY 42101-1026
phone: 270.745.4652 | **fax:** 270.745.4211 | **e-mail:** paul.mooney@wku.edu | **web:** http://ored.wku.edu/Research_Compliance/Human_Subjects/

Equal Education and Employment Opportunities • Printing paid from state funds, KRS 57.375, 2006 • Hearing Impaired Only: 270.745.5389

References

AIMSweb. (2008). *AIMSweb progress monitoring and RTI system*. Retrieved April 1, 2008, from http://www.aimsweb.com/

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

Berninger, V., Nielsen, K., Abbott, R., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *The Journal of School Psychology, 46,* 151-172.

Clark, P., & Kragler, S. (2005). The impact of including writing materials in early childhood classrooms on the literacy development of children from low-income families. *Early Child Development and Care, 175,* 285-301.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

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

Eggen, P., & Kauchak, D. (2013). *Educational psychology: Windows on classrooms* (9th Ed.). Boston, MA: Pearson.

Espin, C., De La Paz, S., Scierka, B., & 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., Shin, J., Deno, S., 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.

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40,* 532-538.

Fewster, S., & MacMillan, P. (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.

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., Noell, G. H., VanDerHeyden, A. M., Slider, N. J., Hoffpauir, L. D., Whitmarsh, E. L., et al. (2004). An examination of the criterion validity and sensitivity to brief intervention of alternate curriculum-based measures of writing skill. *Psychology in the Schools, 41,* 291-300.

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,* 435-450.

Gibb, S., Fergusson, D., & Horwood, L. (2008). Gender differences in educational achievement to age 25. *Australian Journal of Education, 52,* 63-80.

Hessler, T., & Konrad, M. (2008). Using curriculum-based measurement to drive IEPs and instruction in written expression. *Teaching Exceptional Children, 41,* 28-37.

Hosp, M., Hosp, J., & Howell, K. (2007). *The ABC's of CBM. A practical guide to curriculum-based measurement*. New York, NY: Guilford Press.

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34,* 27-44.

Limbrick, L., Wheldall, K., & Madelaine, A. (2010). Estimating gender ratios of poor reading using large-scale assessments. *Australian Journal of Education, 54*, 190-222.

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

Marks, G. (2008). Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review of Education, 34,* 89-109.

McCook, J. (2006). *The RTI Guide: Developing and implementing a model in your schools*. Horsham, PA: LRP Publications.

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

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

McMaster, K., Du, X., Yeo, S., Deno, S., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children, 77,* 185-206.

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

National Commission on Writing. (2003). *The neglected "R": The need for a writing revolution*. Retrieved from http://www.collegeboard.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.* Eden Prairie, MN: Edformation.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The Nation's Report Card: Writing 2007* (NCES 2008–468). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications* (4th ed.). San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York, NY: Guilford Press.

Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York, NY: Guilford Press.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of Curriculum-Based Measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York, NY: Guilford Press.

Shinn, M. R., Knutson, N., Good, R. H., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

Stewart, L., & Silberglitt, B. (2008). Best practices in developing academic local norms. In A. Thomas and J. Grimes (Eds.), *Best practice in school psychology V* (pp. 225-242). Bethesda, MD: National Association of School Psychologists.

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

Youngman, E. A. (2010). *The development of curriculum-based measurement local norms in the area of written expression*. (Unpublished specialist thesis). Western Kentucky University, Bowling Green.