


8-2014

# Test-Retest Reliability of Curriculum-Based Measurement Written Expression Probes

Mallory Hart

Western Kentucky University, mallory.lee@topper.wku.edu

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

 Part of the [Curriculum and Instruction Commons](#), [Educational Methods Commons](#), [Educational Psychology Commons](#), and the [School Psychology Commons](#)

---

## Recommended Citation

Hart, Mallory, "Test-Retest Reliability of Curriculum-Based Measurement Written Expression Probes" (2014). *Masters Theses & Specialist Projects*. Paper 1394.

<http://digitalcommons.wku.edu/theses/1394>

This Other 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 topscholar@wku.edu.

TEST-RETEST RELIABILITY OF CURRICULUM-BASED MEASUREMENT  
WRITTEN EXPRESSION PROBES

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  
Mallory Hart

August 2014

TEST-RETEST RELIABILITY OF CURRICULUM-BASED MEASUREMENT  
WRITTEN EXPRESSION PROBES

Date Recommended July 23, 2014

Carl Myers

Carl Myers, Director of Specialist Project

Fredrick Grieve

Fredrick Grieve

Sally Kuhlenschmidt

Sally Kuhlenschmidt

Carl A. Fro  
Dean, Graduate School

8-7-14  
Date

## ACKNOWLEDGMENTS

I have so many people to thank for helping me get to the finish line in my education and the starting line of my career. First, I am forever grateful for my thesis chair, Dr. Carl Myers, for being the person who opened the door to my future when he accepted me into this rigorous program. His approach to advising was unwavering, from his constant endeavors to maintain a program of the highest standards, to always having his door open to anyone in need of a listening ear. Dr. Rick Grieve and Dr. Sally Kuhlenschmidt, thank you for your willingness to pick up this thesis. Dr. Grieve, I appreciate that you are always engaging your students with such a great sense of humor. Also, thank you Dr. Kuhlenschmidt for being one of the most intuitive teachers I have ever encountered.

Next, I would like to thank Kris Belcher, Kristin Shiflet, and Stacia Wolf for the generosity of your time while helping me score a cumbersome number of exhilarating student writing probes. I would like to thank Ms. Jo Swanson, Assistant Superintendent to Henderson County Schools; Ms. Beth Watson, School Principal of Spottsville Elementary; and Mr. Chad Thompson, School Principal of North Middle School, for allowing me to conduct this thesis in their schools.

Finally, I owe a heart-felt thank you to my family and friends. My husband, James, has been my support system, especially when sacrificing precious vacation days from work to watch our incredible daughter. My thanks go out to Laura McGrail, Lauren Martin, and Stacey Clinard for being invaluable resources to me as a practicing student. Last but not least, this thank you is to my parents for always believing in me and providing guidance no matter what obstacles I came up against.

## TABLE OF CONTENTS

	Page
List of Tables.....	v
Abstract.....	vi
Introduction.....	1
Literature Review.....	5
Method.....	22
Results.....	27
Discussion.....	31
Appendix A: Letter to the Teachers.....	37
Appendix B: Parent Letter and Consent Form .....	38
Appendix C: Assent Form.....	40
References.....	41

## LIST OF TABLES

	Page
Table 1. Summary of Studies Examining Test-Retest Reliability of CBM-Written Expression.....	19
Table 2. Test-Retest Correlations of CBM-Written Expression Scoring Procedures by Grade Level.....	27
Table 3. Means and Standard Deviations of CBM-Written Expression Scoring Procedures by Grade Level.....	29

TEST-RETEST RELIABILITY OF CURRICULUM-BASED MEASUREMENT  
WRITTEN EXPRESSION PROBES

Mallory Hart

August 2014

44 pages

Directed by: Carl Myers, Ph.D., Frederick Grieve, Ph.D., and Sally Kuhlenschmidt, Ph.D.

Department of Psychology

Western Kentucky University

Despite the growing popularity and utilization of Curriculum-Based Measurement for assessing students' academic skills and for progress monitoring, little attention has been devoted to the area of written expression. Very few studies have been conducted to assess test-retest reliability. Only three previous studies were identified that examined the test-retest reliability of written expression curriculum-based measures. To address this issue, the current study examined the test-retest reliability of five common scoring procedures with students in grades 2, 4, and 6. A one-week time interval was used. Results indicated that while test-retest correlations were statistically significant and often at a moderate to moderately strong level, three of the measures showed statistically significant mean differences between the two test administrations in grade 6. The implications of these results are discussed.

## **Introduction**

The abilities to read and write are the primary aspects of being considered literate. Writing could also be considered a vital communication skill, with which a person advocates for one's own thoughts and desires. The skill of writing opens doors for a person's voice. For example, it allows us to document events to be referenced at a future time and even allows us to express our thoughts and feelings toward a given topic. From needing to send an important email to a co-worker, writing an essay for a college application, writing a letter to government officials to voice an opinion, to providing a cover letter and resume to a potential employer, being able to express yourself in an articulate, persuasive, and informative manner will always be an advantage. Therefore, effective writing practices are critical components necessary to communicate successfully.

In school, students must utilize their written expression capabilities almost daily in order to complete a wide range of tasks such as worksheets, writing assignments, and exams. Students must be equipped with strong written expression skills to improve chances of being successful at school and in the work place. The National Commission on Writing (2003) states:

American education will never realize its potential as an engine of opportunity and economic growth until a writing revolution puts language and communication in their proper place in the classroom. Writing is how students connect the dots in their knowledge. (p. 3)

In addition, there is a growing pressure on public schools to ensure students are ready for post-secondary education. Consistent educational standards, referred to as the



Common Core State Standards, have been proposed and adopted by many states to help improve students' college and career readiness (The National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). In part, the Common Core State Standards attempt to shift focus from student writing being based often on student opinion and experience toward writings derived from evidence-based literature and encompassing sequence and detail, capable of shaping informative and even argumentative writing.

With such an important focus being placed on improving student writing ability, there has to be quality, research-based ways to not only assess, but also monitor, the written expression ability level of students. One practice currently used in schools for assessment and progress monitoring of written expression is curriculum-based measurement (CBM). "Curriculum-based measurement is an approach for assessing the growth of students in basic skills" (Deno, 2003, p. 184). Specifically, the areas of reading, math, written expression, and spelling are assessed through CBM probes. Probes are brief and the administration is timed, typically one to seven minutes, depending on the area being assessed and age of the student. These CBM fluency measures are considered indicators of a student's academic health (Shinn & Bamonto, 1998). CBM measures have been shown to have strong criterion validity with lengthier academic assessments in dozens of studies, although only a handful have focused on written expression (Good & Jefferson, 1998).

Most of the CBM research has focused on the area of reading. Little is known even about the basic technical adequacy of CBM in the area of written expression. Few studies have examined the test-retest reliability of CBM – Written Expression. Such

reliability data helps to ensure that CBM measures of writing are stable and appropriate for providing consistent assessment information. CBM probes are often used to monitor the progress of students and are becoming a prevalent part of the Response to Intervention (RtI) framework being implemented in schools across the nation. The RtI framework has now become one of the leading models for structuring and determining student academic needs in school systems, nation-wide (Riley-Tillman, Burns, & Gibbons, 2013). The RtI model encourages universal screening and early interventions. Interventions allow students access to additional help with instruction and curriculum that they would not typically receive with previous models of addressing individual student need. The purpose of RtI is to fill the gap that occurs when we have students who are unable to grasp the content they are required to know. In-turn, these students would receive more specific strategies in addition to the instruction they have already receive in order to maintain pace and progress with peers.

A foundational component to successful RtI implementation is the progress monitoring data that is utilized for students receiving interventions. It has become common practice to use CBM results to determine the meaning of the data being collected. These data can be used for many purposes within the RtI framework, and are detailed in this project's Literature Review. Some states and school districts are now using CBM data to determine student eligibility for special education services in lieu of traditional standardized academic achievement testing, making the importance of these measures to be technically sound, all the more necessary.

Thus, test-retest reliability provides information on the acceptability of CBM to provide accurate growth information over repeated use. With few studies providing test-

retest reliability data for CBM-Written Expression assessments, it makes it uncertain as to whether schools should use such measures in an RtI model. This study will examine the test-retest reliability of five common scoring procedures for CBM-Written Expression probes with a cross-sectional sample of second, fourth, and sixth-grade students. Results will provide insight into the adequacy of these measures to produce reliable results.

## **Literature Review**

### **Curriculum-Based Measurement (CBM)**

It is important to have an understanding of how CBM complements other types of assessments. Norm-referenced, standardized testing is a type of assessment that has specific administration and scoring directions. Statistical analyses are used with large, normative samples in order to allow a comparison of a student's performance with other students of the same age or grade level. This is useful for understanding what is average performance versus above or below average performance. Criterion-referenced testing is a type of assessment that is frequently used with a specific subject matter. Scores may be used as standards for passing or failing or simply to inform teachers as to what concepts are understood or not understood. This type of testing is useful for understanding how much a student knows about the subject matter of interest.

Curriculum Based Assessment (CBA) is a type of criterion-referenced testing that teachers "...use in determining students' skills in various curricula taught in the classroom" (Idol, Nevin, & Paolucci-Whitcomb, 1999, p. ix). Such assessments are designed to reflect content taught by the teachers and include a wide variety of measures. Typically, the teacher creates the CBA measures, which is a practice that is not only time consuming but also raises questions of the results' validity and meaningfulness (Fuchs & Fuchs, 2007). CBM is considered a specific type of CBA. Unlike most CBA measures, CBM consists of specific types of assessments and has standardized administration and scoring procedures (Fuchs & Fuchs, 2007). The standardization allows CBM measures to provide student data that can be compared to other students at the same grade level using national, state or even local norms.

The content areas assessed through CBM include reading, written expression, math, and spelling. Administrations are timed and brief, typically lasting one to three minutes for reading and writing measures and five to eight minutes for mathematics. Reading fluency probes often consist of a one-page passage, using text with a reading level consistent with the students' grade-level. Students read the passage aloud until the time limit has elapsed and their score consists of how many words they were able to read correctly within that time. A CBM reading comprehension assessment requires the student to periodically choose a correct word from a list of three words while reading a passage until time has elapsed. Scores consist of how many correct words were selected.

CBM –Written Expression consists of providing a student with a sheet of lined paper with a partially completed sentence either stated verbally or written at the top of the paper. Students are required to think about the sentence then write about the topic until time has elapsed. Scoring procedures for written expression can vary depending on the purpose of the assessment and will be described in detail later in this document. For spelling, students are required to write a list of grade-level words, provided verbally, until time has elapsed. Scores consist of how many words and how many correct letter sequences were correct. In math, students solve math problems until time has elapsed. Scores consist of how many problems were solved correctly.

### **Uses of CBM**

CBM has a long history that has enabled educators and researchers to discover a variety of beneficial uses. Deno (2003) explained that CBM started in the early 1980s as a special education initiative where it was used as a formative assessment indicator of teacher effectiveness. Deno goes on to describe the development of 14 different uses of

CBM. First, teachers are able to systematically set goals, monitor growth, change programs, and evaluate the effects of changes with use of the formative evaluation model derived from CBM progress monitoring data. Second, high validity coefficients suggest that CBM can aid with (a) classifying age and grade development status, (b) teacher judgment on student proficiency, and (c) discriminating between typically-achieving students and those in compensatory programs. Third, teachers who use CBM for assessment of academic ability level are more capable of identifying appropriate student goals. Teachers who utilize CBM data to determine academic ability level are more likely to change their instruction more often and the changes they make will be more accurate. Fourth, development of local norms is possible when teachers administer probes to normative peer samples. Some CBM databases provide national norms for comparison purposes. However, local norms can be created when students in a school or district are not representative of the national population. Fifth, use of graphs to express CBM data lends itself to help ease understanding during parent conferences, multidisciplinary team meetings, and when teachers are assessing student performance.

Deno (2003) lists several uses of CBM that that are important for identification, placement, and progress monitoring of students' writing skills in both regular and special education. The sixth purpose listed is that CBM is often used as a screening method to flag students who are high-risk for academic failure. Seventh, it has been used as an evaluation tool to determine the effectiveness of pre-referral interventions in the general education setting. Pre-referral interventions consist of more intensive, specific instructional strategies in the general education classroom by the regular education teacher. When providing those additional supports, CBM can be the measure of

effectiveness to make determinations about the need to continue or modify those interventions. Eighth, CBM provides an unbiased data source. Deno explains that CBM data became a component of the problem solving model to help address the concerns with minority students being placed inappropriately in special education. Ninth, the recent use of CBM for the purpose of determining a student's response to treatment has become common practice, and has even replaced traditional standardized testing for identification of students with learning disabilities in some states. Tenth, CBM has been used to make decisions about the least restrictive environment. Teachers have the ability to monitor the progress of special education students included in regular education classes and use those data to determine whether the student benefits from inclusion or whether the student would benefit more from small group instruction.

The 11th use for CBM, described by Deno (2003), is to predict how students might perform on state wide, high-stakes testing. Research suggests a high correlation between CBM scores in math and reading with performance on high-stakes assessments. Research in this area is moving toward the ability to provide trajectories for passing a state assessment if the students achieve specific CBM scores. Twelfth, CBM progress monitoring data can be used as an indicator of growth for later academic skills and content acquisition, as CBM scores have been found to correlate with test scores, grade point averages and teacher judgments. The 13th use states that CBM has been used for assessing English Language Learners (ELL). Deno explains that a problem exists with accurately identifying students who have academic difficulties due to limited English proficiency versus students who are ELL but their academic difficulties derive from a true disability. Some researchers and school systems are using CBM reading scores to

evaluate ELL students' progress in the general education setting, based on local norms developed specifically on their ELL population. Finally, the 14th use of CBM is that oral reading scores can be used as predictive validity measures of early reading skills.

Deno (2003) goes on to note that, as the uses of CBM have evolved to serve a vast array of purposes from special education to general education settings, the special educators will most likely be the greatest beneficiary. He states that special educators not only have the time but the skill to most effectively evaluate and respond to each student's CBM performance data. The majority of these uses have been established based on extensive research in the area of reading. Of the 47 sources referenced in Deno (2003), most were directly related to research literature on reading. For the areas of written expression and mathematics, only one source was listed for each that explicitly mentioned those content areas in the title.

### **Advantages of CBM**

CBM probes provide data about a student or a student population that is sensitive to small amounts of progress or progress over small amounts of time. Traditional standardized academic achievement testing provides an overall picture of a student's current ability level compared to others but these tests are not sensitive enough to measure progress over relatively brief periods of time (e.g., weeks, months). Also, CBM probes are brief and easily administered, which enables entire populations of schools to be screened multiple times throughout the academic year. When entire populations of students are being assessed multiple times per year, that CBM data lends itself to the creation of not only national norms, but state and local norms as well. Such norms can be



used for benchmarking purposes, where students who are not progressing adequately can be identified and provided with interventions.

CBM data, when graphed, provide a clear illustration of growth over time for individual student progress monitoring and for comparison to peers' skill levels. The visual component of CBM data enables the ability of professionals to provide a clear delineation of results to all those concerned about the student's performance, no matter how familiar they are with the measures. CBM graphs are not only easy to understand for the adults invested in a child, but the graphs are also easy to understand for the student. Students can keep track of their own data and view their own performance in relation to their goal with very little effort (National Center on Student Progress Monitoring, n.d.).

CBM data are helpful in making data based decisions regarding student needs. For example, the data can help determine whether a student or group of students need more instruction with particular material, whether a student should be referred for additional testing to consider special education services, and individual goals for students. These data-based decisions can also be used to evaluate the success of the instruction students are receiving and help shape the focus of future instruction (National Center on Student Progress Monitoring, n.d.).

### **CBM-Written Expression**

Curriculum-based measures are available for the areas of reading, mathematics, written expression, and spelling. As the topic of this project is on the written expression area, CBM-Written Expression will be described in depth. In general, the goal behind CBM-Written Expression is to obtain a writing sample from the students. The writing sample is structured by giving all students being assessed the same topic to write about

using a story starter. “Story starters are short oral or written sentences that begin the writing process [and] ...are designed to elicit more than a yes/no or short-answer response” (Hosp, Hosp, & Howell, 2007, p. 85). An example provided by Hosp et al. is, “I was walking to school one day when...” (p. 86). The story starter should be appropriate and relevant to the students for whom it is being administered. The examiner should take into consideration the students' ethnicity, culture, age, and interests when choosing a story starter (Fuchs & Fuchs, 2007).

To assess CBM-Written Expression, the probes consist of either a completely blank sheet of lined paper, if the story starter is given orally, or a lined sheet of paper that has a story starter written at the top. CBM-Written Expression probes can be administered to an individual student or in a group setting (e.g., to an entire class at the same time). To administer a CBM-Written Expression probe, the lined paper with the story starter is placed in front of the students and the examiner reads the standardized directions. The students are given one minute to think about what they would like to write and an allotted amount of time for the students to write their story. Students are then prompted to begin writing.

There is variability among the length of writing time recommended. Some studies allow 5 minutes for writing (e.g., McMaster, Du, & Pétursdóttir, 2009). Fuchs and Fuchs (2007) recommend the following writing times per grade level: mid-elementary - 3 minutes, late-elementary - 5 minutes, middle school - 7 minutes, and high school - 7 minutes. The AIMSweb (2008) program for CBM use utilizes a 3-minute writing time across all grade levels being assessed.

## **Scoring Procedures**

There are several scoring options for CBM-Written Expression probes. All scoring methods fall under one of the following categories: production-dependent and production-independent (Tindal & Parker, 1989) or accurate-production indices (Jewell & Malecki, 2005). Production-dependent scoring methods are a measure of fluency and scores are dependent on how much a student writes. The more a student writes, the greater the likelihood of a higher score. Production-independent scoring methods are a measure of accuracy because the length of the writing sample does not affect the score. The percent correct, based on however much is written, is determined. Accurate-production methods are a measure of both fluency and accuracy.

The production-dependent scoring procedures have been used most frequently in the literature and include: Total words written (TWW), words spelled correctly (WSC), correct writing sequence (CWS). TWW and WSC were the initial methods of scoring CBM-Written Expression when it was created. TWW is the sum of all words written while disregarding spelling and syntax. TWW is considered an indicator of general written expression abilities up to Grade 6 and for older students who demonstrate difficulty with written expression (AIMSweb, 2008). WSC is the calculation of all words correctly spelled with the caveat that words must be used correctly within context. CWS is the sum of sequentially paired words that are mechanically, semantically, and syntactically correct. This procedure is somewhat subjective when scoring, yet provides a better metric of writing skills (AIMSweb, 2008). Further support for the contention that CWS is a good measure of writing comes from Espin et al. (2000), who found that CWS

yielded the strongest alternate-forms reliability and validity coefficients at the middle school level.

Examples of production-independent scoring methods include the percent of words spelled correct (%WSC) and percent of correct writing sequences (%CWS). Percentage of WSC is calculated by dividing WSC by TWW. Percentage of CWS is calculated by dividing CWS by the total number of written sequences. Percentage measures were found to have high correlations with teachers' holistic ratings of writing and for differentiating between special education and remedial students at the secondary level (Espin et al., 2000; Tindal & Parker, 1989; Watkinson & Lee, 1992). The results of one large study, with 2,160 students in grade 2 through 11, suggested that percentage metrics were the most appropriate for screening and eligibility (Parker, Tindal, & Hasbrouck, 1991). However, percentage measures should be used with caution as progress monitoring measures because they do not necessarily show student progress in terms of production (Espin et al., 2000).

Accurate-production scoring methods combine measures of production and accuracy. An example of an accurate-production method would be correct minus incorrect word sequence (CIWS). Correct minus incorrect word sequences is calculated by subtracting the total number of incorrect word sequences from the total CWS. Thus, a higher score would be dependent on writing more, but a student's accuracy is also assessed.

Many additional scoring techniques for CBM-Written Expression have been tested. Examples include: mean length of correct word sequences, number of legible words, percent of legible words, characters written, words spelled incorrectly, sentences

written, characters per word, and words per sentence (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006). However, these techniques have not been validated because they have only been used in one or two studies or they have had poor reliability or validity coefficients associated with their ability to measure written expression accurately (McMaster & Espin, 2007).

### **Gender Differences**

Ideally, males and females would not differ in performance, on average, in the assessment of academic abilities or separate norms may be necessary. However, gender differences in the areas of handwriting automaticity and orthographic coding were identified for children (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008). The CBM-Written Expression scoring procedures have been shown to also demonstrate gender differences. Deno et al. (1982) identified gender differences for the production-dependent methods of TWW and WSC in one of the first studies on CBM-Written Expression, where girls outperformed boys. Similarly, Malecki and Jewell's (2003) results indicated that first- through eighth-grade girls outperformed boys on all three types of scoring indices: production-dependent measures of TWW, WSC, and CWS; production-dependent measures of %WSC and %CWS; and the accurate-production measure of CIWS. However, a follow-up study by Jewell and Malecki (2005) reported significant gender differences only on the production-dependent methods of TWW, WSC, and CWS where, again, girls outperformed boys. They found no gender differences with the scoring methods of %WSC, %CWS, or CIWS. In a recent study, Ferrington et al. (2014) reported that third- through eighth-grade girls scored significantly greater than same-grade boys on TWW and CWS. While girls consistently

score higher on written expression probes than boys, Truckenmiller (2011) examined student growth for writing fluency with students eight to 10 years of age and found no gender differences for rate of growth.

### **Technical Adequacy of CBM-Written Expression**

While not overwhelming in number, the research literature has yielded a number of published studies examining the technical features of CBM-Written Expression. Most of the studies, however, have examined the assessment method's criterion validity, correlating the results of various scoring procedures from CBM-Written Expression with other standardized measures of writing ability. This section will provide a synopsis of criterion validity articles and a review of the test-retest reliability literature. McMaster and Espin (2007) conducted a literature review examining the technical adequacy of written expression measures and was the primary source for literature cited up to 2007. Studies since then were also included in this review of relevant studies.

**Criterion validity.** The bulk of the research on CBM-Written Expression has examined its criterion validity. Measures such as standardized tests and holistic ratings have been correlated with various CBM-Written Expression scoring procedures to determine if CBM-Written Expression is a valid measure of students' writing abilities. McMaster and Espin (2007) examined 21 of these validity studies. The following is a summary of the findings for the production-dependent measures of TWW, WSC, and CWS; the production-independent measure of %CWS; and the accurate-production measure of CIWS.

The TWW scoring procedure has been utilized for almost all criterion validity studies, as it was one of the scoring procedures first developed by Deno and colleagues in

the early 1980s. In four of the early studies completed in the 1980s and reported in technical reports, Deno and his colleagues found moderate to strong correlations between TWW and other writing assessment measures ( $r_s = .41$  to  $.88$ ) with small samples of students in grades 3 through 6. As summarized by McMaster and Espin (2007), later studies by other researchers found wide variations in correlation coefficients both at the elementary level ( $-.02$  to  $.63$ ) and secondary level ( $.10$  to  $.58$ ). The differences among studies may be due to the criterion measure used, the age of the students, or both. Making definitive conclusions is impossible due to so many variations, but the studies reviewed seemed to suggest TWW had lower criterion validity coefficients with students at higher grade levels (McMaster & Espin, 2007).

Much like the TWW scoring procedure, WSC has also been consistently tested in most criterion validity studies of written expression. In three of the four studies completed by Deno and colleagues mentioned previously, WSC was correlated with other standardized writing assessment measures and yielded moderate to strong coefficients ( $r_s = .41$  to  $.88$ ). These studies consisted of small sample sizes and included grades 3 through 6. The McMaster and Espin (2007) review discussed additional studies that have occurred since those original studies that have included WSC as a scoring procedure. Criterion validity coefficients were calculated for the elementary level ( $r_s = -.02$  to  $.64$ ) and secondary level ( $r_s = .08$  to  $.92$ ). Similar to the wide range of correlation coefficients found with TWW, the results for WSC suggest that coefficients were higher with younger students. Studies reporting validity coefficients between holistic ratings and TWW or WSC had the most variation in results (McMaster & Espin, 2007).

The CWS scoring procedure was used by Deno and colleagues in only one of the original CBM-Written Expression studies from the 1980s. They examined the correlations of CWS with five different criterion measures, yielding validity coefficients ranging from weak to strong ( $r_s = -.03$  to  $.85$ ) with holistic ratings yielding the strongest correlation with CWS. Additional studies summarized by McMaster and Espin (2007) yielded validity coefficients between CWS and a wide variety of standardized and holistic scoring methods for elementary students ( $r_s = -.02$  to  $.63$ ) and secondary students ( $r_s = .18$  to  $.99$ ). Again, drawing definitive conclusions from the wide variation of results would be difficult. However, of the studies reviewed, results suggest that CWS has higher criterion validity coefficients for secondary students. Lower validity coefficients were noted on correlations with measures such as English GPA and class grades in English and Social Studies (McMaster & Espin, 2007).

The %CWS scoring procedure was evaluated in only five of the criterion validity studies in McMaster and Espin (2007). One study at the elementary level with third and fourth grade students utilized this scoring procedure in correlations with holistic ratings, yielding moderate results ( $r_s = .43$  to  $.70$ ). At the secondary level, correlations between %CWS and holistic ratings and one standardized writing assessment yielded a range of coefficients from weak to strong ( $r_s = .28$  to  $.92$ ).

Like %CWS, the CIWS scoring procedure was evaluated in only five of the validity studies in McMaster and Espin (2007). For elementary-age students, criterion validity correlations between CIWS and Language Arts tests, Language Arts grades, state accountability assessments, and a research-based scoring system ranged from  $.36$  to  $.62$ .



At the secondary level, correlations between CIWS and holistic measures, district tests, and state accountability tests for grades 7 through 10 ranged from .56 to .82.

**Test-retest reliability.** The focus of this project is on the test-retest reliability of CBM-Written Expression because few studies have examined this basic aspect of technical adequacy. McMaster and Espin's (2007) review of the literature included only one study that examined test-retest reliability and it was from an unpublished technical report by Marston and Deno (1981). Two additional studies were located that examined the test-retest reliability of CBM-Written Expression. Of those three studies, all were at the elementary school level. While all three studies examined the scoring procedures of TWW and WSC, only two included CWS and only one included CIWS. None included any production-independent measures. The Marston and Deno (1981) and the Gansle et al. (2006) studies combined the grade levels in their studies to report the test-retest reliability coefficients. Thus, none of the studies provide such data for multiple grade levels. The results of those three studies are summarized in Table 1. In general, moderate to strong test-retest reliability coefficients were reported.

### **Purpose**

Most of the published studies on CBM-Written Expression have investigated the criterion-related validity of various scoring measures. Surprisingly, very few studies have examined test-retest reliability, a more basic aspect of technical adequacy. If schools are to use these brief assessment tools to identify at-risk students and as measures of progress monitoring, it is important that the measures are consistent over time for students in all grade levels. While three studies were identified that examined test-retest reliability of CBM-Written Expression, there are several limitations with those studies. Not only was

Table 1

*Summary of Studies Examining Test-Retest Reliability of CBM-Written Expression*

Study	N	Grade(s)	Test-Retest Time Interval	Scoring Procedures	Reliability Coefficients	
					3 Min.	5 Min.
Marston and Deno (1981)	28	1-6	1 day and 3 weeks	TWW	.64-.91	
				WSC	.62-.81	
Gansle, VanDerHeyden, Noell, Resetar, and Williams (2006)	190	1-5	1 week	TWW	.80	
				WSC	.82	
				CWS	.78	
McMaster, Du, and Pétursdóttir (2009)	50 47	1	1 month	TWW	.60-.61	.68-.78
				WSC	.63-.64	.65-.81
				CWS	.63-.72	.62-.83
				CIWS	.49-.74	.45-.75

*Note.* McMaster et al. (2009) reported the results of two concurrent studies. TWW = total words written; WSC = words spelled correctly; CWS = correct word sequence; CIWS = correct minus incorrect word sequences.

the Marston and Deno (1981) study conducted over 30 years ago, but they also had only 28 students across six grade levels. Due to the small sample size, results were not reported by grade level. Furthermore, they only evaluated two production-dependent scoring methods. The McMaster et al. (2009) study only included first grade students and had a lengthy time period (i.e., one month) between assessments. Surely, a fair amount of learning occurs in a school setting over a one-month period of time. The Gansle et al. (2006) study was the best in terms of the number of participants and grade levels measured, but only included production-dependent scoring measures and, like Marston and Deno, combined all grades together to report results. Furthermore, no test-retest reliability studies have examined accurate-production measures, such as %CWS.

It is imperative that further research on the test-retest reliability of these measures be conducted. Research data demonstrating the technical adequacy of assessment instruments are, and should be, the driving force behind decisions to utilize a specific assessment tool when determining a student's academic achievement ability and when making decisions for that student based on that assessment. Furthermore, acceptable reliability is the foundation for an assessment instrument's validity. Therefore, it is important that additional research be conducted to ensure that school systems are using valid and reliable measures for assessments.

Thus, the purpose of this study was to provide additional, much needed, test-retest reliability correlation data for the following CBM written expression scoring procedures: TWW, WSC, CWS, %CWS, and CIWS. The selection of these particular scoring procedures was made to ensure all three types of scoring methods (i.e., production-dependent, production-independent, accurate-production) were included. The production-dependent measures of TWW, WSC, and CWS were included because those measures are the ones most commonly used (Hosp et al., 2007). The focus on CWS with the production-independent measure (i.e., %CWS) and the accurate-production measure (i.e., CIWS) was made based on McMaster and Espin's (2007) conclusions from their review of the literature that indicated CWS was one of the stronger measures of written expression. To assess student performance consistency on CBM-Written Expression assessments, two CBM probes using the same story starter were administered to students in grades 2, 4, and 6 with a time span of one week between assessments. The following research questions were addressed:

1. Are TWW, WSC, CWS, %CWS, and CIWS reliable measures of scoring student performance across repeated measures? Test-retest reliability will be demonstrated by the results showing strong, significant correlations between the two administrations. Furthermore, given the relatively brief time span of one week, the scores should not be statistically significantly different between administrations.

2. Are there differences in test-retest reliability across grade levels? McMaster and Espin (2007) summarized studies that suggested certain scoring procedures are better at different grade levels. Large variations in correlations across grade levels would suggest there are age level differences with the scoring procedures.

## Method

### Participants

Students who participated in this study were from a school district located in northwestern Kentucky. The study utilized one elementary school serving 499 students in grades kindergarten through fifth grade and one middle school with approximately 839 students in grades six through eighth. The district had a student body that was comprised of 82% Caucasian students, 9% African American students, less than 1% each of Hispanic and Asian students, and 7% not otherwise specified. This school district had 58% of its students who received free or reduced lunch.

The number of classes per grade varied. In second grade there were four classes consisting of 80 students total, fourth grade had three classes with 67 students, and sixth grade had nine classes with a total of 292 students. Thus, the total number of students being afforded this opportunity to participate equaled 439 students.

The numbers of returned consent forms, indicating permission from parents for their child to participate in the study, were as follows: second grade  $n = 39$  (48.8%), fourth grade  $n = 28$  (41.8%), sixth grade  $n = 78$  (26.7%), with a total sample of  $N = 145$ . Overall, there was a 33.0% return rate. The return rate was hampered by a couple of factors. One fourth grade class that was taught by a long-term substitute teacher yielded only two consent forms returned during the assessment period of the study. Apparently, the teacher sent the consent forms home later than intended, as she provided additional returned forms during the last week of school. At that point, however, there was not time to include those students. One of the sixth grade teachers was absent the day the consent

forms were supposed to be distributed and, apparently, the forms never were sent home with the students because none was ever received from that one class.

### **Instrument**

This study selected an appropriate story starter from those provided by AIMSweb (2008). AIMSweb is a commercial system and database that provides nationally normed CBM assessment and progress monitoring probes. The company also provides percentile scores to allow school personnel to rate student performance. This system provides norms for grades 1 through 6 on 3-minute writing times. All AIMSweb CBM assessments have standardized administration procedures and instructions. The written expression probes can be administered individually or with a group of students. AIMSweb story starters are considered grade independent where the entire list of story starters can be used across multiple grades. AIMSweb CBM-Written Expression measures have yielded alternate form reliabilities ranging from .46 to .86 and inter-scorer agreements between .86 and .96. The story starter, "I opened the door very carefully and..." was selected for this study to use with all grade levels. The story starter was thought to be appropriate for students of all ages, gender, and socio-economic status.

### **Procedure**

**Consent.** Letters to the elementary and middle school principals and teachers were written to provide them with a description of the study's procedures and goals. The participating school district's Assistant Superintendent of Teaching and Learning reviewed those letters, as well as the consent and assent forms, and provided recommendations. The forms were revised to include all recommendations. Approval for this study was then obtained from Western Kentucky University's Institutional Review

Board. Teachers from all second, fourth, and sixth grade Language Arts classes were contacted via letter (see Appendix A) to provide information about the parameters, timelines, and teacher demands that would occur over the course of this study. Teachers were provided packets containing parent letters and consent forms for each student along with instructions for distribution and collection of the forms. Each parent was to receive a letter explaining the study stapled to a consent form (see Appendix B) that was to be signed and returned. Each child received a pencil upon return of the signed consent form, regardless of whether or not the parent or guardian granted permission to participate.

The consent forms were collected on two different occasions. Initially, returned forms were collected nine days after first being sent home. Then a second set of forms was distributed to teachers to give to the students who had not returned a form the first time. Any returned forms were collected a week later. Finally, assent forms (see Appendix C) were provided to those students with permission to participate. The administrator of the assessment read the assent forms aloud to the students, and encouraged the students to read the forms themselves, before indicating their voluntary participation.

**Administration procedures.** The examiners included a school psychology graduate student intern (specialist project author) and an interventionist within the district, both of whom had previous experience with administration of the CBM-Written Expression assessment measure. For the purposes of test-retest procedures, a one week time interval was used for administering the two identical written expression probes. On the assessment days, an examiner administered a story starter, class by class, for each grade level being examined. Upon entering a classroom on the first administration day,

the examiner placed a student assent form (faced up) and a sheet of lined paper with the story starter written across the top (faced down) on each participant's desk. The examiner read the assent form aloud and provided the students with time to sign their names on the assent form and write their names on the sheet of lined paper before beginning the assessment.

Next, the following AIMSweb (2008) instructions were read aloud to the students by the examiner:

You are going to write a story. First, I will read a sentence and then you will write a story about what happens next. 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 don't know how to spell a word, you should guess. Are there any questions? (Pause).

For the next minute, think about..."I opened the door very carefully and..." (p. 8).

The standardized administration instructions to the examiner continued, prompting the examiner to begin a stopwatch, allowing one minute for the students to think about what to write, and then prompting the students to begin writing. After three minutes, the examiner instructed students to stop, put their pencils down and flip their paper over. The examiners then collected all papers.

**Scoring procedures.** Four school psychologists from the school district, one of whom was an intern and author of this study, scored all the CBM probes. The raters spent three hours together to ensure consistent scoring. First, the CBM-Written Expression section of the AIMSweb technical manual (AIMSweb, 2008) was reviewed. Then, raters



practiced scoring probes as a group to determine any inconsistencies in scoring and decide as a group how to resolve any scoring concerns.

The students' names, schools, and grade levels were not visible to the scorers. To determine inter-rater agreement, every fifth probe was pulled from each rater and scored by a second rater. The percent of inter-rater agreement between raters for each of the scoring procedures was determined to be: TWW – 100%, WSC – 97.5%, CWS – 97.6%, %CWS – 94.5%, and CIWS – 93.3%. Such a high level of agreement is excellent.

## Results

The first research question sought to determine if the various CBM-Written Expression scoring procedures had adequate test-retest reliability. To address this question, Pearson correlations were calculated between the scores from the two assessments for each grade and the total sample with the five different scoring methods. The results are presented in Table 2. In order to be consistent with the descriptions of results provided in the literature review, the following descriptive classifications ranges were used: “strong”  $\geq .80$ , “moderately strong”  $.70$  to  $.79$ , “moderate”  $.60$  to  $.69$ , and “weak”  $< .60$  (McMaster & Espin, 2007, p. 69). Such descriptive terminology may be conservative, but higher correlations are expected for assessment instruments. Salvia and Ysseldyke (2001), for example, suggested that group-administered tests should have a minimum reliability coefficient of  $.80$ . McMaster et al. (2009), however, decided upon a minimum coefficient of  $.70$  as “sufficient” for CBM-Written Expression measures.

Table 2

*Test-Retest Correlations of CBM-Written Expression Scoring Procedures by Grade Level*

	TWW	WSC	CWS	%CWS	CIWS
Grade 2 ( $n = 39$ )	.68**	.71**	.68**	.62**	.60**
Grade 4 ( $n = 28$ )	.60*	.58*	.61*	.76**	.67**
Grade 6 ( $n = 77$ )	.71**	.69**	.73**	.71**	.73**
Total Sample ( $n = 144$ )	.82**	.82**	.84**	.74**	.81**

\* $p < .01$ . \*\* $p < .001$ .

Test-retest correlation coefficients for each scoring procedure at all grade levels were statistically significant. Results yielded strong coefficients for the total sample TWW, WSC, CWS, and CIWS. Moderately strong coefficients were the result for total sample %CWS; second-grade WSC; fourth-grade %CWS; and sixth-grade TWW, CWS, %CWS, and CIWS. Coefficients considered at a moderate level were found for second-grade TWW, CWS, %CWS, and CIWS; fourth-grade TWW, CWS, and CIWS; and sixth-grade WSC. A weak correlation was noted for fourth-grade WSC.

To further evaluate the consistency of measurement from Time 1 to Time 2, the means and standard deviations were calculated at each grade level for each scoring procedure and paired-samples *t* tests were completed to determine if any significant differences occurred between the administrations. A Bonferroni correction procedure was applied to the alpha level due to the large number of *t* tests. The correction procedure (i.e.,  $.05/15$ ) results in an alpha level of .003. Thus, to be considered significant, a *p* value needs to be below .003. Results are presented in Table 3. The results revealed that sixth grade students have statistically significantly higher scores with the scoring procedures of TWW, WSC, and CWS on the second assessment. These three scoring procedures are all production-dependent measures. No significant differences were found for any of the measures at grade 2 and 4 and no significant differences were found for %CWS and CIWS at grade 6.

The second research question sought to determine if there were differences in test-retest reliability across grade levels. The correlations do not provide clear-cut patterns across grade levels, although the scoring procedures at the sixth grade level contained twice as many moderately strong correlations (i.e., four of the five) than the other two

Table 3

*Means and Standard Deviations of CBM-Written Expression Scoring Procedures by Grade Level*

	<u>Time 1</u>		<u>Time 2</u>		<i>t</i> value	<i>p</i> value
	Mean	( <i>SD</i> )	Mean	( <i>SD</i> )		
Grade 2 ( <i>n</i> = 39)						
TWW	24.2	(9.3)	26.2	(9.9)	-1.64	.109
WSC	21.4	(9.2)	22.7	(9.6)	-1.17	.249
CWS	18.9	(10.3)	19.1	(9.5)	-0.18	.857
%CWS	70.5	(21.5)	66.7	(17.8)	1.32	.195
CIWS	11.9	(13.3)	10.1	(11.1)	1.04	.305
Grade 4 ( <i>n</i> = 28)						
TWW	40.7	(14.6)	47.1	(11.6)	-2.81	.009
WSC	37.6	(14.5)	43.4	(11.8)	-2.49	.019
CWS	36.3	(15.0)	41.8	(13.3)	-2.32	.028
%CWS	80.3	(14.5)	79.2	(16.5)	0.53	.604
CIWS	28.8	(16.4)	32.3	(16.1)	-1.42	.167
Grade 6 ( <i>n</i> = 77)						
TWW	51.0	(15.1)	56.1	(14.1)	-3.97	.000*
WSC	49.3	(14.9)	53.8	(14.8)	-3.44	.001*
CWS	48.7	(15.8)	53.4	(16.8)	-3.45	.001*
%CWS	87.5	(11.5)	86.0	(13.0)	1.46	.149
CIWS	41.9	(17.8)	45.1	(21.0)	-1.87	.065

\**p* < .003.

grade levels combined. On the other hand, no statistically significant differences in mean scores were found at the second grade level, perhaps suggesting the production-dependent measures might be better with younger grade levels than older grade levels.

## Discussion

The purpose of this study was to further the technical adequacy research base of CBM-Written Expression by determining whether production-independent (i.e., TWW, WSC, CWS), production-independent (i.e., %CWS), and accurate-production (i.e., CIWS) scoring procedures are reliable measures of scoring student performance on CBM probes. Only three previous studies (i.e., Gansle et al., 2006; Marston & Deno, 1981; McMaster et al., 2009) were located that provided test-retest reliability information on this form of assessment and those studies provided incomplete information due to limitations such as small sample size, few grade levels assessed, and few scoring measures tested. The current study was designed to provide test-retest reliability data by grade level.

For pragmatic purposes, having CBM reliability and validity data delineated for each grade-level promotes more accurate and meaningful results when using these forms of measurement. For instance, a second-grade student receiving additional writing instruction in an RtI program may more accurately be assessed with a TWW or WSC scoring procedure, compared to a sixth-grade student with writing difficulties for which use of %CWS may yield more accurate information. Users must be equipped with this knowledge in order to make executive decisions on which procedures to use for specific grade levels.

The CBM-Written Expression scoring methods used in this study resulted in strong reliability scores for four of the five scoring procedures, and moderately strong for the fifth one, when all grade levels were combined. However, combining the entire sample may provide inflated correlations, as grade level differences in scores naturally occurred. Correlations compare the rank order of scores from two administrations. Given

most fourth graders are going to score higher than second graders, and most sixth graders will score higher than the younger students, the rank order of scores for the total sample may provide a misleading picture of the strength of test-retest relationship. With that in mind, an examination of results for the individual grades is likely to provide a better picture of the test-retest relationship and if there are differences by grade level.

In general, the correlation coefficients seem to indicate a high level of test-retest reliability for all measures at all grades. Only one of the correlations was considered weak on the previously cited descriptive categories, but that correlation was just barely below the .60 correlation described as the minimum needed for a moderate level correlation. However, if Salvia and Ysseldyke's (2001) recommendation of a minimum reliability of .80 for group-administered tests is applied to the results, none of the scoring methods at any specific grade level met that level. If using McMaster et al.'s (2009) minimum level of .70 as what is considered sufficient, then only WSC in second grade, %CWS in fourth grade, and all scoring methods except WSC in sixth grade are considered to have sufficient test-retest reliability. The results of the current study yielded results similar to the TWW and WSC reliability coefficients reported by Marston and Deno (1981). This study's total sample's reliability coefficients were very similar to the total sample results for TWW, WSC, and CWS produced by Gansle et al. (2006). Interestingly, when comparing this study's second-grade results with McMaster et al.'s (2009) first-grade sample, a similar pattern of results was noted. Test-retest reliability coefficients were in the moderate range for TWW and CWS, with weak reliability coefficients for CIWS. Such results imply that the method of CIWS may not be the best scoring choice for early primary students.

The correlation coefficients did not yield any particular pattern of results across grade levels or scoring method. The correlations were a little higher at the sixth grade level, but there were exceptions. WSC had a slightly higher correlation at grade 2 than grade 6 and %CWS had a slightly higher correlation in grade 4 than grade 6. Of the five different scoring procedures, none stood out as showing lower or higher correlations.

The analysis of the means at Time 1 and Time 2 provide a different perspective on the results than the correlations provide. Given the small time frame between test administrations, there is a strong likelihood that no changes in curriculum or instruction occurred. Thus, it would be expected that the mean level of student performance would be the same on both administrations, especially considering it was the same story starter or topic given to the students. While the results demonstrated that scores from Time 1 to Time 2 were statistically similar for all scoring procedures at second grade and for the production-independent and accurate-production measures at fourth and sixth grades, the production-dependent measures of TWW, WSC, and CWS yielded significantly greater scores on the second administration for fourth and sixth graders. Such results may provide additional support to previous studies that concluded that TWW and WSC are not sufficient for assessing secondary grades (Parker et al., 1991; Tindal & Parker, 1989) and the production-dependent measures may be less useful in upper primary grades (McMaster & Espin, 2007).

An alternative explanation is that the results with the production-dependent measures may simply be due to practice effects that occur naturally with test-retest administration procedures. Anecdotal evidence for this occurred during the second CBM probe administrations. There were students who asked the examiners if they could write



the same thing they wrote for the initial session. The act of reusing stories may have been an advantageous strategy where students' use of the one minute of allowed time to think about what to write could then be utilized to formulate additional components to the story instead of generating an entirely new story. Given TWW, WSC, and CWS are production-dependent scoring methods where higher scores result from producing more text, it makes sense that the scores would be higher on the second administration of the same probe. The effect may not have been seen with the second-grade students due to less advanced higher level thinking abilities required to yield significantly different performance on the second administration (e.g., remembering the initial story they wrote, creating additional components to a story).

### **Limitations and Future Research**

The main limitation of this study was the small sample of fourth-grade participants and, to a lesser extent, the relatively small sample of second-grade students. A larger number of participants in future studies would provide more definitive results. Similarly, a second limitation to this study was the limited range of grades assessed. The original conceptualization of the study design included eighth grade students. However, the Assistant Superintendent of the district requested that eighth grade students not be included due to the extensive testing demands already being requested of that particular group. A wider range of participants, particularly at the secondary level, may provide more insight into the usefulness of the different scoring methods at all grade levels.

As noted in the Literature Review, gender differences with written expression skills have been found in a few studies; however, nothing is known about potential gender differences in terms of test-retest reliability correlations. For that reason, one

limitation of this study was that gender differences could not be assessed as gender data were not collected effectively for all grade levels in this study. Future research may wish to consider evaluating gender differences of student performance on test-retest reliability measures.

This study evaluated the test-retest performance at the end of a school year. Future research might evaluate if the time of the year that students are being assessed makes a difference. It is possible that students might be more or less motivated to write at certain times of the year. In addition, it was noted that the length of time allowed for students to write varied by author. The length of writing time may have an impact on test-retest reliability. Future research might examine varying lengths of time students are allowed to write.

Finally, one additional area for future research to consider would be the use of word processing programs for student probe completion. Due to some of the writing samples containing illegible printing, there are times when the written expression scoring procedures require inferences about what the student wrote. As examples, there can be scoring errors when there is difficulty with such things as illegible printing, making determinations about correct capitalization of words, and determining if there is a spacing error or simply a misspelled word. Future research is needed to determine if word processing programs might address some of those concerns and lessen the potential for scoring error.

## **Conclusion**

This study appears to be the first test-retest reliability study to provide results for individual grade levels. Results provide empirical support for the use of all five scoring procedures across grade levels. However, correlation coefficients and mean differences

would suggest that some scoring procedures may be better than others to administer at different grade levels. Specifically, production-dependent measures may better assess early-elementary aged students while production-independent and accurate-production measures may be more effective for scoring late-elementary and secondary student performance.

## **Appendix A**

### **Letter to the Teachers**

April 2, 2014

Dear Teacher:

I am planning a written expression research project that will be conducted April 24, 2014 through May 1, 2014. The project has been approved by Ms. Swanson and your principal Ms. Watson. This packet contains enough informative letters and consent forms for each of your students to take one home to his/her parent or guardian. The testing procedures being used will only take about 3 minutes and will be conducted either in your classroom and will take approximately 7 to 10 minutes altogether.

If you are willing to let your students participate, please send a copy of the letter and consent form home with them today (April 2). This is completely voluntary, but in order to stay on schedule and have an accurate count, it is important that the consent forms are signed and returned before April 24, 2014. I will be checking in with you on April 23, 2014 to pick up the forms that have already been returned. After Spring Break, I will provide additional consent forms as a reminder for those that have not been returned. If you need to contact me with any questions or comments, please feel free to email or call me at any time.

Thank you so much for your time, help, and consideration!

Sincerely,

Mallory Hart  
mallory.hart@henderson.kyschools.us  
(270) 831-5040

## Appendix B

### Parent Letter and Consent Form

Dear Parent/Guardian:

Your child is being asked to participate in a project that will be comparing two brief measures of writing. This study is being conducted by Mallory Hart, a school psychology intern employed by Henderson County Schools, and is supervised by Dr. Carl Myers at Western Kentucky University. The University requires that you give your signed agreement for your child to participate in this project.

In this study, we will be looking at a brief measure of writing to see if it is reliable over time. Students will be asked to write for 3 minutes on two different days approximately one week apart. The writing sample would be obtained in your child's classroom during regular school hours. We will be working with your child's teacher to ensure the least amount of disruption from his or her regular school routine.

All information collected in this study will be kept strictly confidential and is accessible only to the project staff. Data will be stored with a code number, not your child's name. Only an overall summary of results will be shared with school personnel, not individual results.

We emphasize that your child's participation in this project is completely voluntary and you may withdraw your consent at any time. In addition, your child will be asked to give his or her assent to participate, and may decide not to participate. If you or your child decides not to participate, it will have no negative outcome for you or your child in any way. Given the task is a commonly expected one in a school setting, we anticipate no discomfort or risks as a result of your child's participation in this study.

The procedures in this study have been reviewed and approved by Western Kentucky University's Institutional Review Board. Any questions about this study may be directed to Mallory Hart at (270) 705-5185 or Dr. Carl Myers at (270) 745-4410. Paul Mooney, Human Protections Administrator (270) 745-2129.

We hope that you will allow your child to take part in our study. We promise to make it a pleasant experience for your child and to schedule the assessment sessions in cooperation with your child's teacher. Please fill in your child's name, your child's date of birth, the name of your child's homeroom teacher and grade level on the attached form. To indicate your consent, check the "yes" box, sign your name and fill in the date. Please return the attached form to your child's homeroom teacher within two weeks. When your child returns this letter to the teacher, whether you check yes or no, your child will receive a small reward (e.g., pencil).

Thank you for your help.

Mallory Hart  
School Psychology Intern  
Henderson County School District

Carl Myers, Ph. D., Supervisor  
Associate Professor of Psychology  
Western Kentucky University



WKU IRB# 14-312  
Approval - 3/31/2014  
End Date - 7/25/2014  
Full Board  
Original - 3/31/2014

WESTERN KENTUCKY UNIVERSITY  
PARTICIPANT CONSENT FORM

Child's Name: \_\_\_\_\_ Date of Birth: \_\_\_\_\_

Teacher's Name: \_\_\_\_\_ Current Grade Level: \_\_\_\_\_

\_\_\_\_\_ Yes, I have read the information provided about this study, and give my consent for my child to participate in the study conducted by Mallory Hart and Dr. Carl Myers of Western Kentucky University. I understand that my child or I may withdraw from the study at any time without penalty.

\_\_\_\_\_ No, I do not give my consent for my child to participate in this project.

\_\_\_\_\_  
Parent/Guardian Signature

\_\_\_\_\_  
Date

\*Please return this form to your child's teacher within two weeks.

When this form is returned, whether it is checked yes or no, your child will receive a small reward.

THE DATED APPROVAL ON THIS CONSENT FORM INDICATES THAT  
THIS PROJECT HAS BEEN REVIEWED AND APPROVED BY  
THE WESTERN KENTUCKY UNIVERSITY INSTITUTIONAL REVIEW BOARD

Paul Mooney, Human Protections Administrator  
TELEPHONE: (270) 745-2129



WKU IRB# 14-312  
Approval - 3/31/2014  
End Date - 7/25/2014  
Full Board  
Original - 3/31/2014

# Appendix C

## Assent Form

### INFORMED ASSENT DOCUMENT

I, \_\_\_\_\_, understand that my mom or dad has said it is okay for me to take part in a project about writing under the direction of Mallory Hart. I will be asked to write a story for 3 minutes.

I am taking part because I want to. I have been told that I can stop at any time I want to, and nothing will happen to me if I want to stop.

Signature \_\_\_\_\_ Date \_\_\_\_\_

THE DATED APPROVAL ON THIS CONSENT FORM INDICATES THAT  
THIS PROJECT HAS BEEN REVIEWED AND APPROVED BY  
THE WESTERN KENTUCKY UNIVERSITY INSTITUTIONAL REVIEW BOARD  
Paul Mooney, Human Protections Administrator  
TELEPHONE: (270) 745-2129



WKU IRB# 14-312  
Approval - 3/31/2014  
End Date - 7/25/2014  
Full Board  
Original - 3/31/2014

## References

- AIMSweb. (2008). *AIMSweb progress monitoring and RTI system*. Retrieved from <http://www.aimsweb.com/>
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E. W., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology, 46*, 151-172.
- Deno, S. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184-192.
- Deno, S. L., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study* (Vol. IRLD-RR-87). University of Minnesota, Institute for Research on Learning Disabilities.
- Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The Journal of Special Education, 34*, 140-153.
- Fearrington, J., Parker, P., Kidder-Ashley, P., Gagnon, S., McCane-Bowling, S., & Sorrell, C. (2014). Gender differences in written expression curriculum-based measurement in third-through eighth-grade students. *Psychology in the Schools, 51*, 85-96.
- Fuchs, L., & Fuchs, D. (2007). *Using CBM for progress monitoring in written expression and spelling*. Retrieved from <http://studentprogress.org/weblibrary.asp>
- 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., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York, NY: The Guilford Press.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: The Guilford Press.
- Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1999). *Models of curriculum-based assessment: A blueprint for learning* (3<sup>rd</sup> ed.). Austin, TX: Pro-ed.
- 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.
- Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools*, 40, 379-390.
- Marston, D., & Deno, S. (1981). *The reliability of simple, direct measures of written expression* (Vol. IRLD-RR-50). University of Minnesota, Institute for Research on Learning Disabilities.

- McMaster, K. L., Du, X., & Pétursdóttir, A. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities, 42*, 41-60.
- McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education, 41*, 68-84.
- National Center on Student Progress Monitoring. (n.d.). *Fact sheet: Benefits of curriculum-based measurement*. Retrieved from <http://www.studentprogress.org/family/factsheet.asp>
- National Commission on Writing. (2003). *The neglected R: The need for a writing revolution*. New York, NY: College Board.
- National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common core state standards*. Washington DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.
- Parker, R., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality, 2*, 1-17.
- Riley-Tillman, T. C., Burns, M. K., & Gibbons, K. (2013). *RTI applications: Assessment, analysis, and decision-making*. New York, NY: Guilford Press.
- Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston, MA: Houghton Mifflin.
- 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: The Guilford Press.

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

Truckenmiller, A. J. (2011). *Modeling elementary aged students' fluency growth in written expression: Predicting fluency growth for girls and boys in general education*. (Order No. 3454429, Syracuse University). *ProQuest Dissertations and Theses*, 174. Retrieved from <http://search.proquest.com/docview/869990114?accountid=15150>

Watkinson, J. T., & Lee, S. W. (1992). Curriculum-based measures of written expression for learning-disabled and nondisabled students. *Psychology in the Schools*, 29, 184-192.