4-2010

The Development of Curriculum-based Measurement Local Norms in the Area of Written Expression

Elizabeth Anne Youngman
Western Kentucky University, elizabeth.youngman590@wku.edu

THE DEVELOPMENT OF CURRICULUM-BASED MEASUREMENT LOCAL NORMS IN THE AREA OF WRITTEN EXPRESSION

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
Elizabeth Anne Youngman
May 2010

THE DEVELOPMENT OF CURRICULUM-BASED MEASUREMENT LOCAL NORMS IN THE AREA OF WRITTEN EXPRESSION

Date Recommended 5-19-10

Carl Z Myers
Director of Specialist Project

Lisa C. Duffy

Pepi Barrera

Richard H. Bowles June 7, 2010
Dean, Graduate Studies and Research Date

Acknowledgements

As I near the end of this three year journey, it is bittersweet to complete this milestone of my life. Along the way, there have been numerous individuals who have taken this journey with me and I owe them much thanks. First, I'd like to thank Dr. Carl Myers, my chairperson, for all of his support, patience, and guidance. Without him, not only would this project not be possible, but neither would my career as a school psychologist. I am eternally grateful that he took a chance on me. I'd also like to thank Dr. Jacqueline Pope-Tarrence and Dr. Lisa Duffin, my thesis committee members, for their support and insight on this project.

I would like to thank my amazing family for their never ending support. Thank you to my Pop for instilling the importance of education at a young age and constantly reminding me what "#1" is in my life. Thank you to my Grandma for always being my greatest fan and teaching me to be pretty on the inside. I know that you are looking down with great pride. Thank you to my brothers, Sam and Brad, for always believing in me and for providing words of encouragement during all aspects of my life. Thank you to Brenda, for the support and empathy you shared with me every step along the way.

Thank you to my Mom. Your immense support and unconditional love has made me the person I am today. Thank you for always telling me I could do anything I wanted if I set my mind to it...and meaning it. Leading by example, you have shown me the ultimate balance between strength and compassion in helping others. Seeing how you changed so many kids' lives through your life and career inspired me to follow your path. You have truly made the world a better place for so many just by being in it.

Table of Contents

List of Tables

Page

THE DEVELOPMENT OF CURRICULUM-BASED MEASUREMENT LOCAL NORMS IN THE AREA OF WRITTEN EXPRESSION

Elizabeth A. Youngman May 2010 61 pages

Directed by: Dr. Carl Myers, Dr Jacqueline Pope-Tarrence, and Dr. Lisa C. Duffin

Department of Psychology Western Kentucky University

This project used Curriculum-Based Measurement (CBM) in the area of Written Expression to establish district norms for Bowling Green City Schools. CBM uses brief fluency measures as indicators of students' academic performance. With the use of CBM, it is possible to identify students who are considered to be at-risk for educational performance. AIMSweb probes were used to assess 1,565 first through fifth grade students from five elementary schools within the Bowling Green Independent School District. Performance was scored using the three most common scoring indices: Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Word Sequence (CWS). Data collected from this study are presented as norms at the building and district-wide levels. Each table (see Tables 2 -16) indicates student performance from each school on TWW, WSC, or CWS for a specific grade level. The tables illustrate what raw score corresponds to percentile ranks at the 10^{th}, 16^{th}, 25^{th}, 50^{th}, 75^{th}, 84^{th}, and 90^{th} percentiles. The 50^{th} percentile would be considered typical performance for an "average" student at that particular time of year and grade level. Findings from this study will be beneficial in making educational decisions regarding students potentially at risk for difficulties in the area of written expression.

Introduction

The No Child Left Behind Act of 2001 (NCLB, 2001) intended to reform education and raise academic expectations on a national scale. Two major goals that are emphasized in the NCLB Act are to improve student performance and to hold school districts accountable for success. Ideally, these two goals are attainable by setting high measurable goals, rewarding success, and determining causes for failure to meet them. According to the NCLB Act, schools are required to set individual standards and develop assessment procedures at the state level in order to measure annual student progress for grades 3 through 8.

With federal funding pending upon student performance, it is important for schools to monitor student performance and record progress. Curriculum-Based Measurement (CBM) consists of brief, fluency measures of basic academic skills in the areas of reading, math, spelling, and written expression. Although CBM has been used in a number of ways, it has two primary purposes. The first primary purpose of CBM is to screen all students to see which students might be at risk for academic difficulties. Students scoring below a predetermined level (e.g., below the 10^{th} or 20^{th} percentile) are targeted for interventions to address academic difficulties. Development of local norms is a very useful aspect of the screening process. Local norms are developed after administering the probes to all students in a district and determining percentile ranks based on the raw scores. By knowing how students from a specific school building or from a specific school district perform in an academic area, school personnel are better able to determine which students are truly at risk academically.

The second main purpose of CBM is for progress monitoring of students' academic skills. CBM probes are very brief and easy to administer so it is a simple and efficient method for determining if a student is making progress. A lack of progress implies an intervention is not working and needs to be modified or changed. If local norms have been determined, outcome goals to target are more easily developed and progress toward that goal can be measured. For example, if a third grade student is considered at risk and the average score for a third grader at the end of a school year is 40, then 40 becomes the goal for that particular student to reach by the end of the school year. The student's progress is then monitored to see if he or she is likely to reach that goal.

Unfortunately, most publications and research on CBM have been conducted in the academic area of reading. Few research studies have provided information on CBM measurement of written expression. The Bowling Green Independent School district requested assistance from Western Kentucky University with establishing local norms in the written expression area using CBM probes. This Specialist Project is part of a larger project. The purpose of this part of the project is to describe the process of establishing local norms in a school district. The primary outcome of this project will be the development of CBM written expression norms for grades one through five. Additional areas of study that may stem from the larger project could analyze more specific aspects of the data, such as differences in gender, socio-economic status, English-Language Learners, or additional measures to examine concurrent validity.

Literature Review

No Child Left Behind Act

The No Child Left Behind Act of 2001 (NCLB, 2001) was designed to reform education. It raised the bar of expectations to enhance performance and increase accountability of schools. The desired outcome of the NCLB Act is for all students to reach "proficiency" in reading and math by 2014. The NCLB Act mandates that states set individual standards and develop assessment procedures or tests to measure annual student progress and achievement in reading and math for grades 3 through 8, which are referred to as Annual Yearly Progress (AYP, The White House, 2002). Funding is allotted from the federal government to reward states for student success. If a school district is not meeting desired standards, then the NCLB Act grants parents the right to seek education for their children elsewhere.

While the act has been a large motivator for schools, it has also become controversial. Although the overall goal of the NCLB Act (2001) is focused on improving education with research-based methods, it only emphasizes reading and math, and fails to incorporate other curriculum areas (The White House, 2002). The NCLB Act established and increased federal funding for reading programs, such as Reading First, but neglects to fund improvements in other areas of education. The White House suggested that, to ensure increased AYP and secure future funding, teachers often focus on teaching items that will be on the state exam rather than on teaching other skills that should be taught. The competing argument to this criticism points out that by testing students in schools across the nation, it can be detected which schools are not meeting desired standards of education. If data can show which schools are not covering certain

basic skills, then it will allow for early interventions that can help students receive the education they need. To ensure that adequate education is being provided to all students, a system for measuring student progress is crucial (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002).

School systems across the nation are attempting to address the requirements of the NCLB Act (2001) in order to receive federal funding. To fulfill their commitment to assessment and accountability, Kentucky developed the Commonwealth Accountability Testing System (CATS) in 1998. The desired outcome of CATS was to assist Kentucky schools in achieving AYP and to serve as an additional academic indicator in grades 3 through 8. The CATS assessment included a norm-referenced test in multiple-choice format, as well as standards-based tests such as open response format, on-demand writing, and writing portfolio assignments (Kentucky Department of Education, 2009).

The writing portfolio portion of the CATS played a central and time-consuming role. Students were required to produce multiple pieces based upon writing core content in fourth, seventh, and twelfth grades. After writing several drafts, students would conference with faculty members to polish and select pieces to include in portfolios. Performance evaluation was based on Kentucky's "core content" in the area of Written Language (Kentucky Department of Education, 2009). However, critics insist that writing portfolios assessed students' ability to write drafts and make corrections rather than measuring ability based upon core content (Kentucky Department of Education).

In 2009, Kentucky's General Assembly discontinued the CATS and required Kentucky's Department of Education to establish a new method for assessment and accountability. Kentucky has been permitted by the federal government to delay

accountability until the new system has been approved and set in place (Kentucky Department of Education, 2009). Whatever the new statewide written language assessment system will be, schools will want to emphasize written language instruction starting at the very beginning of primary school, similar to current educational programs in place designed to ensure all children learn to read at proficient levels. Best practices in early literacy instruction include explicit instruction of core reading principles, daily reading of texts, school-wide screening of early literacy skills, and early intervention for children showing at-risk reading skills (National Research Council, 1998). Many schools use Curriculum-Based Measurement (CBM) procedures to screen all students' early literacy skills and determine which students may be at risk for reading difficulties (Ikeda, Neessen, & Witt, 2008). CBM procedures could be used in a similar manner in the written expression area.

Curriculum-Based Measurement

Curriculum-Based Measurement (CBM) uses brief fluency measures as indicators of students' academic performance. Customarily, the administration of CBM includes a set of standardized directions and scoring principles, a stopwatch, and the assessment materials such as reading passages, pages of math problems, and lined sheets of paper (Hosp, Hosp, & Howell, 2007; Powell-Smith & Shinn, 2004). The procedure begins with orally administered directions that are direct and precise. CBM measures are familiar to students because they resemble common academic tasks presented regularly during class (e.g., reading, writing, or solving math problems). CBM assesses students' "fluency" with the academic skills by timing them to determine a level of performance based on

correct responses within a brief time period. CBM began with measuring skills in reading, writing, and spelling, and later incorporated mathematics (Deno, 1992).

Curriculum-Based Measurement (CBM) has been used to assess student performance since the mid-1970s. Deno and Mirkin developed the assessment tool at the Minnesota Institute for Research on Learning Disabilities (Deno, 1992). CBM developed out of Deno and Mirkin's program called Data-Based Program Modification (DBPM). Individually administered, norm-referenced tests are very lengthy, require a highly trained examiner, and provide limited information about instructional practices. DBPM was an effort to develop an alternative approach in determining outcomes of education, and it was thought that the process should measure concepts directly from the curriculum being taught, as well as evaluate teacher instruction. According to Deno, it was necessary to develop a system that would provide individual student progress data for educators to determine if changes in curriculum and instruction were necessary. CBM became a preferred assessment method because it is a quick, inexpensive process that can be administered frequently with multiple forms. The CBM probes take just one to three minutes to administer, depending on the academic area.

Deno (2003) described nine attributes of CBM that set it apart from other forms of assessment. These attributes consist of the following: (a) assessment content and curriculum are identical; (b) CBM has empirical demonstrations of reliability and validity; (c) criterion-referenced measures are utilized instead of norm-referenced measures; (d) standard procedures are used so that results may be used for multiple purposes, including program evaluation; (e) student progress can be easily charted; (f) rules for judging performance are available to help interpret what the results indicate; (g)

CBM involves multiple assessments over time to provide adequate data for progress monitoring; (h) CBM is efficient because training evaluators is quick, inexpensive, and simple; and (i) data from CBM assessments can be summarized in multiple ways, including paper and pencil graphing or computer-programmed charting, which provides a straightforward visual method to analyze progress.

Initially, the primary purpose for CBM was to measure student progress. When congress passed the No Child Left Behind Act in 2001, however, an additional use of CBM quickly became popular. CBM is now frequently used in setting benchmarks at each grade level for universal screening purposes. That is, all students' scores on the brief CBM measures can be used to identify at-risk students by comparing their performance and progress against established benchmarks (Ikeda et al., 2008).

CBM provides benchmarks for typical student performance using of local norms, which present educators with an overall picture of how students are performing with respect to fellow classmates, curriculum, grade, school, or district (Stewart & Silberglitt, 2008). The development of local norms will be described in more detail in a later section of this literature review. In essence, however, typical performance (e.g., 50th percentile) is determined for the local school building or district for each grade level. That level or range of typical performance for each grade level becomes the benchmark comparison to determine whether an individual student might be at risk in a specific academic area. If a student is scoring below a predetermined level of performance (e.g., below the 10th percentile), then interventions can be developed for that student.

Once a student is identified as being at risk in an academic area, the use of CBM for progress monitoring becomes important. Ultimately, the main goal of CBM is to

improve the quality of educational decision-making (Deno, 1992; Ikeda et al., 2008). CBM's usefulness for progress monitoring is a very advantageous attribute as it is sensitive enough to document subtle gains in student progress (Gansle et al., 2002). Decisions about whether a student is making adequate progress, or if a particular academic intervention is effective, are based on data rather than based simply upon someone's opinion. Due to the brief nature of CBM probes, they can be given frequently (e.g., weekly) which allows educators to determine relatively quickly whether a student is making progress in that academic area. A lack of progress indicates a student may need some type of intervention (e.g., extra instruction, different curriculum). For a student already receiving an intervention, progress monitoring is used to determine if the curriculum being taught and instructional methods being used are adequate and effective. Progress monitoring data are extremely helpful in evaluating quality of instruction and whether or not modifications are necessary (Hosp et al., 2007).

CBM Measures of Specific Academic Areas

Reading. Oral reading fluency simply measures the number of words read correctly within a one-minute period. After providing the student with a copy of the passage to read, standardized directions are read aloud to the student. The directions instruct the student to read aloud across the page until told to stop. The examiner follows along using another copy of the reading passage and marks through any words the student misreads or does not know. At the end of one-minute, the examiner directs the student to stop reading (Hosp et al., 2007).

Traditionally, the reading probes were selected from the reading materials used for that grade level in a particular school. Currently, most school districts use grade-level

reading probes provided by one of two national agencies that promote the use of CBM assessment procedures. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002) and AIMSweb (2008) provide numerous reading probes for assessment purposes and allow computerized monitoring of students' progress.

There is a long history of research studies supporting the validity of CBM measures in the area of reading. Marston (1989) provided details on 14 early research studies supporting the validity of CBM reading measures. Additional research studies conducted since then have continued to support the reliability and validity of CBM reading measures (e.g., Elliott, Lee, & Tollefson, 2001; Shinn, Good, Knutson, & Tilly, 1992). As summarized by Hosp et al. (2007), reading CBM is considered a valid and reliable tool for identifying students at risk for reading failure, students who are not making sufficient progress with current instruction, instructional levels for individual students, and students in need of an additional evaluation.

Mathematics. When the math CBM is administered, the student (or entire classroom of students) receives sheets of paper filled with math problems (i.e., addition, subtraction, multiplication, and/or division). The examiner reads standardized directions aloud informing the student(s) to work across the page and answer as many problems as possible until told to stop. After two minutes, the examiner directs the student(s) to stop and the number of correct digits is scored (Hosp et al., 2007).

Foegen, Jiban, and Deno (2007) reviewed the available literature regarding progress monitoring using math CBM probes. After conducting a search for all related studies, 32 were determined to be applicable to progress monitoring. After an analysis of the types of measures used, skills assessed, and results, it was concluded that all studies

provided sufficient reliability and validity coefficients, suggesting math CBM to be a valid tool for progress monitoring. However, further research in math CBM was considered necessary as the majority of research studies only used elementary level students. Little attention has been devoted to mathematics at the middle school level while no studies have been conducted at the high school level (Foegen et al.).

Spelling. For the spelling CBM, the student (or entire classroom of students) is asked to number a paper 1 to 12 for first and second grade level, or 1 to 17 for third grade level and above. The teacher then reads one word at a time from a pre-selected appropriate grade-level spelling word list. After reading the first word, the teacher starts the timer and reads one word every 10 seconds (for first and second grade) or 7 seconds (for all other grades) until the end of 2 minutes. After 2 minutes, the teacher directs the students to stop. When scoring spelling CBM, the examiner counts correct letter sequences (CLS) and words spelled correctly (WSC). CLS includes the total number of adjacent pairs of letters that are arranged in the correct sequence while WSC is the total number of words that are spelled correctly (Hosp et al., 2007). No studies were found regarding the reliability and validity of spelling CBM indicating that further research in this area is necessary.

Written expression. Written Expression-Curriculum Based Measurement (WE-CBM) skills are assessed on an individual basis or with entire groups or classes at the same time. Students are presented with lined sheets of paper that include a developmentally appropriate grade level story starter. Shinn (1989) notes that story starters are open-ended statements that are written to be part of the first sentence of a story and are intended to generate students' production of a story (e.g., "One night the

lights went out and..."). Story starters should not consist of questions (e.g., "Do you like pizza?") or statements that would generate a list (e.g., "What I want for my birthday is..."). How story starters are written does not appear to have changed since they were first developed as Hosp et al. (2007) still describes the same criteria for their development as originally provided by Shinn. Hosp et al. do note that different grade level story starters (i.e., primary, intermediate, advanced) are now available from the AIMSweb Internet resource, but there is no mention of how such story starters are determined to be at a particular level.

Students are read directions and are instructed to write for three minutes. Standardized directions include reading a sentence (story starter) to the students and directing them to write about what happens next. The teacher allows the students one minute to think about what they intend to write. After one minute, the students are directed to write for three minutes. Performance is then scored using specific guidelines (Hosp et al., 2007). Gansle et al. (2002) noted that the three most common ways to score students' performance on WE-CBM are Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Writing Sequence (CWS).

To measure TWW, each word or grouping of letters is counted, even if not spelled correctly. Scoring WSC consists of counting the number of words that are accurately spelled, regardless of the context in which they are used. According to Videen, Deno, and Marston (1982), a CWS is two adjacent, correctly spelled words that are acceptable within the context of the phrase to a native speaker of the English language. CWS also depends upon punctuation, syntax, semantics, spelling, and capitalization (Hosp et al., 2007). These guidelines require that for two consecutive words to be considered as a

CWS, each word must spelled correctly, capitalized only if appropriate, and punctuated correctly, although the use or lack of use of commas is ignored. Also, a CWS must be syntactically accurate. For example, sentences beginning with a conjunction, such as "and," are not considered to be correct, even when capitalized (Hosp et al.).

Scoring of math and reading CBM probes is considerably less difficult than the scoring of written expression probes. There is no dispute regarding a student's ability to solve a math problem or read a word correctly because each problem or word has only one correct response. WE-CBM presents many challenges due to numerous possibilities of responses as well as difficulty applying scoring guidelines (Gansle et al., 2002). For example, a CWS is not counted if a word is inappropriately capitalized. However, it sometimes can be very difficult to determine if some letters (e.g., c, s, o, w, v) are written as capital letters or just printed large. A lack of spacing between words can make it difficult to discern individual words. Even poor handwriting can contribute to scoring interpretation difficulties.

Studies Examining Written Expression Curriculum Based Measurement

The evaluation of students' written expression skills can address many components. Howell, Fox, and Morehead (1993) consider fluency, syntactic maturity, vocabulary or semantic maturity, content, and language conventions to be most relevant when assessing students' written expression abilities. While assessing such factors may be ideal, the intent of WE-CBM is to provide a brief indicator of written expression abilities, not a comprehensive assessment. As previously mentioned, the conventional method for assessing WE-CBM counts the total words written, words spelled correctly, and correct word sequences (Shinn, 1989; Shinn, 1998). Only four studies were found

that have examined WE-CBM, but the majority have examined additional scoring methods of WE-CBM. Those studies will be briefly reviewed next.

A study by Fewster and Macmillan (2002) compared the CBM results in reading and written expression for 465 sixth and seventh graders with their final grades in English and Social Studies at eighth, ninth, and tenth grade. Their results indicated significant correlations between the CBM measures of Words Read Correctly (WRC) and Words Spelled Correctly (WSC) with later academic performance. The written expression measure, WSC, resulted in higher correlations with English grades (ranging from .28 for 10th grade to .34 for 8th grade) than Social Studies (ranging from .16 for 9th grade to .24 for 8th grade). Their study provided some initial support that a CBM written expression measure is related to, and even predictive of, academic success.

Gansle et al. (2002) developed a study to address teachers' face validity concerns regarding WE-CBM measures (i.e., TWW, WSC, CWS) as indicators of students' writing ability. A 3-minute WE-CBM probe was administered to 179 third and fourth graders. Results from standard WE-CBM measures were compared with two subscales related to language usage from the Iowa Tests of Basic Skills (ITBS), two subtests from the Louisiana Educational Assessment Program (LEAP), and teacher ratings. In addition to the standard methods of scoring the CBM probes, Gansle et al. also tried numerous other scoring methods to evaluate the writing samples from the probes (e.g., number of nouns, verbs, and adjectives; number of long words; total punctuation marks; correct punctuation marks; correct capitalization; complete sentences).

Results of the Gansle et al. (2002) study indicated that TWW did not significantly correlate with the ITBS, teacher rankings of writing abilities, or one subtest from the

LEAP. TWW only had a significant correlation (r = .28) with one subscale from the LEAP (i.e., Write Competently). WSC was significantly correlated with the LEAP subscales (r = .29 and .26) and teacher ratings (r = .21), but not with the ITBS. CWS was significantly correlated with all of the measures, with correlations ranging from .28 to .43. A few of the other measures resulted in statistically significant correlations, but none were considered strong. Gansle et al. were surprised to discover the measure of correct punctuation marks received the highest correlation (r = .44) between any of the measures, and that was with one subscale of the ITBS. In summary, the authors stated the TWW was not a good choice for evaluating written language skills but the CWS was a good measure. Furthermore, they recommended additional research on the possibility of using correct punctuation marks, especially because it took much less time to determine than CWS.

Jewell and Malecki (2005) conducted a study using 203 second, fourth, and sixth grade students to determine the validity of various scoring measures for WE-CBM probes, including production-dependent indices (i.e., TWW, WSC, CWS), production-independent indices, and an accurate-production index. Production-independent methods include the percentage of words in the sample that are spelled correctly (%WSC) and percentage of correct word sequences out of the total number of possible correct word sequences (%CWS). The accurate-production measure involves a formula for correct minus incorrect writing sequences (CMIWS).

Jewell and Malecki's (2005) study examined the variety of WE-CBM scoring methods as compared to the Stanford Achievement Test (SAT) Language and Spelling subtest scores, language arts grades, and Total scores from the Tindal and Hasbrouck's

(1991) analytic scoring system called the THASS. Their results indicated significant correlations, ranging from .37 to .58, between CWS and all other measures at the second and fourth grade levels but non-significant correlations at the 6th grade level (except with the Total THASS score). Very few comparisons for TWW and WSC and all the criterion measures resulted in significant correlations. An interesting result from this study indicated that at all grade levels, production-independent and accurate-production indices resulted in significant correlations (ranging from .34 to .67) with all the criterion measures of written expressions skills. The authors also noted significant gender differences for the production-dependent measures, but not for the other indices. Cautions about using production-dependent measures, particularly at the upper primary grade levels, were expressed by the authors due to the gender differences.

Gansle, VanDerHeyden, Noell, Resetar, and Williams (2006) conducted research with 538 first through fifth graders to study the technical characteristics of WE-CBM. This study specifically looked at reliability and criterion validity of these measures. Student writing samples were scored using standard WE-CBM methods (i.e., TWW, WSC, CWS) and the alternative scoring methods of correct punctuation (CP), correct capitalization (CC), complete sentences (CS), and words in complete sentences (WICS). Results were compared to the Stanford-9, which is a norm-referenced standardized test that examines prewriting, composing, editing, and total language skills. Gansle et al.'s results indicated that CWS resulted in the highest correlation with the Stanford-9 Total Language score of any of the measures (r = .43), although all of the other standard and alternative scoring methods also resulted in statistically significant correlations (ranging from .28 for CC to .41 for WICS). WSC also yielded significant results (r = .38) with

the Stanford-9 Total Language score. It is interesting to note that while previous studies did not find TWW to be significantly correlated with criterion measures such as the ITBS (Gansle et al., 2002) or the SAT (Jewell & Malecki, 2005), the current study did find a significant correlation between TWW and the Stanford-9 Total Language score ($r = .34$). The authors concluded that WE-CBM does provide reliable and valid results.

Summary of WE-CBM Studies

The limited number of available research studies indicates that the evaluations of reliability and validity of the three most common scoring indices (i.e., TWW, WSC, CWS) have varying results. Most studies have demonstrated that CWS is a useful measure of assessing written expression performance. Some support is available for the WSC measure while there seems to be little support for TWW. A wide variety of other measures have been attempted to assess production-independent skills (e.g., correct punctuation, percent of CWS) and a few have shown good potential. However, none of those results has been replicated. More research addressing the various WE-CBM measures is clearly necessary to determine the best techniques for scoring written expression probes.

Local Norms

While typical norm-referenced tests provide helpful information in comparing student performance to same-aged peers at a national level, local norms are beneficial in illustrating individual performance in comparison to more immediate factors, such as goal setting and progress monitoring (Stewart & Silberglitt, 2008). Shinn (1988) and Shapiro (2004) note that nationally norm-referenced tests are not that helpful in making many student-specific decisions because such norms do not always represent the local

student population and the content assessed on the tests is not always comparable to the local district's curriculum. For example, using local norms to compare a Bowling Green, Kentucky student's performance to other students in the same grade and district that have been taught with the same curriculum can be more meaningful than using nationally published norms for some types of decisions. Local norms aid in identifying problems, determining instructional focus and intensity, benchmarking, setting performance goals, progress monitoring, determining appropriate student placement, and allocating resources (Stewart & Silberglitt).

With local norms, individual student progress can be measured against fellow classmates. This information can be used to determine if a lack of progress stems from problems within the curriculum or are more student-specific. When comparing below average individual student performance to average classroom performance, teachers are better able to eliminate curriculum or instruction as possible reasons for low performance. Local norms help teachers determine instructional groups of students with similar abilities or difficulties. The teacher may wish to implement a specific intervention or different curriculum for a similar group of students. By monitoring student progress and determining if goals are attained, teachers can determine if an intervention being implemented results in changes in performance. In a Response to Intervention model, local norms are also capable of being used to classify students with a learning disability (Stewart & Silberglitt, 2008).

Therefore, the development of local norms is important for determining how students are performing compared to other students within the same classroom, building, or district. With a measurement tool like CBM, it is very feasible for schools to uphold

NCLB mandates in prevention, early identification, and accountability. Establishing local norms is critical in determining if student performance and progress is consistent with local educational expectations (Shinn, 1998).

Developing local norms involves the systematic collection of reliable and valid data from a local population (Stewart & Silberglitt, 2008). Local norms can be developed on classroom, school, or district-wide levels (Shinn, 1989). It is best to choose the largest group possible as district-wide norms will be much more helpful than classroom norms (Shinn, 1998). Classroom level norms are developed for individual teachers wanting to assess and track their own students. When individual schools within a district prefer to remain autonomous, have diverse populations of students, or use different curriculums, then school-independent norms are more appropriate. If a common curriculum is used across schools within a district or a comparison of performance between schools within a district is desired, then district-wide norms may be more appropriate (Shinn, 1989).

Shinn (1998) suggests establishing norms three times during the school year (i.e., fall, winter, spring). Shinn does not provide any type of empirical support for his recommendation of three times per year. However, three times per year provides a reasonable set of norms in which to compare an individual student's performance to an expected level of performance. Assessing students less than three times a year poses difficulty in comparing a student's performance to the norms because early elementary students typically make rapid academic gains. The comparison of a third grader's performance in October, for example, to third grade norms based on a May testing date would provide information of limited value. Developing norms on a very frequent basis (e.g., monthly) would require an extensive amount of resources and personnel and the

benefits would be questionable. Stewart and Silberglitt (2008) noted that some districts choose to collect norms four times per year due to state reporting or other requirements, but that even four times a year is difficult because of limited time in the typically over-loaded academic calendar. Together, this research suggests collecting data three to four times per year would be efficient and effective for evaluating students' progress.

After deciding upon what level (i.e., classroom, school, district-wide) to assess and how often to norm, the next phase of developing local norms used to be establishing a "measurement net" (Shinn, 1989). Establishing a measurement net involves collecting or developing grade appropriate materials to assess the skill of interest. This includes, but is not limited to, determining the skills to be assessed (e.g., addition or multiplication), creating or gathering the probes for assessment, preparing the standard instructions, and determining administration times during the school day. It should be noted that since the development of commercially available resources, such as DIBELS (Good & Kaminski, 2002) and AIMSweb (2008), this step is not as complicated or time-consuming as it used to be because most of the materials needed for each grade level are readily available.

The final phase of preparing to develop local norms is to determine a sampling plan (Shinn, 1989). A sampling plan involves a decision as to how many students to assess to create the local norms. All students could be assessed or a strategic sample of students may be used if resources are more limited or if the student population is huge. At a classroom level, all students in that particular classroom are typically assessed. Even at an individual school level, all students are likely to be assessed, assuming the student population is not too large for personnel to handle. For very small schools, it is

recommended two years worth of data collection be used to develop norms (Shinn, 1989). For example, if a school building only has around 50 students per grade level, data from CBM probes would be collected for two consecutive years and combined to establish norms. For district-wide norms, or schools with large populations of students, a sample of students can be used to develop norms. According to Tindal, Germann, and Deno (1983), a sufficient sample should include five to ten students for every group of forty to forty-five, which converts to approximately 11 to 25%. However, it is recommended that at least 100 students per grade level be used to establish norms so a larger percentage of students may be necessary for school districts with fewer students (Shinn, 1998).

Purpose of the Current Project

With recent changes to educational law and new responsibilities brought on by the NCLB (2001), it is becoming increasingly imperative school districts implement methods for monitoring student performance and progress. CBM has been determined an effective tool for screening students who may be considered at risk for learning problems, as well as monitoring students' progress through the curriculum or as a result of an intervention. CBM measures of reading have established growth rates, benchmarks, and norms. However, "there currently is no research on Writing CBM related to growth rates and benchmarks" (Hosp et al., 2007, p. 93). Hosp et al. reprinted WE-CBM norms from the AIMSweb website in their book. However, it is unspecified as to what demographic characteristics, region of the country, or sample size from which these norms were derived. The current AIMSweb website does not appear to post such norms, at least as public information (i.e., without paying to join). Because schools do not have ready

access to norms in order to detect students who are at risk for performance difficulties in the area of written expression, it is necessary to develop local norms.

The lack of research in this area is surprising. Some studies that have been conducted have resulted in important revelations about WE-CBM; however, few studies have been conducted to replicate or verify the results. As an example, Gansle et al. (2006) found TWW to correlate significantly with a criterion measure of written language but TWW was not significantly correlated with criterion measures for Gansle et al. (2002) or Jewell and Malecki (2005). In addition, Jewell and Malecki found gender differences with a sample of 203 second, fourth, and sixth graders when using standard WE-CBM measures (i.e., TWW, WSC, CWS). At all grade levels, girls wrote more on average than boys and generated more words spelled correctly and correct word sequences. Additional research is clearly needed in the area of WE-CBM to examine its scoring methods, growth rates, and criterion validity.

Bowling Green Independent Schools sought assistance from personnel in the Psychology Department at Western Kentucky University to help with the development of district norms at the elementary level in the area of written expression. The purpose of this Specialist Project is to illustrate how local norms are developed within a district using WE-CBM. The primary outcome of this project will be the development of CBM written expression norms and growth rates for grades one through five for each elementary building and the district as a whole. This project is considered an introduction to the process of developing norms; additional projects may follow from the data collected. Such projects could examine other factors to examine possible differences

in gender, socio-economic status, English Language Learners, or additional measures to examine concurrent validity.

Method

Participants

This project did not involve collecting data directly from the participants, but rather the scoring of probes provided by the school district. Permission to examine and score the written expression probes was obtained from Western Kentucky University's Human Subjects Review Board (see Appendix A). Typically, a representative sample of students is assessed when an entire district is used to develop norms (Shinn, 1989). However, it was decided to assess all students grades one through five because (a) such data would provide the most accurate representation of the district, (b) adequate numbers of personnel were available for this project, and (c) determining a representative sample would have been very complex and time consuming, given the diversity of individual school buildings related to ethnicity and students whose English is a second language.

Participants in this project included 1,565 first through fifth grade students from five elementary schools within the Bowling Green Independent School District. An attempt was made to obtain 100% of all students; however, the actual number of students participating during any one assessment (i.e., fall, winter, or spring) was roughly 1,300 students. The students assessed over the three time periods were not exactly the same group of students each time due to students moving in and out of the district, absences, etc. Demographic information consisting of gender, ethnicity, presence of a disability, English as a second language, and the percentage of students receiving free or reduced lunch, is presented in Table 1. As can be seen in Table 1, demographic characteristics varied greatly among the school buildings. As an example, the percentage of students where English was a second language ranged from 5.1% to 43.8%.

Table 1

Demographic Characteristics of Participants by School Building

	School Building					
	Parker-Bennett	Dishman	McNeill	Potter-Gray	T.C. Cherry	Total
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Gender						
Males	191 (54.4)	166 (52.4)	165 (50.6)	204 (50.4)	92 (47.9)	818 (52.3)
Females	160 (45.6)	151 (47.6)	161 (49.4)	175 (46.2)	100 (52.1)	747 (47.7)
Ethnicity						
Caucasian	74 (22.7)	144 (46.9)	254 (77.9)	326 (86.5)	146 (76.0)	944 (61.8)
Hispanic	106 (32.5)	53 (17.3)	7 (2.1)	9 (2.4)	7 (3.6)	182 (11.9)
African-American	111 (34.0)	92 (28.2)	44 (13.5)	26 (6.9)	33 (17.2)	306 (20.0)
Asian	14 (4.3)	6 (2.0)	12 (3.7)	11 (2.9)	2 (1.0)	45 (2.9)
Other	21 (6.4)	12 (3.9)	9 (2.8)	5 (1.3)	4 (2.1)	51 (3.3)

	School Building					
	Parker-Bennett	Dishman	McNeill	Potter-Gray	T.C. Cherry	Total
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Disability?						
No	226 (77.4)	217 (71.6)	271 (88.0)	325 (86.4)	135 (70.3)	1174 (79.8)
Yes	66 (22.6)	86 (28.4)	37 (12.0)	51 (13.6)	57 (29.7)	297 (20.2)
English Language Learner?						
No	164 (56.2)	229 (74.4)	291 (89.3)	357 (94.9)	182 (94.8)	1223 (81.9)
Yes	128 (43.8)	79 (25.6)	35 (10.7)	19 (5.1)	10 (5.2)	271 (18.1)
Free/Reduced Lunches	95%	97%	25%	25%	63%	

Instrument

Fifteen AIMSweb (2008) story starters were selected for this project (five grade levels, three times per year). Lists of story starters are provided on AIMSweb that are separated into three groups: primary, intermediate, and advanced. Primary level story starters were selected for grades 1, 2, and 3 while intermediate story starters were selected for grades 4 and 5. The selection of specific story starters was based on the author's preference and clinical judgment as to the likely interest value for students; they were not randomly selected. A list of the story starters used in this project is provided in Appendix B. Each story starter was printed on a sheet of lined paper given to the students to write their responses.

Procedure

Permission to conduct this project and assist with the scoring of the WE-CBM probes was obtained from the Bowling Green Independent Schools' school board. Approximately 15 school personnel (i.e., school psychologists, guidance counselors) responsible for administering the WE-CBM probes attended an hour-long training at the school district's central office. A Western Kentucky University psychology professor and I conducted the training session. A description of WE-CBM and an emphasis on the importance of adhering to the standardized procedures and directions were made. The personnel also engaged in a mock administration where they were presented with a sheet of lined paper, which included a story starter. Personnel were read the standardized directions and given the allotted time of one minute to think about what they were going to write. After one minute, the personnel were directed to write for three minutes. After

three minutes, the personnel were told to stop and they were given the opportunity to ask questions or express any concerns.

The WE-CBM probes were administered on three separate occasions (fall, winter, spring) throughout the 2008-2009 academic year. It was determined that three assessments would be adequate for normative purposes and that three times per year was typical based on previous research. All students in the district were assessed within the same two-week period of time. Students were assessed in groups, by classroom. Each student received the lined sheet of paper containing a grade-level appropriate story starter. All students at each grade level, at each assessment time, received the same story starter. Standardized directions were read aloud to the classroom allowing one minute for students to consider what they will write. After one minute, students were instructed to begin writing. Ninety seconds after they were told to begin writing, students were prompted to continue to write. At the end of three minutes, students were directed to stop writing and their papers were collected.

All the probes were collected and given to the involved Western Kentucky University personnel (i.e., three school psychology graduate students) for scoring. Probes were divided among the three scorers who determined Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Word Sequence (CWS) for each probe in each grade at each school. After all probes were initially scored, 20% of the probes were randomly selected from each classroom and were scored a second time by a different rater as an inter-rater agreement check. That amount, 20%, was chosen based on clinical judgment as a reasonably sized subsample. If any set of classroom probes did not have at least 80% inter-rater agreement, then all probes for that class were re-scored

and differences were discussed and resolved between two raters. A minimum of 80% agreement between observers was used as that level is considered the standard minimum level of acceptable inter-rater agreement (Alessi & Kaye, 1983; Kazdin, 2001; Sattler, 2002).

As a result of inter-rater agreement checks, no probes were re-scored for TWW. TWW had an inter-rater agreement ranging from 81% to 100% with a median of 100%. Inter-rater agreement for WSC ranged from 58% to 100% with a median of 87.5%. As a result, all probes from 19 classes were re-scored. CWS inter-rater agreement ranged from 50% to 100% with a median of 62.5%. As a result, all probes from 36 classrooms were re-scored to determine the CWS. CWS is the most difficult area to score due to a variety of reasons. During this particular project, the following factors were anecdotally noted as difficulties in accurately scoring CWS: difficulties in deciding if a letter was capitalized or not; little or no spacing between words (e.g. "cameout" or "adoor"); compound words that were separated (e.g. "flash light" or "home work"); partial erasures that made it difficult to tell what was there; capital letters placed in the middle of a word; capitalization of sentences beginning with "and" or "but"; and use, or lack of use, of quotations.

Results

Data collected from this study are presented as norms at the building and district-wide levels. Norms were determined by converting raw scores for each grade level into percentile ranks (using SPSS statistical software). Each table (see Tables 2 - 16) indicates student performance from each school on TWW, WSC, or CWS for a specific grade level. The tables illustrate what raw score corresponds to percentile ranks at the 10^{th}, 16^{th}, 25^{th}, 50^{th}, 75^{th}, 84^{th}, and 90^{th} percentiles. Such a wide variety of percentile ranks provides greater flexibility in the application of the data to individual students. The 50^{th} percentile would be considered typical performance for an "average" student at that particular time of year and grade level. Percentile ranks of 16 and 84 were included because they correspond to plus and minus one standard deviation from the mean on a normal curve. Plus/minus one standard deviation is typically considered the average range of performance (Sattler, 2002). The 10^{th} and 16^{th} percentiles, and sometimes the 25^{th}, are often used as cutoff points for determining which students are "at risk" and may need additional instruction or interventions (Hosp et al., 2007). The 75^{th} and 90^{th} percentiles are included simply to balance corresponding percentile ranks at the lower end of the normal curve and provide information on scores for high performing students. School personnel will be able to use these data to make educational decisions to address students at risk for academic concerns in the written expression area, and determine appropriate intervention goals for students.

Tables 2 through 6 illustrate student performance from each school and district-wide in the area of Total Words Written during the fall, winter, and spring assessments. Tables 7 through 11 indicate achievement from all schools in the area of Correct Word

Table 2

Grade 1 Raw Scores Corresponding to Percentile Ranks for Total Words Written

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	1	5	5	1	5	3	4	10	10	3	5	12	1	2	6	1	5	6
16th	1	7	8	1	7	5	5	13	14	4	7	15	2	5	9	3	7	8
25th	3	9	9	3	9	6	7	16	16	6	9	19	3	7	11	3	9	12
50th	5	17	13	5	16	13	10	23	28	9	14	25	6	16	19	7	16	19
75th	8	23	19	8	24	23	17	30	36	13	20	31	11	30	23	11	24	28
84th	10	26	23	11	26	27	20	33	39	16	24	34	14	35	31	14	27	31
90th	11	30	24	14	30	29	24	36	46	17	26	40	15	38	35	17	31	36

Note. F = Fall, W = Winter, S = Spring.

Table 3

Grade 2 Raw Scores Corresponding to Percentile Ranks for Total Words Written

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	4	11	12	4	9	15	10	13	16	11	16	14	5	13	12	6	12	14
16th	6	13	14	6	10	16	14	15	20	13	18	23	7	14	16	8	14	17
25th	8	17	18	6	14	22	18	20	23	16	21	27	9	19	20	11	18	22
50th	14	25	24	13	24	28	23	27	31	21	26	33	18	28	35	18	26	29
75th	19	36	32	19	30	35	29	33	38	26	35	41	24	45	45	24	35	39
84th	24	47	38	22	31	40	33	37	39	30	42	43	30	52	49	29	40	42
90th	30	51	41	27	35	42	36	43	45	33	45	46	34	58	55	33	46	46

Note. F = Fall, W = Winter, S = Spring.

Table 4

Grade 3 Raw Scores Corresponding to Percentile Ranks for Total Words Written

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	15	17	11	5	9	14	21	24	26	20	28	24	1	10	15	14	17	17
16th	17	23	15	9	14	17	24	25	29	23	33	31	8	15	15	18	23	21
25th	20	32	20	11	17	20	25	33	32	27	38	36	15	23	19	21	30	27
50th	27	42	30	20	27	34	35	41	47	36	46	47	31	45	33	30	40	38
75th	35	56	38	26	37	47	44	55	57	41	56	58	40	54	49	39	54	50
84th	40	62	43	28	43	52	47	59	63	45	60	63	47	59	51	44	59	55
90th	43	64	46	34	54	56	49	66	66	47	67	67	49	64	52	47	64	63

Note. F = Fall, W = Winter, S = Spring.

Table 5

Grade 4 Raw Scores Corresponding to Percentile Ranks for Total Words Written

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	11	19	12	12	22	18	28	27	33	29	23	37	7	20	16	14	23	18
16th	15	27	14	14	26	21	29	31	36	32	30	41	11	25	20	18	27	24
25th	18	31	19	15	28	26	33	36	38	34	35	44	14	31	31	23	32	30
50th	29	43	27	22	39	34	45	46	49	42	42	51	22	37	40	33	42	42
75th	41	60	38	29	54	43	53	58	58	49	53	59	27	54	46	46	55	54
84th	46	66	40	32	68	53	58	62	64	55	58	65	33	57	49	50	62	58
90th	48	69	44	37	77	58	65	64	68	59	65	69	40	62	55	56	65	64

Note. F = Fall, W = Winter, S = Spring.

Table 6

Grade 5 Raw Scores Corresponding to Percentile Ranks for Total Words Written

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	20	15	20	15	23	18	29	26	21	28	27	25	20	26	22	21	26	22
16th	23	32	23	18	34	28	33	33	29	34	32	32	23	29	25	26	33	26
25th	27	37	26	20	40	39	36	40	35	38	43	37	29	34	29	32	37	32
50th	45	52	32	32	63	60	50	50	44	43	54	52	43	50	41	43	53	43
75th	56	69	39	46	79	76	62	62	57	57	63	63	51	60	53	54	63	60
84th	66	73	43	52	83	82	70	66	63	61	66	69	54	63	53	60	69	64
90th	75	82	46	54	91	87	73	72	65	65	70	72	62	63	54	66	76	72

Note. F = Fall, W = Winter, S = Spring.

Table 7

Grade 1 Raw Scores Corresponding to Percentile Ranks for Correct Word Sequence

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	0	0	0	0	1	0	0	3	1	0	2	6	0	0	0	0	1	1
16th	0	1	1	0	1	0	1	6	4	0	2	8	0	0	1	0	2	1
25th	0	2	1	0	2	1	2	10	8	1	3	10	1	2	2	0	3	3
50th	1	6	4	1	6	4	3	15	17	2	8	16	1	6	10	1	7	9
75th	2	10	8	2	16	11	6	22	27	6	13	24	4	12	14	4	14	18
84th	2	12	9	3	18	16	10	25	32	7	14	26	5	19	18	5	18	23
90th	3	16	12	5	19	19	12	32	39	8	18	29	6	23	19	7	20	26

Note. F = Fall, W = Winter, S = Spring.

Table 8

Grade 2 Raw Scores Corresponding to Percentile Ranks for Correct Word Sequence

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	1	2	3	1	2	5	2	5	14	4	11	12	2	5	4	1	4	7
16th	1	3	5	1	4	7	6	9	17	5	12	14	3	7	9	2	7	9
25th	2	6	7	2	5	9	11	13	21	7	15	20	4	10	13	3	10	12
50th	4	11	12	4	10	17	16	22	26	11	20	27	8	23	24	8	17	22
75th	8	24	19	11	18	25	22	33	36	16	30	35	16	32	38	16	27	31
84th	12	27	23	14	21	29	25	36	39	21	35	39	22	41	42	20	33	36
90th	19	34	27	16	23	35	27	40	43	27	38	44	23	44	51	23	37	41

Note. F = Fall, W = Winter, S = Spring.

Table 9

Grade 3 Raw Scores Corresponding to Percentile Ranks for Correct Word Sequence

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	6	10	5	3	6	5	11	16	20	16	22	22	0	1	6	5	10	9
16th	7	15	8	3	7	8	14	18	25	20	27	26	1	8	9	8	16	14
25th	9	19	11	5	10	9	18	24	30	27	32	31	9	14	14	12	20	19
50th	16	28	19	11	20	23	28	37	40	34	44	43	19	32	27	21	34	30
75th	22	39	23	18	28	33	39	46	52	41	55	56	32	42	36	33	46	44
84th	25	48	24	21	32	39	40	54	59	44	58	59	36	45	41	39	52	52
90th	27	53	32	25	46	46	41	58	65	49	62	62	42	51	49	41	55	56

Note. F = Fall, W = Winter, S = Spring.

Table 10

Grade 4 Raw Scores Corresponding to Percentile Ranks for Correct Word Sequence

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	5	10	5	5	11	9	19	21	28	23	18	29	4	13	12	8	16	11
16th	8	16	7	7	13	11	25	25	30	26	22	35	5	18	16	12	19	15
25th	13	23	9	9	18	14	29	30	31	32	29	39	8	21	21	17	24	21
50th	20	33	19	16	26	24	37	39	42	38	38	48	15	29	29	29	35	33
75th	30	47	28	22	42	31	45	49	51	48	52	57	25	46	42	39	48	48
84th	37	53	33	27	49	36	51	54	60	51	56	61	28	52	45	45	53	53
90th	39	55	35	29	52	45	58	60	61	57	60	63	31	57	54	50	57	59

Note. F = Fall, W = Winter, S = Spring.

Table 11

Grade 5 Raw Scores Corresponding to Percentile Ranks for Correct Word Sequence

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	14	14	11	5	13	11	25	20	19	19	18	22	9	14	16	13	15	17
16th	17	24	17	8	19	14	28	26	23	23	27	24	12	16	17	19	23	19
25th	25	32	20	12	25	24	31	34	29	28	36	35	15	26	22	25	31	25
50th	32	44	27	22	48	45	41	44	39	40	49	49	37	41	35	36	45	37
75th	51	55	36	36	65	59	53	60	56	54	62	59	43	52	50	48	59	53
84th	54	64	40	40	74	64	65	62	61	60	68	65	46	54	53	54	64	58
90th	64	67	44	43	83	73	73	65	66	62	74	70	56	61	55	61	69	64

Note. F = Fall, W = Winter, S = Spring.

Table 12

Grade 1 Raw Scores Corresponding to Percentile Ranks for Words Spelled Correctly

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	0	1	2	0	3	1	2	7	6	1	3	10	0	1	3	0	3	3
16th	0	3	4	0	4	2	3	10	9	1	3	13	1	2	5	0	4	5
25th	0	5	6	0	6	4	4	14	13	2	6	16	1	6	7	1	6	8
50th	2	11	9	2	11	11	6	20	24	5	10	21	3	11	14	3	12	15
75th	5	16	14	5	19	19	12	26	31	9	17	28	9	20	19	7	19	23
84th	6	21	16	6	23	23	15	28	35	10	19	31	11	26	22	10	23	28
90th	8	23	20	7	25	24	17	33	44	11	20	35	11	33	26	11	27	31

Note. F = Fall, W = Winter, S = Spring.

Table 13

Grade 2 Raw Scores Corresponding to Percentile Ranks for Words Spelled Correctly

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	2	7	8	2	6	11	6	9	16	7	12	13	4	9	9	3	9	12
16th	3	9	12	3	9	12	9	13	18	8	16	20	5	12	14	5	12	14
25th	5	12	14	5	12	19	14	17	23	10	18	24	7	15	16	7	15	18
50th	9	21	19	7	19	24	19	25	29	15	24	31	13	27	29	14	23	27
75th	14	32	27	17	24	31	25	33	36	21	30	37	19	40	43	20	31	36
84th	19	41	35	19	26	36	28	36	38	26	39	41	25	49	47	23	37	39
90th	23	46	38	20	30	38	32	42	42	29	43	42	29	52	51	27	43	42

Note. F = Fall, W = Winter, S = Spring.

Table 14

Grade 3 Raw Scores Corresponding to Percentile Ranks for Words Spelled Correctly

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	11	11	10	5	8	10	17	21	24	17	27	23	0	6	10	10	15	14
16th	12	16	14	6	11	16	19	23	27	20	31	30	4	9	11	15	20	19
25th	17	26	19	9	14	18	22	29	31	26	36	34	13	19	15	18	27	24
50th	22	37	27	18	27	31	33	37	44	34	44	44	28	38	31	27	38	35
75th	31	51	36	24	35	45	39	52	52	40	55	56	36	49	44	37	50	48
84th	35	57	38	26	37	49	43	58	62	43	59	61	40	53	47	39	56	53
90th	37	61	40	30	46	52	45	60	66	45	63	64	45	57	50	43	60	59

Note. F = Fall, W = Winter, S = Spring.

Table 15

Grade 4 Raw Scores Corresponding to Percentile Ranks for Words Spelled Correctly

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	8	16	9	10	21	15	23	26	33	26	22	35	5	20	16	12	21	17
16th	13	22	13	12	22	17	28	29	34	31	27	37	11	23	19	15	25	21
25th	16	27	17	14	25	21	32	35	37	33	32	42	11	27	24	21	30	29
50th	26	41	24	20	35	31	42	43	48	39	41	49	20	34	36	31	40	39
75th	37	55	37	27	52	41	50	55	55	47	51	56	26	51	45	43	52	51
84th	45	59	39	30	62	47	55	59	62	53	56	62	33	55	47	47	58	55
90th	46	67	42	33	71	53	63	62	65	58	63	66	35	60	53	54	63	61

Note. F = Fall, W = Winter, S = Spring.

Table 16

Grade 5 Raw Scores Corresponding to Percentile Ranks for Words Spelled Correctly

	Parker-Bennett			Dishman			McNeill			Potter Gray			T.C. Cherry			District-Wide		
Percentiles	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F	W	S
10th	18	15	19	14	21	17	28	26	19	27	23	24	14	20	21	19	22	21
16th	21	30	21	16	30	26	32	33	29	31	30	29	20	24	22	24	30	25
25th	26	36	23	19	37	35	34	39	33	34	39	35	27	32	26	30	36	30
50th	37	48	31	28	58	57	47	48	42	42	51	51	41	44	36	41	50	42
75th	52	68	39	44	76	71	59	60	57	56	60	60	47	57	49	52	61	56
84th	62	69	42	50	79	77	65	63	62	60	65	66	54	60	53	59	68	61
90th	74	74	46	52	82	78	72	66	63	64	68	71	57	61	54	62	73	68

Note. F = Fall, W = Winter, S = Spring.

Sequence from each of the three administrations. Lastly, tables 12-16 demonstrate performance in the area of Words Spelled Correctly from each school.

In reviewing the normative data in the tables, noticeable differences between schools are apparent, particularly in the lower elementary grade levels. These differences appear to be consistent with differences in demographic variables among schools. A visual analysis of the 50th percentile generally indicates that students attending McNeill and Potter-Gray elementary schools had the highest performance. These two particular schools are regarded as having students from a higher socio-economic status as indicated by the low percentage of students receiving free or reduced lunch (25%). In addition, the percentage of students who are identified as English Language Learners is relatively low at McNeill (10.7%) and Potter-Gray (5.1%). Students at Dishman and Parker-Bennett-Curry schools produced lower written expression scores. Those schools are regarded as having students from a lower socio-economic status as indicated by the high percentage of students receiving free or reduced lunch (97 % and 95%, respectively). Dishman and Parker-Bennett-Curry also have numerous students who are English Language Learners (25.6% and 43.8%, respectively). For obvious reasons, being an English Language Learner will impair written language performance in English. Students at T.C. Cherry elementary school produced results in between the other four schools. T.C. Cherry has a low percentage of English Language Learner students (5.2%) but the percentage of students receiving free or reduced lunch was relatively high at 63%. By fifth grade, however, the large gaps between schools do not appear as great across the board. For example, fifth grade CWS scores at the 50th percentile at the winter test date only ranged from 41 to 49.

From reviewing the data, it is noted that in some tables there is a drop in raw score points at particular percentile ranks from the winter to the spring test dates, where growth would be expected. A visual analysis of the tables suggests the biggest drops occurred at Parker-Bennett-Curry for grades 3 through 5, although numerous other examples of scores dropping from winter to spring can be found across grade levels and schools. For example, the CWS drop at Parker-Bennett-Curry for 5th grade at the 50th percentile was 17, while the drop ranged from 0 to 6 for the other four schools. The reason for this is unknown and puzzling. Several of the probes from different grade levels were reviewed to make sure the data were entered correctly and that the winter and spring data were not transposed. All data were entered correctly. In addition, the statistical analyses were run again and the results were compared to data in the tables. No differences or errors were found.

It is worth mentioning that a large improvement in performance occurred from the fall to the winter test dates in first grade. On a district-wide basis, TWW went from 7 to 16, WSC went from 3 to 12, and CWS went from 1 to 7 at the 50th percentile rank. This positive result indicates that after just a few months of classroom instruction, first grade students made vast improvements. On the other hand, it is noted that the fall scores in all three areas (TWW, WSC, and CWS) for first graders are so low that little useful information is revealed from that particular administration.

Another interesting finding is that the data from this study indicate that little to no progress in written expression is made over summer months. Although the comparison is being made between different groups of students, when the normative data from the spring of one grade level are compared to the fall of the next grade level, there appears to

be no progress made and some regression. As an illustrative example, when comparing district-wide CWS scores at the 50th percentile between spring of first grade (score = 9) to fall of second grade (score = 8), it appears no progress was made and some regression occurred. However, no statistical analyses of such differences were conducted to test for significance. Such a pattern of results was fairly consistent from one grade level to the next across schools.

Discussion

The purpose of this project was to develop local norms in the area of written expression for grades 1 through 5 in Bowling Green City Schools. Data were compiled at an individual building level and at the district-wide level to provide the most accurate representation of the diversity of the district. A wide range of raw scores, corresponding to the 10^{th}, 16^{th}, 25^{th}, 50^{th}, 75^{th}, 84^{th}, and 90^{th} percentile ranks, were provided for each grade level for three points in time during the school year. Such data can be used for future cohorts of elementary students to determine which students are at risk in the written expression area (e.g., below the 10^{th} percentile) and to help with appropriate goal-setting (i.e., data are available to indicate at what level a typical student could be expected to perform at the end of the school year).

The normative data indicate that the raw scores corresponding to specific percentile ranks varied greatly by school building. Shinn (1989) recommended that norms be developed for individual school buildings within a district if the schools were diverse. These data illustrate why such a recommendation is so important. Schools with students from presumably higher socio-economic status levels (i.e., those with fewer students receiving free or reduced lunch) and with few English Language Learners tended to score much higher than other schools with large numbers of students receiving free or reduced lunch and English Language Learners. Such results are not surprising. It is interesting, however, that by 5^{th} grade some of the gaps between the school buildings minimized. Perhaps there is some type of "ceiling" effect occurring whereby a student can only write so much in a three-minute period of time. Students from diverse backgrounds may be able to catch up eventually in the writing area.

It was puzzling that sometimes the scores dropped from the winter to spring administrations. This phenomenon did not occur in the data presented by Hosp et al. (2007). The largest drops occurred at Parker-Bennett-Curry although drops were noted throughout all grade levels and school buildings. In speculating about reasons for such an unexpected result, one possibility, albeit unlikely, is that teachers did not provide much, if any, instruction on written expression skills from mid school year to the end of the school year. Another remote possibility is that students were less motivated to write at the end of the school year. One factor to consider is that the data from winter and spring do not include the same students.

Rates of inter-rater agreement for TWW, WSC, and CWS provided an unexpected, but important, implication from this project. Agreement for determining a student's TWW was very high, implying such a scoring procedure can be easily done. A consistently high level of inter-rater agreement was a little more difficult for WSC, but agreement on whether a word was spelled correctly was still at an acceptable level. Initial inter-rater agreement on CWS, however, was at an unacceptable level. Scoring procedures, such as provided by Hosp et al. (1997) and Shinn (1989), are simply inadequate to provide guidelines for all the types of writing students may produce. A much more extensive scoring guide is needed to ensure consistent scoring.

Limitations

While the administration portion of WE-CBM was relatively quick and easy, scoring the probes was a much more difficult process. Each probe was scored in three areas, TWW, WSC, and CWS, for each classroom, grade, and school. Scoring so many probes (and re-scoring many of them) made the process difficult and time consuming.

Limited research in the WE-CBM area also makes it hard to determine if the three most common scoring indices used in this project (i.e., TWW, WSC, & CWS) are the best indicators of writing proficiency.

Difficulties in accurately scoring CWS complicated the scoring process of this project. CWS is more difficult to score because it involves numerous aspects of writing, such as punctuation, syntax, semantics, spelling, and capitalization (Hosp et al., 2007). However, despite the acknowledgement that CWS scoring can be difficult, extensive scoring guidelines do not exist. Thus, it is possible that even after re-scoring so many probes, the number of CWS's reported in the normative tables may still not be exactly correct.

Strengths

The sampling plan from this project included all students in the entire district, which provided the most accurate representation of the Bowling Green City Schools' student population. Although Shinn (1989) indicated it was acceptable to assess a percentage of the population, he encouraged obtaining the largest sample size possible for development of norms. Previously published studies of WE-CBM (i.e., Fewster & Macmillan, 2002; Gansle et al. 2002; Jewell & Malecki, 2005) only examined students from two or three grade levels and used much smaller sample sizes than the current project's sample size. The one study that did use students from first through fifth grade had only 538 students (Gansle et al., 2006). The current project assessed well over 1,300 students.

Suggestions for Future Research

This project explains the overall process of developing WE-CBM norms, provides normative data for the Bowling Green City Schools, and describes how such data can be used. It may be beneficial to follow these students in a longitudinal study and compare current raw scores with future scores in the area of written expression (e.g., language arts grades or standardized test scores). In such a manner, it can be determined if there is a particular level under which students are truly at risk and need intensive early intervention.

Conducting written expression assessments in higher-grade levels will provide additional information on how students progress over their school years. Examining individual differences among populations may also be enlightening. For example, it may be beneficial to examine differences in gender or with English-Language Learners to look for ways to improve or modify classroom instructional techniques for any groups that may be having greater difficulty with written expression. Finally, additional scoring rules for CWS are definitely needed so that consistent scoring can occur for teachers and researchers using WE-CBM.

References

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

Alessi, G. J., & Kaye, J. H. (1983). *Behavior assessment for school psychologists.* Kent, OH: National Association of School Psychologists.

Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36,* 5-10.

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

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills - Modified. *School Psychology Review, 30,* 33-49.

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.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41,*121-139.

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,* 435-450.

Good, R. H., & Kaminski, R. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to Curriculum-Based Measurement.* New York: Guilford.

Howell, K. W., Fox, S. L., & Morehead, P. K. (1993). *Curriculum-based evaluation: Teaching and decision making* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 103-114). Bethesda, MD: National Association of School Psychologists.

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.

Kazdin, A. E. (2001). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth Thomson Learning.

Kentucky Department of Education. (2009, September 24). *No Child Left Behind briefing packet and reports.* Retrieved October 21, 2009, from http://www.education.ky.gov/KDE/Adminstrative+Resources

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

National Research Council (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

No Child Left Behind Act (2001). Public Law No. 107-110, §115, Stat. 1425 (2002).

Powell-Smith, K. A., & Shinn, M. R. (2004). *Administration and scoring of written expression curriculum-based measurement (WE-CBM) for use in general outcome assessment.* Eden Prairie, MN: Edformation Inc.

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

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York: Guilford Press.

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). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

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

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

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

The White House. (2002, January 8). *Fact sheet: No Child Left Behind Act.* Retrieved August 15, 2008, from http://www.whitehouse.gov/news/release/2002/01/print/20020108.html

Tindal, G., Germann, G., & Deno, S. L. (1983). *Descriptive research on the pine county norms: A compilation of findings* (Research Report No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., & Hasbrouck, J. (1991). Analyzing student writing to develop instructional strategies. *Learning Disabilities Research and Practice, 6,* 237-245.

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

Appendix A

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

cc: HS file number Myers HS10-172

HUMAN SUBJECTS REVIEW BOARD
WKU
APPROVED

Appendix B

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 …