8-2014

Alternate Forms Reliability for Written Expression Probes

Ashley Carey
Western Kentucky University, ashley.carey@hardin.kyschools.us

ALTERNATE FORMS RELIABILITY FOR 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
Ashley Carey

August 2014

ALTERNATE FORMS RELIABILITY FOR WRITTEN EXPRESSION PROBES

Date Recommended July 29, 2014

Carl Myers

Carl Myers, Director of Specialist Project

Anthony Paquin

Ronda Talley

Dean, The Graduate School 8-7-14 Date

TABLE OF CONTENTS

LIST OF TABLES

Page

ALTERNATE FORMS RELIABILITY FOR WRITTEN EXPRESSION PROBES

Ashley Carey August 2014 41 pages

Directed by: Dr. Carl Myers, Dr. Anthony Paquin, and Dr. Ronda Talley

Department of Psychology Western Kentucky University

The use of Curriculum-Based Measurement (CBM) in schools continues to increase, as it is a useful brief assessment of students' basic academic skills. CBM measures are used for multiple tasks such as identifying students at-risk, creating local norms, monitoring students' progress during interventions, and assisting with special education eligibility determinations. Much of the research has focused on CBM in the areas of math and reading. Relatively few studies have examined the area of CBM-Written Expression. Even fewer studies exist exploring the reliability among alternate writing forms. This study determined alternate form reliability coefficients for written expression probes at the second, fourth, sixth, and eighth grade levels using production-dependent, production-independent, and accurate-production scoring methods. When all grade levels are combined, alternate forms reliability coefficients are at a sufficiently high level. However, some scoring methods resulted in much higher correlations at younger grade levels than older grade levels. In general, the correlations were lower at the eighth grade level. Implications of the results for school personnel are discussed.

Introduction

Just as reading and math are essential to daily functioning as adults in society, writing has become an essential tool of everyday life. Writing is utilized in school and work environments; thus, it is vital for people to learn to effectively communicate with others through written expression. More than a decade ago, the National Commission on Writing called for a "writing revolution," requiring states to set comprehensive writing standards and to incorporate writing instruction (National Commission on Writing, 2003). According to assessments completed by the National Assessment of Educational Progress in 2002, 72% of fourth-graders, 69% of eighth-graders, and 77% of twelfth-graders were writing below the proficient level in writing (National Commission on Writing, 2003). The proficient level was defined as clearly demonstrating the ability to accomplish the communicative purpose of writing. More recently, in 2011, 76% of students in 8th and 12th grades performed below the proficient level in writing (National Commission on Writing, 2011). Thus, no progress is being made at increasing students' writing proficiency.

In Kentucky, the Kentucky Performance Rating for Educational Progress (K-PREP) state assessment results indicated that 67.4% of elementary school students, 63.6% of middle school students, and 58.3% of high school students did not meet criteria for the proficient level in writing during the 2012-2013 school year (Kentucky Department of Education, 2014). It is necessary for educators to address this issue as demands for writing continue to increase in the classroom, in the workplace, and on state-mandated assessments for all grade levels. To help develop the writing skills of students who demonstrate difficulties effectively writing, early identification and intervention is

fundamental to their success as a student and their future as contributing adults in society (McMaster et al., 2011).

The Response to Intervention (RTI) approach emphasizes early identification and intervention in school systems and has been widely adopted in U.S. schools (Brown-Chidsey & Steege, 2010). Early identification of struggling students is accomplished through universal screening, where all students are assessed with brief measures, usually three times a year. Those identified received successively intense interventions. Progress monitoring, consisting of frequent and brief measures of academic skills, has become the preferred method to track student progress in response to an intervention (Riley-Tillman, Burns, & Gibbons, 2013). RTI requires brief, but valid tools for universal screening and progress monitoring purposes (Cocker & Ritchey, 2010).

McMaster and Espin (2007) note that one of the most extensively researched universal screening and progress monitoring assessment methods, at least in the area of reading, is Curriculum-Based Measurement (CBM). CBM uses brief fluency measures to assess students' basic academic skills in the areas of reading, math, spelling, and writing (Shinn, 1998). McMaster and Espin described CBM as a procedure in which multiple probes of equivalent difficulty are administered repeatedly, yielding time-series data that reflect student growth. School systems are increasingly relying on CBM methods to monitor students' academic growth. CBM data can also be used to create local norms as a way to measure students' achievement (Jewell & Malecki, 2005). Deno (2003) further notes CBM data have been used in a wide range of assessment activities such as screening, pre-referral evaluation data, placement in remedial and special education programs, formative evaluation, and evaluation of reintegration and inclusion.

Recently, RTI and CBM have become a necessity for the identification of students with learning disabilities, based on the changes in the diagnostic criteria for a learning disability in the Individuals with Disabilities Act of 2004 (IDEA-04). The change directed schools to focus more on helping all children learn by addressing problems earlier, before the child is so far behind that a referral to consider special education services is needed. The main components of an RTI approach are the provision of scientific, research-based instruction and interventions in general education; monitoring and measurement of student progress in response to the instruction and interventions; and use of these measures of student progress to shape instruction and make educational decisions (Klotz & Canter, 2006).

One of the academic areas addressed through the process of RTI is writing. Many aspects contribute to successful and effective writing. It is important for educators to know what factors can assist student writing skills, and how to improve those skills. Although an abundance of supporting evidence exists for CBM in the area of reading, and a moderate number of studies have focused on mathematics, very few studies exist for the area of written expression. As stated previously, CBM measures are administered repeatedly on a relatively frequent basis (e.g., weekly to monthly). These frequent administrations involve multiple alternate forms of probes. Thus, it is important to determine if these alternate forms are consistent measures of the same construct. Indeed, a review of the CBM-Written Expression literature by McMaster and Espin (2007) indicated that very few studies have been conducted that examined the basic technical adequacy components (e.g., reliability, validity) of CBM-Written Expression. Therefore,

an examination of one of the basic technical adequacy components, specifically alternate forms reliability, is the focus of the current study.

Specifically, this study assessed the alternate forms reliability coefficients for CBM-Written Expression probes using a sample of elementary and middle school students in the second, fourth, sixth, and eighth grades. Within one week, two CBM writing probes were administered to the participants. Those probes were then scored with five of the most popular scoring methods, and the correlations between the two probes were determined to evaluate the consistency of measurement. Using multiple scoring methods and having a sample of multiple grade levels allows an evaluation of potential differences depending on the method used and/or age of the students.

Literature Review

The literature reviewed in this thesis focuses on aspects of CBM, with an emphasis on the area of Written Expression. First, an overview of Response to Intervention (RTI) and CBM is provided. Then, a descriptive depiction of CBM-Written Expression procedures is presented. Because this project focuses on the reliability of alternate test forms, research studies exploring the validity and reliability of CBM-Written Expression are reviewed, with an emphasis on studies that have assessed alternate forms reliability.

Response to Intervention

It is important to review RTI to provide a context for the importance of CBM. RTI can be defined as a high-quality teaching and assessment method, in a data-based systematic way, in which students who are not successful when presented with one set of instructional methods can be given a chance to succeed using other instructional practices (Brown-Chidsey & Steege, 2010). RTI typically encompasses a three tier process. The first tier contains approximately 80% of students that are able to be successful with high quality research-based general instruction. For students that are not successful with Tier 1 instruction alone, Tier 2 consists of interventions added to the general instruction. Interventions typically consist of additional instruction, usually in small groups, but can be defined as any behavior and/or academic activities used to help students (Brown-Chidsey & Steege, 2010). Tier 3 is utilized for students that are not demonstrating significant success with Tier 1 and Tier 2 interventions. Typically, interventions increase in intensity, and/or the amount of time provided, as the student moves through the tiers. Tier 3 may consist of Tier 1 and Tier 2 procedures with increased intensity (e.g., the same

interventions are provided more frequently and/or for longer periods of time) or may consist of completely different strategies taught by highly skilled educators.

Throughout the tiers, progress monitoring is utilized to make systematic, data-based decisions about students' progress at achieving educational goals. A lack of progress indicates the intervention is not effective and needs to be changed in some manner. Methods used to track the progress students make at achieving their goals during RTI must be brief, so as not to interfere with instructional time, and consist of multiple versions due to the frequency of measurement. According to Fuchs and Fuchs (1997), CBM is commonly used because its qualities meet the specifications of a good progress monitoring tool. CBM will be described in more detail in the next section but its utility extends to many uses, such as estimating rates of improvement, identifying students who are not demonstrating adequate progress and therefore require additional or alternative forms of instruction, and comparing the effectiveness of different forms of instruction (Hosp, Hosp, & Howell, 2007). The use of CBM as a progress monitoring tool in a RTI model has also become a staple in identifying students to refer for comprehensive evaluations to determine eligibility for special education services.

Curriculum-Based Measurement

An overview of CBM in general will be provided and then CBM-Written Expression will be described. CBM is used to assess basic skills in reading, mathematics, spelling, and written expression (Shinn, 1998). Each academic area assessed by CBM has its own materials, instructions, and scoring guidelines to ensure standardized administration and scoring. Information regarding the use of CBM can be found in multiple sources such as AIMSweb (2008) and Hosp et al. (2007).

Stanley Deno is one of the pioneers in creating CBM. According to Shinn (1998), in the late 1970s, Deno wanted to provide his special education student teachers with efficient yet accurate methods of assessing the effects of their instruction on students' academic skills. Deno strived to provide his teacher trainees academic measures that could be collected daily, graphed, and evaluated for evidence of student learning within short periods of time. While at the Minnesota Institute for Research on Learning Disabilities, and with the assistance of Phyllis Mirkin, Deno developed the first formal conceptualization of CBM called Data Based Program Modification.

According to Deno (1992), the purpose of CBM is to enable teachers to improve student performance. Additionally, the primary assumption of CBM is that it will be used to create a database for each student to allow teachers to evaluate the effectiveness of an individual student's educational program. Furthermore, the goal of this individual student monitoring is to create a formative evaluation framework in which teachers can systematically test alternative approaches to instruction for individual students. CBM allows for setting goals, monitoring growth, changing programs, and evaluating the effects of the changes made for students (Deno, 2003).

Riley-Tillman et al. (2013) explained the process of CBM. When implementing CBM methods, the administrator samples the student's performance on parallel probes as much as two to three times a week and graphs the data in a chart. The teacher then inspects the graph and uses a set of decision criteria to determine if the student is making sufficient progress, and if the instructional program is effective. If the student is not making sufficient progress, the student's teacher implements a change in the instruction and uses additional data to evaluate that change.

Deno (2003) described the specific characteristics of CBM as: (a) generally, it is technically adequate, (b) it has standard administration and scoring guidelines, (c) it has procedures for stimulus selection and performance sampling, (d) it consists of multiple equivalent samples, (e) it is time efficient, and (f) easy to teach to those using it. The standard CBM tasks for reading includes reading aloud from text for one minute to determine the number of correct words read per minute to assess reading skills. Typical CBM writing tasks encompass writing a story within a 3-minute time limit when given a story starter or picture, and writing orally dictated spelling words. CBM math assessments involve students correctly answering computational math problems for two minutes. Administrators of CBM are required to follow standardized administration and scoring procedures (e.g., instructions, time limits). All CBM scores are obtained by counting the number of correct responses given during a fixed time period, resulting in a fluency measure of academic skills.

CBM serves multiple purposes. CBM can be used to create norms for a specific school building or district, measure students' achievement, and monitor progress in the academic areas of reading, written expression, spelling, and math (Jewell & Malecki, 2005). Deno (2003) listed the common uses of CBM as: (a) improving individual instructional programs, (b) predicting performance on important criteria, (c) enhancing teacher instructional planning, (d) developing local norms, (e) increasing ease of communication (e.g., using CBM graphical data to explain to parents their student's progress), (f) screening the identification of students academically at risk, (g) evaluating classroom pre-referral interventions, (h) reducing bias in assessment (e.g., minority students being inappropriately placed in special education based on possible bias in

assessment measures), (i) offering alternative special education identification procedures, (j) recommending and evaluating inclusion of students with disabilities in the general education setting, (k) predicting performance on high-stakes assessment, (l) measuring growth in secondary school programs and content areas, (m) assessing English Language Learner students, and (n) predicting success in early childhood education.

CBM has numerous advantages and benefits for its many uses in the educational setting. For example, CBM procedures are easily taught to professionals, paraprofessionals, and parents. Evidence exists that increased measurement frequency is directly related to improved test scores (Mirkin, Deno, Tindal, & Kuehnle, 1982). According to research conducted by Fuchs and Fuchs (1986) with mildly and moderately disabled students, the use of systematic measurement and data-evaluation procedures increases average achievement by seven-tenths of a standard deviation over performance of students whose teachers do not use these procedures. Moreover, CBM provides clarity for teachers and students about what the key indicators of growth and basic skills are in academic skills (Deno, 1992).

Commercially available, norm-referenced achievement tests have also been designed to assess students' academic skills. However, achievement tests are lengthy to administer and the administrator must have specialized expertise. Thus, the use of such tests may not be feasible to monitor students' growth. Furthermore, norm-referenced achievement tests are designed to compare a student's performance to a national norm sample and, as such, are not designed to assess short-term growth (Deno, 1992; Shinn & Bamonto, 1998). For example, a student could be administered an achievement test resulting in skills at the 5th percentile. If that same child is tested again a few months later,

results will most likely be the same, at the 5th percentile. This information does not provide a demonstration of growth in the specified area. An advantage of CBM measures is that they are designed to assess short-term growth.

Another advantage of CBM is the graphic representation of changes in student's performance over time. These graphs reveal the past, present, and probable future growth rate of an individual student. Additionally, the graphs provide multiple avenues of interpretation. CBM graphs pictorially depict the student's current performance in comparison to recent and long-term past performances, provides a goal reference in that the student's current level of performance and rate of improvement can be viewed relative to his or her goal, provides a means of analysis of performance which can decipher whether or not the goal is attainable if the conditions remain constant, and when peer performance data are presented on the graph, norm referencing is possible (Deno, 1992).

CBM-Written Expression

CBM-Written Expression can be used with students in grades 1-12 (Fuchs and Fuchs, n.d.). According to Hosp et al. (2007), minimal materials are needed to conduct CBM-Written Expression. Once a quiet environment for students to work in has been established, the administrator needs a stop-watch or some type of time keeping tool, standardized directions for the administration, and writing materials for the student(s) to use (e.g., lined paper and pencil). Administrators can then record the student's performance on an equal-interval graph or a graphing program. Story starters are used as prompts to give the students a topic to write about. Story starters are short, oral or written sentences that begin the writing process. Hosp et al. recommends that story starters be

equivalent in grade level, difficulty, and interest; however, guidelines on how to ensure the equivalency of story starters are not mentioned.

Administration can be done individually or in a group. To administer CBM-Written Expression, lined paper and a pencil is provided to the student. The administrator reads a script indicating what the child is required to do (i.e., write a story), and how long they have to complete the task (e.g., 3 minutes). After the story starter is given, the administrator allows 1 minute for the child to think about their answer. At the end of that minute, the child is instructed to begin writing, and is typically allotted 3 minutes to write (Hosp et al., 2007).

According to Hosp et al. (2007), the scoring of CBM-Written Expression is typically completed utilizing three scoring procedures: Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Writing Sequences (CWS). TWW is the number of words, or groups of letters, written regardless of spelling or context. WSC is defined as the number of correctly spelled words, regardless of context. CWS is two adjacent, correctly spelled words that are acceptable within the context of the written phrase to a native speaker of the English language. The three measures are called production-dependent measures, because higher scores are dependent upon writing or producing more words and sentences.

While not as common, production-independent indices (e.g., Percentage of Words Spelled Correctly, Percentage of Legible Words, and Percentage of Words Correctly Sequenced) have also been developed. These scoring methods are a measure of writing accuracy, as scores are independent of the length of the writing sample. Tindal and Parker (1989) examined production-independent indices. Results of their study indicated that

Percentage of Correct Word Sequences (%CWS) was strongly correlated to teachers' holistic ratings of students' writing (r = .75). Percentage of Legible Words and Percentage of Words Spelled Correctly were weakly correlated to teachers' holistic ratings of students' writing (r = .10 and .24, respectively).

Furthermore, accurate-production indices (e.g., Correct Minus Incorrect Writing Sequence, CIWS) measure both writing fluency and accuracy (Jewell & Malecki, 2005). Few studies, however, have examined the CIWS scoring method. One study that did include CIWS found internal consistency coefficients ranging from .72 to .78 for sixth and seventh grade students (Espin et al., 2000).

Other mechanisms of scoring written expression (e.g., characters per word, number of words per sentence, number of sentences written) have been attempted at different grade levels, but have not been found to be valid and useful measures of writing (McMaster & Espin, 2007). In general, Espin et al. (2000) indicated that research at the elementary school level has demonstrated that TWW, WSC, and CWS are valid and reliable indicators of students' general performance in written expression. However, research has indicated that at the secondary level, two of the most commonly used CBM scoring metrics (TWW and WSC), are not appropriate for the secondary level. The research is too sparse to make any such conclusions about the production-independent and accurate-production measures.

One additional interesting finding about CBM-Written Expression is that a few studies have found gender differences when comparing the writing ability of boys and girls. However, findings have been consistent. A study conducted by Malecki and Jewell (2003) indicated that girls outperform boys on production-dependent, production-

independent, and accurate-production indices for grades 1 through 8. Jewell and Malecki (2005) also reported that girls outperform boys on writing fluency tasks at the second, fourth, and sixth grade levels.

The Technical Adequacy of CBM-Written Expression

Validity and reliability are essential to an assessment measure's technical adequacy. Thorndike (2005) defined validity as how well a measure assesses what it is intended to measure. The validity of CBM is essential to the many uses of CBM. Research has been conducted to explore the criterion validity of CBM-Written Expression and some of those studies will be highlighted in this section. Few studies address the area of reliability concerning writing. Thorndike (2005) referred to reliability as the precision, accuracy, and consistency of measurement procedure. It is imperative that CBM measures are valid and reliable as important decisions are made regarding students' education based on that data. For this literature review, a few of the research studies evaluating the criterion validity of CBM-Written Expression are briefly reviewed to establish its relationship to other measures of reading. Then, the studies including alternate forms reliability of CBM-Written Expression are reviewed.

A limited amount of studies address the area of CBM-Written Expression, especially ones involving reliability. McMaster and Espin (2007) conducted a literature review concerning research articles examining technical features of CBM-Written Expression. Out of the 172 articles found discussing CBM in the areas of reading, spelling, math, and writing, only 28 articles were found to include research about written expression in regards to reliability and validity. Only three of the published studies addressed alternate forms reliability and most of those only assessed a few of the CBM

scoring methods. Three additional studies that assessed alternate forms reliability since McMaster and Espin (2007) are also reviewed.

Criterion validity. Espin et al. (2000) examined the validity of CBM-Written Expression with middle school students. This study was one of the earliest studies addressing this topic. The researchers also analyzed if different types of writing samples would result in more valid measures. Two probes with story starters and two probes with descriptive writing samples were administered to 112 seventh and eighth grade students. The students completed their replies on a computer. The participants were given the prompt, 30 seconds to think of what they wanted to write, and three minutes to write. The following methods were used to score the writing passages: TWW, WSC, CWS, CIWS, Words Spelled Incorrectly, Characters Written, Sentences Written, Characters Per Word, Words Per Sentence, and mean length of correct word sequences. The criterion variables included teacher ratings of the students' writing skills, and scores obtained on a district-wide writing assessment. Results indicated that CIWS and CWS were the strongest measures of writing compared to teachers' ratings of writing (r = .66 and .59, respectively). Furthermore, results indicated that the scoring methods of Words per Sentence (r = .74), Sentences Written (r = .72), CIWS (r = .69), and CWS (r = .61) correlated the highest with the district-wide writing assessment. On the district-wide writing assessment, students' writing was scored on a scale from 1 to 4, with 1 signifying "poor writing" and 4 symbolizing "excellent" writing. The ratings were based on 3 categories that included mode and organization, sentence structure, spelling, and handwriting.

Fewster and MacMillan (2002) examined whether middle school students' written expression performance was predictive of high school performance. Participants, consisting of 465 students initially in grades 6 and 7, were administered CBM-Written Expression probes. Teachers used the scoring methods of TWW and WSC to score the probes. The students' CBM scores were later compared to their end of the year English and Social Studies grades in grades 8, 9, and 10. Results indicated that CBM scores reliably distinguished among students in special education, remedial, general education, and honors classes. The authors concluded the CBM measures were valid indicators of academic achievement.

Gansle, Noell, VanDerHeyden, Naquin, and Slider (2002) conducted a study to assess teachers' anecdotal concerns that the traditional scoring methods of CBM-Written Expression (i.e., TWW, WSC, & CWS) were not useful. Participants were administered two 3-minute writing probes to 179 students in grades 3 and 4 in a suburban school in the Southeast. Scores were correlated with teachers' rankings of students' writing ability and the language score on the Iowa Tests of Basic Skills (third graders), or the Louisiana Educational Assessment Program writing subtests (fourth graders). The correlations for the third grade students with the Iowa Tests of Basic Skills ranged from .15 to .43. For fourth graders who took the Louisiana Educational Assessment Program, the correlations ranged from .08 to .36. The CWS method consistently had the highest correlations with the other measures of writing.

A study conducted by Jewell and Malecki (2005) examined the production-dependent (i.e., TWW, WSC, and CWS), production-independent (%CWS and Percentage of Words Spelled Correctly), and accurate-production (CIWS) scoring

methods as compared to the language test on the Stanford Achievement Test. Participants were 203 second, fourth and sixth grade students from three schools in one rural northern Illinois school district. One 3-minute writing probe was provided for participants to complete. Results indicated that the highest correlations with the SAT language test were the %WSC (r = .46 to .50), CIWS (r = .41 to .62) and %CWS (r = .52 to .67) scoring methods across all grades. Moreover, across grades TWW was not significantly related to the SAT language test (r = -.14 to .24). CWS was significantly correlated with the SAT language test for second and fourth grade (r = .46 to .57) but not significantly correlated at grade 6. WSC was significantly correlated with the SAT language test only for second grade.

Alternate forms reliability. In the Espin et al. (2000) study described earlier, with 112 seventh and eighth grade students, alternate forms reliability was also explored. Results for their 3-minute CBM writing probes revealed reliability coefficients of .73 for TWW, .72 for WSC, .76 for CWS, and .74 for CIWS. Alternate forms reliability coefficients are considered to be at a sufficient level at or above .70, based on previous studies (i.e., McMaster, Du, & Pétursdóttir, 2009; McMaster et al., 2011).

Alternate forms reliability was also assessed in the Gansle et al. (2002) study with 83 students in grade 3 and 96 students in grade 4 described earlier. Both grade levels were combined for the results they provided. A correlation coefficient of .62 was obtained for TWW, .53 for WSC, and .46 for CWS.

Weissenburger and Espin (2005) evaluated the alternate forms reliability of CBM-Written Expression probes with 484 students in grades 4, 8, and 10 from three school districts in west central Wisconsin. Two probes were given to the students within a 2-

week period, and scored for TWW, CWS, and CIWS. The correlations ranged from .55 to .80 for TWW, .59 to .79 for CWS, and .61 to .73 for CIWS. All alternate forms correlation coefficients were significant at the $p < .001$ level for all three types of scoring methods. The highest correlations were always at the lowest grade level (i.e., fourth grade) and the lowest correlations were always at the highest grade level (i.e., 10^{th} grade).

McMaster and Campbell (2008) administered several types of measures to assess writing in their study. For comparison purposes, only the results for the standard CBM-Written Expression administration method (i.e., use of a story starter and 3-minute writing sample) will be reported. Participants included students in grades 3 ($n = 25$), 5 ($n = 43$), and 7 ($n = 55$). The alternate form reliabilities for TWW ranged from .60 to .73, for WSC the range was .54 to .78, for CWS the range was .58 to .86, and for CIWS the range was .67 to .86. For all scoring methods, the highest correlations were at the third grade level.

McMaster et al. (2009) reported on the results of two studies, both consisting of 50 first grade students. Multiple types of written expression assessments were administered but only the standard CBM-Written Expression results are reported here. The TWW alternate forms correlations for the two studies were .56 and .66. For WSC the correlations were .47 and .63, for CWS the correlations were both .58, and for CIWS the correlations were .50 and .67.

McMaster et al. (2011) studied the technical features of slopes produced from CBM-Written Expression probes 84 first grade students over a 12-week period. For the standard CBM administration with a story starter, the alternate forms reliability was .61

for TWW, .64 for WSC, and .64 for CWS. The authors note the similarity of results with the previous study with first grade students (McMaster et al. 2009).

Purpose of Present Research

RTI procedures are currently popular in schools. RTI frequently relies on CBM measures for progress monitoring purposes. Thus, it is important to know the alternate forms reliability of CBM-Written Expression probes because results from progress monitoring are used in the determination of whether an instructional intervention is successful. Six studies were located and reviewed that examined alternate forms reliability of CBM-Written Expression. When considering the five most popular scoring methods, all six examined TWW and CWS and five studies examined WSC. Four of the six studies included CIWS. None of the studies evaluated any production-independent measures (e.g., %CWS).

In terms of grade levels, two of the six studies (i.e., McMaster et al. 2009; McMaster et al., 2011) only included first grade students. Two studies (i.e., Espin et al., 2000; Gansle et al., 2002) had participants from multiple grade levels, but reported results based on the grade levels combined. This fact is important to note because the other two studies (i.e., McMaster & Campbell, 2008; Weissenburger & Espin, 2005) found younger students had higher alternate forms correlations than older students.

Thus, additional research is needed to evaluate differences in alternate forms correlations across grade levels and particularly with a production-independent scoring method. Specifically, this study explores potential differences in the production dependent (i.e., Total Words Written, Words Spelled Correctly, Correct Word Sequence), production-independent (i.e., Percentage of Correct Word Sequence), and accurate-

production (i.e., Correct Minus Incorrect Word Sequence) scoring methods across multiple grade levels (i.e., Grades 2, 4, 6, and 8). Given previous research noted gender differences on CBM-Written Expression scores, gender differences in alternate forms correlations will also be explored in an informal manner. This research addresses the following questions:

Research question 1: Is the alternate form reliability of CBM-Written Expression at a sufficient level? The previously used criteria for a sufficient reliability coefficient of at least .70 (McMaster et al., 2009; McMaster et al., 2011) will be used as the criteria in this study. It is hypothesized that the alternate forms reliability coefficients will be sufficient at all grade levels.

Research question 2: What method or methods of CBM-Written Expression scoring show the highest and lowest correlations at specific grade levels? This question seeks to determine if the correlation coefficients for each of the scoring methods show a pattern across grade levels (e.g., higher at lower grade levels and lower as grade levels increase). Given the previous research literature, it is hypothesized that TWW and WSC will show higher correlations at the younger grade levels and the CWS, %CWS, and CIWS measures will show higher correlations at the upper grade levels.

Research question 3: Are there gender differences in the correlations? Given previously reported gender differences in mean scores on CBM-Written Expression measures, this research question explores potential gender differences affects alternate forms correlations as well. It is hypothesized that the correlations will not differ.

Method

Participants

The participants in this study were obtained from one elementary school (grades 1-5) and one middle school (grades 6-8) within a single district in central Kentucky. Students from grades 2, 4, 6, and 8 were included in this study to examine potential differences across grade levels. Participants included 225 students (52% females and 48% males). Overall, 12% of the participants received special education services. The number of students receiving special education services was fairly equal for grades 2 (14%), 4 (16%), and 8 (13%). Only 4% of the students in grade 6 received special education services. Ethnicity of participants encompassed Caucasian (53%), African American (26%), Hispanic/Latino (12%), two or more races (7%), Asian (1%), and Native American (1%).

Instrument

Eight CBM-Written Expression story starters (4 grade levels x 2 administrations) were randomly selected from story starters listed in Hosp et al. (2007). The story starters can be found in Appendix A. Hosp et al. (2007) provide multiple CBM-Written Expression story starters for primary, intermediate, and advanced levels. For this study, primary level story starters were used for grade 2, intermediate story starters were selected for grades 4 and 6, and advanced story starters for grade 8.

Procedure

The elementary school selected for this study was chosen for convenience reasons. All four second-grade and all four fourth-grade classrooms in the building were included in this study. In the middle school, also chosen for convenience reasons, three sixth-grade and three eighth-grade homeroom teachers were randomly selected from the 15 (total)

homerooms. Opt-out consent forms were sent home to the students' parents and guardians requesting permission for the students to participate in this study. The students whose parents indicated that they did not want their child to participate were asked to partake in another activity during the administrations of the CBM probes. Permission to conduct this study was granted by the school district's Superintendent and Western Kentucky University's (WKU) Institutional Review Board (see Appendix B).

One school psychologist and one school psychology intern, trained by a WKU psychology professor who had previously received extensive training on CBM, administered the CBM-Written Expression probes in the spring of the 2013-2014 school year. The second set of probes was administered four to six days later. During the administrations of the CBM probes, all students participating in the study first completed an assent form indicating their own agreement to participate. After the assent forms were collected, lined sheets of paper with the story starter typed at the top were passed out facedown. The students were instructed to write their name and homeroom on the back. Then the administrators read the standardized directions to the subjects. After the directions were read, the students were instructed to flip the paper over and begin writing. The examiners gave the following standardized instructions from Hosp et al. (2007) before students began writing:

- Say: "Today I want you to write a story. I am going to read a sentence to you first and then I want you to compose a short story about what happens. You will have 1 minute to think about what you will write and 3 minutes to write your story. Remember to do your best work. If you do not know how to spell a word,

you should guess. Are there any questions?" (Pause) "Put your pencils down and listen. For the next minute, think about … [insert story starter]."

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

For the first administration of the writing probes, the participants replied to the one probe chosen for their grade level. For the second administration, participants were provided a different writing probe with the same level of difficulty per grade level. After the administration of both probes, student's probes were matched together based on their names by homeroom and grade level.

After the two administrations of the writing probes, they were collected and scored by the school psychology. The writing probes were scored using the production-dependent measures of Total Words Written (TWW), Words Spelled Correctly (WSC), and Correct Word Sequence (CWS), the production-independent measure of Percentage of Correct Word Sequence (%CWS), and the accurate-production measure of Correct Minus Incorrect Word Sequences (CIWS).

After scoring the writing probes, 20% of the probes from each grade level were re-scored by a certified school psychologist familiar with CBM-Written Expression to

evaluate inter-rater agreement. As recommended by Sattler (2002) a minimum inter-rater agreement of 80% was used. The inter-rater agreement was calculated by adding the number of probes that fell within the agreed upon standard error of measure for each scoring method, and then dividing that number by the total number of probes re-scored for that grade level. Inter-rater agreement across the grade levels ranged from 91% to 100% for the three methods of TWW, WSC, and CWS, and ranged from 81% to 94% for both %CWS and CIWS. When differences in scores occurred, the scoring differences were discussed between raters to reach an agreement on the correct score. All the inter-scorer agreements were above the minimum acceptable levels and suggest the scores are accurate.

Results

The first research question addresses whether the CBM-Written Expression alternate form reliability coefficients are at a sufficient level using the minimum criterion of .70. Pearson correlations were calculated between the scores of each scoring method from the two administrations for each grade and the total sample and are presented in Table 1. McMaster and Espin's (2007) descriptive terminology will also be used in interpreting the results: strong ≥ .80, moderately strong .70 to .79, moderate .60 to .69, and weak < .60. All correlations were statistically significant at $p < .001$. Eleven of the 25 correlations met the "sufficient" criteria of being at least .70. Of the correlations that were considered sufficient, none was at the strong level. All 11 considered sufficient were at the moderately strong level. Six of the correlations were at the moderate level and eight were at the weak level. Thus, the hypothesis that all the correlations would be at a sufficient level was only partially supported. Only 44% of the correlations were considered sufficient.

For descriptive purposes and as an indirect, secondary method of evaluating the results of the CBM-Written Expression alternate forms, the means and standard deviations of each assessment were calculated at each grade level for each scoring method. Those results are presented in Table 2. A repeated measures ANOVA was completed for each grade level to determine if any significant differences occurred in the scores between the administrations. A significant difference occurred only in the eighth grade, $F(1, 53) = 15.25$, $p < .001$, $h2 = .22$. Post-hoc analyses indicated all scoring methods, except %CWS, in grade 8 had significantly different mean scores between

Table 1

Alternate Forms Correlation Coefficients by Grade Level and Scoring Method

	TWW	WSC	CWS	%CWS	CIWS
Grade 2 (n = 57)	.72	.70	.64	.63	.51
Grade 4 (n = 62)	.72	.73	.76	.50	.72
Grade 6 (n = 52)	.59	.65	.62	.79	.60
Grade 8 (n = 54)	.46	.46	.53	.50	.54
Total Sample (n = 225)	.68	.70	.73	.71	.73

Note. All correlations were significant at $p < .001$.

Table 2

Means and Standard Deviations by Grade Level and Scoring Method

	Time 1		Time 2	
	Mean	(*SD*)	Mean	(*SD*)
Grade 2 (*n* = 57)				
TWW	18.5	(9.9)	20.1	(11.0)
WSC	16.0	(9.1)	17.6	(10.6)
CWS	12.8	(9.1)	14.0	(10.1)
%CWS	62.6	(25.4)	63.9	(24.6)
CIWS	7.0	(9.6)	7.3	(12.1)
Grade 4 (*n* = 62)				
TWW	36.7	(15.0)	40.8	(14.6)
WSC	34.7	(14.6)	37.9	(13.8)
CWS	31.7	(14.7)	34.2	(14.4)
%CWS	79.8	(17.1)	78.8	(15.1)
CIWS	25.0	(14.9)	25.6	(17.0)
Grade 6 (*n* = 52)				
TWW	37.0	(14.3)	37.8	(19.8)
WSC	35.2	(14.0)	35.6	(17.9)
CWS	33.8	(13.7)	34.6	(17.8)
%CWS	82.6	(13.8)	83.4	(17.8)
CIWS	27.8	(13.6)	29.5	(17.6)
Grade 8 (*n* = 54)				
TWW	38.0	(15.8)	50.0	(22.3)
WSC	36.9	(15.8)	48.4	(22.1)
CWS	37.6	(16.6)	48.3	(23.3)
%CWS	90.8	(9.5)	90.3	(8.2)
CIWS	34.2	(16.8)	43.6	(22.9)

Note. There were no significant differences in grades 2, 4, or 6.

assessments. Thus, these results seem to suggest that all scoring methods consistently measure students' writing skills, except above the sixth grade.

The second research question evaluates whether any particular pattern of results occurred, either across scoring methods or grade level. A visual analysis of the correlations in Table 1 was used to examine the existence of any patterns of results. Many of the correlations varied greatly among the methods and grade levels, making it difficult to see any clear-cut patterns of results. However, it appears the TWW and WSC scoring methods had correlations higher at grades 2 and 4 than at grades 6 and 8. At grades 2 and 4, those four correlations were at a moderately strong level while three of those four correlations at grades 6 and 8 were at a weak level, with the fourth one at a moderate level. The only other noticeable pattern is that the correlations were lower in grade 8 than the other grade levels. The correlations for all scoring methods in grade 8 were at a weak level.

Using a Fisher *r*-to-*z* transformation, the significance of the difference between all pairs of correlation coefficients was calculated. The results indicated that correlations at grades 2 and 4 for TWW were significantly higher than grade 8. For WSC and CWS, the grade 4 correlations were significantly higher at grade 8. For %CWS, the grade 6 correlation was significantly higher than at grade 8. No significant differences between grade levels were found for CIWS.

The third research question sought to explore the possibility of gender differences in the alternate forms reliability coefficients. The correlation coefficients and means were determined separately for boys and girls by grade level and scoring method. The

correlation coefficients are presented in Table 3. A visual analysis of the correlations in Table 3 suggests boys and girls had the most similar correlations at grade 4. In grades 6 and 8, however, the girls frequently had lower correlations than the boys. In those two grades, boys had higher correlations than the girls in nine of the 10 comparisons. The correlations for the girls were often much lower as well, particularly in the sixth grade. To illustrate the difference, the means of the correlations were determined for grades 6 and 8 separately for the boys and girls. In grade 6, the boys' mean correlation coefficient was .71 while for the girls it was .40. In grade 8, the correlation was .53 for the boys and .38 for the girls. In the eighth grade, the correlations for the boys and girls were similar for the production-dependent measures of TWW, WSC, and CWS. The gender differences occurred primarily with the other two measures (i.e., %CWS and CIWS). In fact, boys had higher correlations on %CWS and CIWS in eight of the 10 comparisons. Using a Fisher *r*-to-*z* transformation, the significance of the difference between the boys' and girls' correlation coefficients was calculated. Only two statistically significant differences occurred and those were at the sixth grade level for TWW and CWS (both $p = .04$). When all grade levels are combined, the total correlations were remarkably similar between boys and girls and not statistically significantly different.

For descriptive purposes, the means for the girls and boys are presented in Table 4. The girls' mean scores were higher than the boys' mean scores in 36 of the 40 comparisons. Girls' mean scores were always higher on the production-dependent measures. For three of the four times boys, on average, scored higher than girls, it was on the production-independent measure of %CWS.

Table 3

Alternate Forms Correlations for Girls and Boys by Grade Level and Scoring Method

	TWW		WSC		CWS		%CWS		CIWS	
Grade	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
2	.75	.55	.73	.56	.63	.73	.56	.77	.45	.75
4	.78	.66	.78	.69	.82	.68	.51	.60	.73	.69
6	.23	.69	.37	.71	.27	.70	.83	.77	.28	.69
8	.41	.42	.40	.45	.46	.54	.24	.61	.39	.62
Total	.70	.64	.73	.66	.75	.71	.69	.76	.72	.75

Table 4

Mean Scores by Grade Level, Gender, and Scoring Method

	Time 1 Means		Time 2 Means	
	Girls	Boys	Girls	Boys
Grade 2 (36 girls, 21 boys)				
TWW	19.9	16.1	21.6	17.5
WSC	16.9	14.4	19.0	15.2
CWS	13.3	12.1	15.2	12.1
%CWS	60.0	67.1	64.3	63.3
CIWS	6.6	7.6	8.6	5.1
Grade 4 (34 girls, 28 boys)				
TWW	39.0	33.9	43.6	37.5
WSC	36.9	31.9	40.9	34.3
CWS	34.1	28.8	37.8	29.8
%CWS	79.1	80.7	81.5	75.5
CIWS	27.2	22.4	29.6	20.8
Grade 6 (25 girls, 27 boys)				
TWW	41.4	32.9	46.6	29.7
WSC	39.6	31.1	43.1	28.7
CWS	39.4	28.7	42.8	27.0
%CWS	85.0	80.3	85.6	81.3
CIWS	33.8	22.2	37.1	22.5
Grade 8 (23 girls, 31 boys)				
TWW	41.7	35.2	58.9	43.4
WSC	40.5	34.2	57.7	41.6
CWS	41.2	35.0	58.3	40.9
%CWS	90.7	90.9	92.3	88.8
CIWS	37.7	31.5	54.2	35.6

Discussion

CBM is an assessment measure that can be utilized in different ways (e.g., identification of at-risk students, assist with special education eligibility decisions, monitor student progress in academic areas, etc.). The purpose of this study was to add to the literature on the technical adequacy of CBM-Written Expression by assessing its alternate forms reliability. This study investigated the consistency of student performance on alternative writing probes on production-dependent (i.e., TWW, WSC, and CWS), production-independent (i.e., %CWS), and accurate-production (i.e., CIWS) CBM-Written Expression scoring methods across multiple grade levels (i.e., grades 2, 4, 6, and 8).

The first research question sought to determine whether the CBM-Written Expression alternate form reliability coefficients are at a sufficient level. It was hypothesized that alternate forms reliability coefficients would be sufficient at all grade levels. From the comparisons, less than half of the correlations are considered "sufficient" at r = .70 or above. Such results suggest school personnel should be cautious about using a single CBM-Written Expression probe to determine a student's writing skills, as the different scoring methods did not consistently measure student's writing from one probe to another across all grade levels. Thus, one writing probe does not depict a student's true writing ability. An additional probe or probes might be necessary if a student's CBM score is not supported by other sources of information about his or her writing skills.

On the other hand, the additional analysis of the means revealed no statistically significant differences between the alternate forms for grades 2, 4, and 6. It is also important to point out that the correlations for the total sample were mostly at a sufficient

level. (TWW was close to being considered sufficient with a correlation of .68.) Those analyses suggest consistent results can be obtained from alternate forms of CBM-Written Expression probes, at least below eighth grade.

The second research question looked for patterns of high and low correlations across scoring methods and grade level. Based on previous research (McMaster & Espin, 2007; Weissenburger & Espin, 2005), it was hypothesized that correlations for TWW and WSC would be higher at lower grade levels. This hypothesis was confirmed. Correlations for those measures in grades 2 and 4 were all at a sufficient level (i.e., > .70) while none of those correlations in grades 6 and 8 were at a sufficient level. In fact, three of the four correlations at those grade levels were considered weak.

The pattern of correlations obtained in this study, at least for TWW and WSC, were similar to those found by Weissenburger and Espin (2005) and McMaster and Campbell (2008), who reported higher correlations at younger grade levels. The second part of the hypothesis, that the other scoring methods would have higher correlations in the upper grade levels, was not confirmed. Those correlations varied greatly across the grade levels, with no clear pattern of results. The exception appears to be with the eighth grade students, who tended to have lower correlations for all scoring methods. Support for weaker correlations at the eighth grade level also comes from the results of the repeated measures ANOVA. For four of the five scoring methods, the mean scores were statistically different between probe administrations in grade 8. One possibility of the differences among the eighth grade students, compared to the second, fourth, and sixth grade students, could be an issue with the writing probes provided (e.g., the second probe was easier to write about or it had greater interest for the students). Another possible

explanation could be related to carry-over effects with test-retest methods. Students at all grade levels had mean scores on their second administration higher than their first administration, with the exception of %CWS in grades 4 and 8. Eighth grade students may have scored significantly higher because they benefitted more from practice effects being at a higher developmental level than the younger students.

Although research indicates girls outperform boys on writing tasks, no gender differences were found for the combined grade levels. When examining gender differences at individual grade levels, fourth grade yielded the most similar correlations between boys and girls while sixth grade yielded the most disparate correlations. It is difficult to explain such results. Most likely, the results are variable due to the small sample size of boys and girls at each grade level (range: 21 to 36). Thus, the results of the combined sample may be the most accurate. Even if girls outscore boys, there is no reason to expect one group to be less consistent with their skills.

Strengths and Limitations

The current study has multiple strengths. It evaluated a fairly large sample of participants across four grade levels. Other studies that evaluated CBM-Written Expression alternate forms reliability only used students from one to three grade levels. The current study included second and sixth grade students. None of the other studies included students from those grade levels. The overall sample size of 225 students is only surpassed by one other study (Weissenburger & Espin, 2005). Another strength was that integrity checks were used during the administrations of the CBM writing probes. During the administrations, a trained person in standardized procedures of administering CBM-Written Expression observed the author of this study and verified that the administrations

had consistent timing, and instructions given to the participants were verbatim to the instructions provided in the Hosp et al. (2007).

A limitation of the present study is the issue of generalizability. The participants were from two schools in a school district in central Kentucky. Although the district is the fourth largest district in Kentucky, and the sample was diverse, the results will not necessarily generalize to other school districts across the nation. The order of the administration of probes was not counterbalanced, so it is not clear if any differences were due to the probes themselves. Furthermore, the sample may not have been large enough to evaluate gender differences. Another limitation to this study is although scoring guidelines provided by Hosp et al. (2007) were used to score the writing probes, several instances arose that were not specifically included in the scoring guidelines. For example, it is not specified how a word should be scored when it has an incorrectly added apostrophe (e.g., "a lot of tree's"). Additionally, no directions are given for what to do when it cannot be determined if the student wrote an upper or lowercase letter due to the student's handwriting.

Future Research

Some avenues of future research could include student's performance on writing probes that are deemed more interesting to them. In this study, anecdotal evidence suggests some participants wrote more when they demonstrated an interest in the provided prompt. How does the interest level in the topic affect how much a student writes? Moreover, research on writing probes in general should be conducted by alternating the order of the administration of probes among students. For instance, on the first administration, half of the participants would receive probe one, while the other half

would receive probe two; and then the probes would be reversed for the second administration. This should rule out the writing probe as a factor for differing scores. Furthermore, more specific and detailed scoring procedures should be developed to address all possible writing characteristics and scoring errors.

Conclusion

The current study is an important addition to the current literature on alternate forms reliability for CBM-Written Expression. It is important for school personnel to realize that some of the scoring methods elicit more reliable results when scoring writing probes and should consider this when analyzing the student's performance. Appropriate scoring methods are necessary for school personal to accurately assess students' progress so appropriate writing goals for intervention purposes can be developed.

References

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

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to Intervention: Principles and strategies for effective practice* (2nd ed.). New York, NY: The Guilford Press.

Cocker, J. L., & Ritchey, K. D. (2010). Curriculum-based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional Children*, *76*, 175-193

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

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

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

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

Fuchs, L. S., & Fuchs, D. (n.d.). Using CBM for progress monitoring in written expression and spelling. *Introduction to Curriculum-Based Measurement,* Retrieved from http://www.studentprogress.org/summer_institute/2007/written/writing_manual_2007.pdf

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199-208.

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

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.

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

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

Kentucky Department of Education (2014). *Kentucky Performance Rating for Educational Progress*. Retrieved from http://applications.education.ky.gov/SRC/AssessmentByState.aspx

Klotz, M. B., & Canter, A. (2006). Response to intervention (RTI): A primer for parents. *NASP Resources.* Retrieved from http://www.nasponline.org/resources/factsheets/rtiprimer.aspx

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

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

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

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

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

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. (1982). Frequency of measurement data and data utilization strategies factors in standardized behavioral assessment of academic skill. *Journal of Behavior Assessment, 4,* 361-370.

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

National Commission on Writing. (2011). *National Assessment of Educational Progress NAEP 2011 Writing Assessment.* Retrieved from http://nces.ed.gov/whatsnew/commissioner/remarks2012/09_14_2012.asp

Riley-Tillman, T. C., Burns, M. K., & Gibbons, K. (2013). *RTI applications: Assessment, analysis, and decision-making*. New York, NY: Guilford Press.

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

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

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

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

Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education.* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Weissenburger, J., & Espin, C. A. (2005). Curriculum-based measures of writing across grade levels. *Journal Of School Psychology*, *43*(2), 153-169. doi:10.1016/j.jsp.2005.03.002

APPENDIX A

Story Starters

Second Grade

1st administration: My favorite game to play at recess is…

2nd administration: The best part of school is…

Fourth Grade

1st administration: When the alarm sounded I…

2nd administration: My day was going bad until…

Sixth Grade

1st administration: Instead of going to bed last night I decided to…

2nd administration: I can't believe I had been voted class president! My first item of business was…

Eighth Grade

1st administration: The teenagers were hiking through the forest when they came across an old rundown cabin that was...

2nd administration: The clerk at the store was annoyed because…

APPENDIX B

Institutional Review Board Approval

INSTITUTIONAL REVIEW BOARD
OFFICE OF RESEARCH INTEGRITY

DATE:	March 31, 2014
TO:	Carl Myers, Ph.D.
FROM:	Western Kentucky University (WKU) IRB
PROJECT TITLE:	[580126-3] Investigating the technical adequacy of Curriculum Based Measurement-Written Expression
REFERENCE #:	IRB 14-312
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	March 31, 2014
EXPIRATION DATE:	July 25, 2014
REVIEW TYPE:	Full Committee Review

Thank you for your submission of Amendment/Modification materials for this project. The Western Kentucky University (WKU) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Full Committee Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a *signed* consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of July 25, 2014.

Please note that all research records must be retained for a minimum of three years after the completion of the project.

If you have any questions, please contact Paul Mooney at (270) 745-2129 or irb@wku.edu. Please include your project title and reference number in all correspondence with this committee.