6-2009

Curriculum-Based Measures in Writing: A School-Based Evaluation of Predictive Validity

Christina M. Terenzi
Western Michigan University

CURRICULUM-BASED MEASURES IN WRITING: A SCHOOL-BASED EVALUATION OF PREDICTIVE VALIDITY

by

Christina M. Terenzi

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology
Advisor: Galen Alessi, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2009

CURRICULUM-BASED MEASURES IN WRITING: A SCHOOL-BASED EVALUATION OF PREDICTIVE VALIDITY

Christina M. Terenzi, Ph.D.

Western Michigan University, 2009

Recent research in the area of Curriculum-Based Measures (CBM) in writing has shown that traditionally used metrics, such as total words written and total words correct, may not be the best tools for measuring writing performance, for both secondary and elementary aged children (e.g., Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Tindal & Parker, 1989a; Watkinson & Lee, 1992). Evidence suggests that more advanced measures, such as production-independent measures (e.g., percentage of correct word sequences) may be stronger predictors of student skill level in the area of writing. The present study replicated portions of a recent seminal study and investigated the predictive validity of CBM in the area of writing for the Michigan Educational Assessment Program (MEAP) writing and ELA assessments and the Iowa Test of Basic Skills (ITBS) reading assessment. Participants included 700 fourth grade students in a Midwest urban school district who completed a three-minute writing probe, which was scored for 20 independent variables. Dependent variables included assessments administered in the same year and in years following the administration of the writing probes. Correlations were calculated between each of the independent and dependent variables. Interscorer reliability was calculated, with all variables above .80. Alternate form reliability (n=199) was above .40 for all but two independent variables. Stepwise multiple regressions were run with two sets of independent variables with each of five dependent variables. The independent variables which appeared to be the most promising indices for predicting

performance on dependent measures included percentage of correct word sequences, correct punctuation marks, and words in complete sentences. Implication of analyses, limitations, and implications for future practice and research are discussed.

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

I would like to start by thanking two of the most influential people in my life: my parents. Their continued support and encouragement through the many years of my studies helped me to persevere, despite the many obstacles I encountered.

Secondly, I would like to thank my mentor, Dr. Ruth Ervin, whose enthusiasm and forward thinking have been inspirational to me, in both my studies and practice. I would also like to sincerely thank the members of my dissertation committee: Dr. Galen Alessi, Dr. Mary Anderson, Dr. John Austin, and Dr. Margaret McGlinchey, for taking the time to review and evaluate my work. I would like to express my utmost gratitude to Dr. Anderson for her continued support and coaching throughout this study. I am so appreciative of her expertise and guidance, as well as her enthusiasm for this project.

Finally, I would like to thank my friends who dedicated their time and positive energy to help me with scoring the writing probes, specifically Nancy Burr, Janet Green, Carrie Koptizki, and Autumn Terenzi.

Christina M. Terenzi

TABLE OF CONTENTS

Table of Contents—continued

LIST OF TABLES

List of Tables—Continued

CHAPTER I

INTRODUCTION

Purpose

Written expression requires a complex set of interwoven skills, including spelling, handwriting, planning and organizing material, and transcribing thoughts into words. As students progress from elementary through secondary grades, the focus of their education will shift from learning basic language arts skills to applying these skills, requiring them to draw on their knowledge in the area of written expression, and integrate it with other basic language arts skills (Tindal & Parker, 1989b). Therefore, in order for students to have later academic success, the development of these fundamental skills (e.g., reading, writing) in early grades is essential. A student's progress in the area of written expression is dependent upon frequent corrective feedback and practice, as well as the need for formative evaluation to guide instruction in writing. Although a number of written expression assessments exist, many are subjective in nature, difficult to administer and score, or require large amounts of teacher or student time (Watkinson & Lee, 1992). Educators need reliable and valid measures that can be used to provide frequent feedback to students and to guide instruction. Frequent monitoring of student progress is critical to students' success (Deno, 1992), and curriculum based measures are technically adequate measures that are reliable, time-efficient, easy to administer and provide objective indicators of student progress (Deno, 1985).

The purpose of this project was to assist a large Midwestern urban school district in evaluating the utility of Curriculum-Based Measures (CBM) in the area of writing and to add to the literature on the utility of CBM writing measures. In the following section,

CBM and its uses, including a historical background, are described. Next, an extensive review of the available research on CBM in the area of writing is provided. Finally, a practical and empirical rationale for the proposed project is presented.

Curriculum-Based Measures

Curriculum-Based Measures (CBM) are "a set of standard simple, short-duration fluency measures of reading, spelling, written expression and mathematics computation" (Shinn & Bamonto, 1998, p. 1). These measures were initially designed to monitor students' progress in skills that are critical in achieving academic success (i.e., reading, math, spelling, and written expression) (Deno, 1985; Marston, 1989) and to allow teachers to formatively evaluate their instruction to make improvements (Deno, 2003a, 2003b). In reading, for example, students are asked to read out loud for one minute. Oral reading fluency, the number of words read correctly per minute, is the metric that is then used to guide instructional decision-making. Oral reading fluency has been shown to predict performance on later high-stakes testing (McGlinchey & Hixson, 2004), as well as reading comprehension and later reading fluency (Good, Gruba, & Kaminski, 2002). In mathematics, CBM involves having students complete math probes for 3- to 5-minutes. The decision making metric is the number of digits correct per minute. Similarly, in the area of spelling, students write words that are dictated every 7 seconds for 2 minutes. The number of correct letter sequences written is determined, and is the metric used in decision making. Finally, in the area of written expression, students are given a story starter, instructed to think and plan their story for 1 minute, and then to write for 3 minutes. The number of words written, number of words spelled correctly,

and number of correct word sequences are typical metrics used for educational decision-making.

Historical Overview of Curriculum-Based Measures

CBM was developed during the late 1970s and early 1980s by Deno, Mirkin, and colleagues at the University of Minnesota's Institute for Research on Learning Disabilities (IRLD) (Marston, 1989). Deno sought to provide his students, future special education teachers, measures that could be collected daily, graphed, and used to evaluate student progress (Shinn & Bamonto, 1998). Since the development of CBM, extensive research has focused on establishing its technical adequacy and practical utility as a formative and summative evaluation tool for individual students as well as larger groups of students (Martston; Shinn & Bamonto). One common use for CBM is to monitor the academic progress of individual students over time in order to evaluate the effectiveness of implemented interventions. CBM improves the data-base for making educational decisions because measures of student achievement can be collected frequently (Shinn, 2002). For example, a teacher can monitor a student's performance on a regular basis and make instructional decisions about whether student progress is satisfactory, or if the instructional plan should be changed. In addition to being a useful tool for individual student progress monitoring, CBM has been employed with large groups of students as a formative tool to screen for students who are at-risk for having difficulties in various academic areas (Deno, 2003a). For example, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; http://www.dibels.uoregon.edu) are research-based CBM measures of pre-reading and reading skills (Good, Gruba, & Kaminski, 2002) that are administered on a schoolwide level three times per year to screen for reading difficulties.

Based on scores on the DIBELS measures, students are identified as meeting *benchmark* goals (i.e., research based goal levels), or being at risk for reading difficulties and in need of either *strategic* or *intensive* intervention. In other words, when CBM tools, such as the DIBELS, are used to screen for risk of learning problems, they assist in identifying which students need additional intervention in order to achieve expected learning outcomes in basic skill areas.

CBM was developed based on a number of salient criteria that are considered important when designing progress monitoring data collection procedures, including: (a) being tied to the students' curricula, (b) of short duration to make frequent administration possible, (c) capable of having many forms, (d) inexpensive to produce, and (e) sensitive to small improvements in student performance over time (Marston, 1989). Frequently, evaluation of student learning occurs at the conclusion of the instructional period (i.e., summative evaluation), when it is often too late to change teaching methods (Shinn & Bamonto, 1998). In contrast, by using CBM as a *formative* evaluation tool, performance can be assessed continuously, and decisions regarding student progress and proper adjustments to curriculum and/or instruction can be made more frequently. In addition, research in the area of CBM has been expanded to include not only monitoring of student progress and formative evaluation, but also screening and identification of at-risk students (Good & Kaminski, 1996), predicting performance on high-stakes assessments (McGlinchey & Hixson, 2004), and developing schoolwide accountability systems (Deno, 2003b).

Assessment using curriculum-based measures in reading, writing, and mathematics has been determined to be reliable and valid in several research studies. An

initial study by Deno, Mirkin, and Chiang (1982) correlated reading CBM with criterion-based norm-referenced tests, and results indicated that listening to students read out loud for one minute was a valid measure of reading skill. Most correlation coefficients in that initial study were above .80, with a range of .73 to .91. Through separate studies, reliability estimates were determined using test-retest, parallel forms, and interrater methods. Results indicated that most coefficients were above .90, and provide compelling evidence of the reliability of CBM in the area of reading fluency (Marston, 1989). In the areas of math, spelling, and written expression, results of initial research studies indicate that CBM are valid measures of student skill level, and are highly correlated (usually above .80) with criterion measures. In addition, studies indicate high reliability with CBM in these areas, using test-retest, parallel forms, and interrater agreement (Marston). Although less attention has been paid to the validity of CBM in the area of written expression, results of studies indicate that reliability and validity are at a high enough level for it to be useful (Marston; Shinn, Ysseldyke, Deno, & Tindal, 1986) (See Marston for a complete review of initial research of reliability and validity in the areas of reading, math, and spelling). More recently, Hosp and Fuchs (2005) examined the relation between reading CBM and several subtests of the Woodcock Johnson Reading Mastery Test – Revised (Woodcock, 1987). Results indicated strong correlations between words read correctly and all subtests at all grade levels assessed (i.e., grades 1 through 4). Correlation coefficients ranged from .71 to .91 (Also review Shinn, Good, Knutson, Tilly, & Collins, 1992).

Despite published research, which supports the reliability and validity of using CBM to monitor students' progress and to make educational decisions, many educators

do not accept CBM as an adequate measure (Fewster & MacMillan, 2002; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Shinn et al., 1992). This has been termed a "face validity" issue by Shinn and colleagues. In other words, teachers may view the measures as being too simplistic to be effective at measuring constructs as complex as reading or writing. This may be a bigger concern in the area of written expression, because, as indicated by Tindal and Marston (1990), writing as a construct is more difficult to measure. In reading, for example, there is a single, identifiable measure that is directly related to the students' reading fluency (i.e., number of words read correctly) (Marston, 1989). In the area of math, it is also possible to obtain a single, discrete measure (e.g., number of digits correct per minute) of a targeted skill. In the area of writing, however, there are an infinite number of possible responses to writing demands and an equally large number of scoring rubrics (Gansle et al.). Perhaps due to this complexity, reports from educators indicate that typical CBM writing metrics (e.g., total words written) appear too simplistic (Gansle et al.). Tindal and Parker (1989a) indicate that direct assessments of writing are thought to have stronger content validity than more indirect methods, such as published tests which use objective multiple-choice questions or sentence-combining formats. Research has shown that CBM in the area of written expression is indeed a useful tool; however, a relatively small number of studies have been conducted in the area of CBM in written expression. These research studies are described in the following section.

Validity of Curriculum-Based Measures in the Area of Written Expression

Initial research studies on curriculum-based measures in writing were conducted with elementary aged students in grades 3 to 6 (Deno, Mirkin, and Marston, 1980). These

studies focused on the validity of six basic measures, including total words written (TWW), total words spelled correctly (TWC), correct letter sequences (CLS), number of mature words (i.e., words not found on Finn's (1977) undistinguished word list), number of large words, and Hunt's (1966) average t-length, which describes grammatical maturity. These measures were correlated with norm-referenced standardized measures, including the *Test of Written Language* (TOWL: Hammill & Larsen, 1978), the Developmental Sentence Scoring System (Lee & Canter, 1971), and the Language subtest of the Stanford Achievement Test (SAT: Madden, Gardner, Rudman, Karlsen, & Merwin, 1978). The results of the initial studies indicated that TWW, TWC, CLS, and mature words were all highly correlated with the criterion measures (Marston, 1989). Additionally, Deno and colleagues demonstrated that equivalent results could be obtained when using either 2- or 5-minute written samples, and when using story starters, picture stimuli, or a topic sentence. Deno, Marston, and Mirkin (1982) replicated the results of the initial studies, finding that TWW, TWC, CLS, and mature words were, again, most highly related to the criterion measures. Vindeen, Deno, and Marston (1982) expanded the research by exploring the relation between the curriculum based measures and teacher holistic ratings, as well as to the standardized criterion measures (i.e., TOWL, Developmental Sentence Scoring System). Their results indicated a correlation of .85 between TWW and teacher holistic ratings, and a correlation of .84 between TWC and teacher holistic ratings. The correlations between the TOWL and the curriculum based measures were also strong, with a correlation of .66 for TWW, and .92 for TWC. Correlations between the curriculum based measures and the Developmental Sentence

Scoring System were not as strong, with correlations of .51 for TWW and .52 for TWC (Marston).

More recently, Gansle, VanDerHeyden, Noell, Resetar, and Williams (2006) examined the technical adequacy of previously studied curriculum-based measures in writing. Results supported previous findings, showing moderate correlations between the Total Language score on the Stanford-9 Achievement Test and several traditional curriculum-based measures for elementary school students, including TWW ($r = .34$), TWC ($r = .38$), and CWS ($r = .43$). Additionally, technical adequacy data on less frequently studied variables also were presented, supporting the validity of these measures, including words in complete sentences ($r = .41$), correct punctuation ($r = .39$), correct capitalization ($r = .28$), and complete sentences ($r = .36$). Further studies have been completed which explored the discriminative validity (e.g., Parker, Tindal, & Hasbrouk, 1991; Watkinson & Lee, 1992) in addition to criterion-related validity (e.g., Espin, Scierka, Skare, & Halverson, 1999) of CBM in the area of writing. Several researchers have examined the validity of using curriculum-based measures with secondary school students (Tindal & Parker, 1989a; Watkinson & Lee; Espin et al.). Results have shown that traditional measures, such as TWW and TWC, are not as strong of predictors of writing performance for students in secondary grades as for elementary school students, and production-independent measures, which focus on accuracy as opposed to fluency, are often cited as being better predictors for secondary students (Malecki & Jewell, 2003; Tindal & Parker; Watkinson & Lee). Tindal and Parker demonstrated that percentage of correct word sequences, percentage of correctly spelled words, and mean length of correct word sequences were more highly related to teacher

holistic ratings of student writing performance than their production-dependent counterparts for middle school students. These studies are described in detail in the following sections.

The reliability of CBM in writing was also examined, with positive results. Results of initial studies indicated reliability estimates ranging from .42 to .91 for TWW, .46 to .81 for TWC, and .51 to .92 for CLS, using test-retest reliability (Marston, 1989). Using parallel form reliability, studies indicated reliability estimates for TWW ranging from .42 to .96, with most estimates above .70. For TWC, indicated reliability estimates ranged from .41 to .95, and for CLS, reliability estimates ranged from .49 to .96. For all three measures, interrater scoring was very high (Marston).

Using Curriculum-Based Measures to Differentiate Between Students With and Without Disabilities

In order to examine its discriminative validity, researchers have explored the possibility of using CBM in writing to differentiate between students with disabilities and students without disabilities, including high- and low-achieving general education students, and non-disabled students in remedial programs. Traditional identification procedures (e.g., commercial nationally norm-referenced tests) have not been consistently reliable in distinguishing between different educational groups (Shinn & Marston, 1985); however, studies have shown that curriculum-based measures reliably differentiate between students in different educational groups in the areas of reading (e.g., Shinn & Marston), math (e.g., Shinn & Marston), written expression (e.g., Watkinson & Lee, 1992, Tindal & Parker, 1989a), and spelling (e.g., Shinn & Marston). Shinn and Bamonto (1998) indicate that interest in curriculum-based measures "exploded in the mid-1980s" due to a decrease in the confidence of the test-and-place model using nationally norm-

referenced tests (p.3) as well as an increased interest in alternative assessment practices. This ability to distinguish between groups of students supports the use of curriculum-based measures for the purpose of special education screening and eligibility decision-making (Shinn & Marston; Fewster & MacMillan, 2002); however, initial studies in the area of written expression indicated that typical measures (e.g., TWW, TWC) may not be sensitive enough to differentiate between groups of general education students reliably as reading or spelling measures (Shinn et al., 1986). More recent research, however, has demonstrated that alternate CBM writing measures could differentiate between groups of students reliably (Parker et al., 1991; Watkinson & Lee, 1992).

Two of the first studies to compare the performance of groups of students examined CBM in the areas of reading, spelling, and written expression (Shinn & Marston, 1985; Shinn et al., 1986). Results of both of these studies indicated that the typically used CBM in writing (e.g., TWW, TWC) did not reliably differentiate between groups of students; however, results supported the use of writing CBM to distinguish students with disabilities from non-disabled peers. Shinn and Marston compared the writing performance of students in fourth through sixth grades who were in the general education population, received Chapter 1 services, or who had an educational diagnosis of mildly handicapped. Students completed CBM in the areas of reading, spelling, math, and written expression. In the area of written expression, students were given a story starter and 3 minutes to write, and the passages were scored for TWC. In all grades, the performance of the students with mild handicaps was significantly different from the students in the other two groups in all areas of CBM. For students in grade 4, all three groups were significantly different from each other on the writing measure; however, for

students in grades 5 and 6, Chapter 1 students and general education students did not show any differences in performance from one another. This study is notable because it showed that CBM reliably differentiates between students receiving Chapter 1 services and those students in special education, as traditional identification procedures (i.e., test and place model) have not been able to reliably differentiate between these groups (Shinn & Marston). In a similar study, Shinn et al. included weekly progress monitoring for five weeks in the areas of reading, spelling, and written expression in addition to using CBM to differentiate between students with and without disabilities. They addressed the following research questions: (1) Do students receiving learning disability services perform more poorly on average than low-achieving students, (2) Do students receiving learning disability services show less academic gain over a five-week period than low-achieving students, and (3) Is the rate of learning less for students receiving learning disability services than for low-achieving students? Dependent measures included, in reading, number of words read correctly (WRC) and number of words read incorrectly; in spelling, number of words spelled correctly and number of correct letter sequences (CLS); and in written expression, TWW and TWC. First, an analysis of statistically significant differences between the groups on each of the measures was conducted. Results indicated that reading and spelling measures consistently demonstrated a high relationship with the students' group membership. The written expression measures, however, did not support the differentiation between the groups. They appeared "less reliable in differentiating students in learning disability programs and those who are low achieving" (Shinn et al., p. 549). Second, a week-by-week comparison was made between the groups on each of the measures to determine if the students with learning

disabilities showed fewer gains over the five-week period. Results indicated slopes of improvement in the areas of reading and spelling were not significantly different between the groups of students, and in the area of written expression, the group of students with learning disabilities showed more gains in written expression than the low achieving group. The results of these studies did support the use of CBM in the areas of reading and spelling to differentiate between groups of students. Nonetheless, in terms of using CBM in the area of writing, results were less promising, and indicated that further research was needed.

Parker et al. (1991) expanded on the existing CBM writing research by exploring the suitability of using the six initially-studied measures for the purposes of special education screening and eligibility decisions through two separate research studies with students in grades 2 through 5 and grades 6, 8, and 11. These studies were designed to build upon previous research and correct several limitations of those studies, including small samples on which distributions were based, lack of research on the sensitivity of the measures around and at the cut-off scores, reliability calculated within, rather than across, grade level, and limited research regarding teachers' holistic ratings of student writing. The objective variables explored included TWW, TWC, correct word sequences (CWS), percentage of correctly spelled words (%TWC), and percentage of correct word sequences (%CWS). Additionally, each passage was rated holistically. Five descriptive analyses were completed, including comparing mean scores across grades and from fall to winter within a school year, producing histograms to describe score distributions, comparing percentile ranks for correct word sequences and percentage of correct word sequences, placing standard error of measurement bands on the percentile graphs for

grade 5, and correlating the five countable indices with teachers' holistic ratings of writing quality.

Parker et al. indicated that there were five major findings. First, based on mean score increases across grade levels, all five indices (i.e., TWW, TWC, CWS, %TWC, %CWS) appeared suitable to make eligibility and screening decisions. Second, analysis based on grade-level histograms suggested that at certain grade levels, some indices were less suitable than others. Only the %TWC was suitable at grade 2, and at grade 3, only %TWC and %CWS. At grade 4, %TWC, %CWS, and TWW appeared to be suitable. The remaining indices were deemed unsuitable "because of positive skewness or clustering of the scores at the low end of the scale" (Parker et al., p.13). Third, findings from analyses of percentile line graphs indicated that only %TWC and %CWS were suitable for screening-eligibility decisions. The fourth major finding of the studies indicated that neither CWS or %CWS could distinguish well between percentile ranks near the bottom of the score scale, as they could not distinguish between a 30-percentile point spread (CWS) or a 20-percentile point spread (%CWS). The final major finding of the studies was related to the agreement between the countable indices and the teacher holistic ratings of the writings. These agreements confirm both the validity of CWS, and the weakness of TWW, which is in accordance with findings from previous studies (e.g., Tindal & Parker, 1989a). Overall, Parker et al. recommend the %TWC when making screening-eligibility decisions; however, they indicate it has only been proved to be moderately effective due to the range of uncertainty at the bottom of the percentile ranks.

Watkinson and Lee (1992) expanded on the limited research conducted with students in middle school by examining differences on writing CBM with students with

learning disabilities in the area of written expression and students without learning disabilities. Writing probes were administered to approximately 600 sixth through eighth grade students during their language arts classes. Students were provided with story starters and given one minute to think and three minutes to write. The measures collected included TWW, legible words (Leg: identifiable letters, must approximate known words), TWC, CWS, incorrect word sequences (IWS), percentage of legible words (%Leg), %TWC, and %CWS. Statistically significant differences between the groups were found between both production-dependent measures (i.e., CWS, IWS) and production-independent measures (i.e., %Leg, %TWC, and %CWS). All measures were determined to demonstrate differences between students with learning disabilities and students without learning disabilities, with the exception of TWW, Leg., and TWC. Statistically significant differences between groups existed for all of the production-independent measures (i.e., %Leg., %TWC, and %CWS), and for two production-dependent measures (i.e., CWS and IWS). This study was one of the first to include alternate measures (i.e., production-independent measures). Results indicated that typically-used metrics, TWW and TWC, may not be adequate CBM measures in writing and production-independent measures may be more appropriate measures, indicating the need for further research to determine the ability of production-independent measures to discriminate between groups of students and to reflect student progress over time.

In addition to examining the ability of CBM to differentiate between groups of students, a number of studies have examined the comparison between CBM and teachers' perceptions and ratings of student performance. In one such study, Tindal and Parker (1989a) included a comparison of CBM to teachers' holistic ratings of student

performance. Specifically, they examined the relationship between middle-school students' performance on CBM in writing in order to determine the relation to teachers' holistic ratings of the students' writing products. The performance of students in special education and students in remedial programs were compared. The objective measures collected included TWW, Leg., TWC, CWS, mean length of correct word sequences (MLCWS), %Leg., %TWC, and %CWS. The subjective measures included holistic judgments made by the students' teachers regarding the communicative effectiveness of the writing, which was based on a 7-point Likert scale. Results indicated highly statistically significant differences between the groups with the holistic ratings. Additionally, on the objective measures (i.e., the three production-independent indices of %CWS, %TWC, and MLCWS), there were also statistically significant differences between the groups of students. No statistically significant differences were found on the production-dependent counterparts (e.g., TWW, Leg.) to those measures between the two groups of students. In other words, the production-independent measures, which were computed as ratios (e.g., %CWS), were highly related to the teachers' holistic ratings, while the production-dependent variables (e.g., TWW) were weakly related to the teachers' holistic ratings.

Fewster and MacMillan (2002) sought to find evidence to support the validity of using CBM in the schools by using teachers' future course grades and placement decisions. They tracked the performance of students in grades 2 through 7 over three years in order to determine if CBM scores in reading and writing can reliably predict future academic outcomes in English and social studies classes, as well as future placement in special education, learning assistance, general education, and honors

classes. In the area of reading, words read correctly (WRC) was counted, and in the area of written expression, TWC were counted. Results of the study indicated statistically significant positive correlations between the elementary CBM scores in reading and writing, and secondary school English and social studies grades. Overall, WRC was correlated more highly with the grades than TWC. Correlations were highest in grade 8, and decreased gradually over time. In addition to correlations, a discriminative analysis was completed in order to examine the ability of a combination of CBM scores to differentiate between students classified in program placement groups. Results from this analysis indicated that "there were reliable separation of the four program groups using both CBM measures" (Fewster & MacMillan, p. 153). All differences between groups were statistically significant at the .05 level, with the exception of the differentiation between the special education and the remedial support groups. In addition to corroborating previous research which supports using CBM to predict membership in special education and remedial groups, this study demonstrates the utility of using CBM to predict membership in honors classes as well.

Overall, the research supports the use of CBM to predict performance in school curricula, including group membership (e.g., special education, Chapter 1, general education, honors classes) (e.g., Shinn & Marston, 1985; Watkinson & Lee, 1992). In the area of written expression, however, results of initial studies did not always indicate the ability to differentiate between groups (e.g., Shinn et al., 1986). The initial studies used the most basic writing measures (e.g., TWC, TWW); however, when more "advanced" production-independent CBM writing measures were introduced (e.g., %CWS, %TWC), the validity of these measures to differentiate between groups was determined. In

addition to demonstrating that CBM could reliably differentiate between groups of students, results of studies demonstrated that CBM in the area of writing related closely to teacher ratings of student performance, both through holistic ratings of writing and semester grades.

Using Curriculum-Based Measures with Secondary Students

The literature regarding the use of CBM with secondary students is limited, and more research with this population of students is needed. Results of studies by Tindal, Parker, and colleagues using CBM in writing with secondary students have indicated that measures which were proven to be valid at the elementary level may not be appropriate for secondary level students (e.g., Tindal & Parker, 1989b). Citing the increasing complexity of writing as a possible reason for the inadequacy of the "simple" production-dependent CBM in writing used at the elementary level, such as TWW, TWC, and number of characters written, Espin et al. (1999) explored the criterion-related validity of seven different indicators for predicting secondary students' writing proficiency. Additionally, they explored the possibility that a combination of measures may be a better predictor than a single measure for predicting student performance. The indicators examined included TWW, number of characters written, number of characters per word, number of sentences written, TWC, CWS, and MLCWS. A correlational analysis was completed with these variables, using the standard scores of the language subtest of the California Achievement Test (CAT: CTB/McGraw-Hill, 1985), students' first and second semester English grades, independent ratings of the students' writing samples, and the students' group placement (i.e., LD, basic English, regular English, enriched English) as dependent variables.

Correlational analysis revealed that the number of sentences and number of characters per word were most strongly and consistently correlated with the writing performance measures. Number of CWS and MLCWS also consistently correlated with the writing performance measures, with many of the correlations being statistically significant. Interestingly, the measures most often used at the elementary level, TWW and TWC, were observed to have the weakest correlation with the writing performance measures. A stepwise multiple regression revealed that a combination of variables may be a better predictor of student performance in writing than any single variable. Three variables, characters per word, number of sentences, and MLCWS, accounted for 38% of all variance of the Language Total score on the CAT. Overall, the correlational results of this study indicate that two measures, number of sentences and characters per word, are potentially more adequate predictors of student writing performance at the secondary level than traditional writing CBM (e.g., TWW). The results of the regression analysis supported the hypothesis that a combination of variables would be better than any single predictor. Three variables, characters per word, number of sentences, and MLCWS may be better predictors of student writing performance when used in combination than any measure used alone.

A follow-up study by Espin et al. (2000) sought to further extend the research conducted at the secondary level. Their study examined three different issues: the reliability and validity of the quantitative indicators, the effects of the type of writing and sample duration on reliability and validity, and whether a combination of measures would be a better indicator of student writing performance than a single measure. The predictor variables examined in this study included TWW, TWC, words spelled incorrectly (TWI),

total number of characters written, sentences written, characters per words, words per sentence, CWS, correct word sequences minus incorrect word sequences (CWS-IWS), and MLCWS. Correlation coefficients were determined by computing the correlation between each CBM measure and the teacher's rating of the student's writing proficiency. The correlations all proved to be moderate with the exception of CWS-IWS, which had moderately strong correlations. This pattern was consistent when comparing the CBM measures to the district writing test. When looking at a combination of measures, no other variable added to the strength of the predictive validity beyond CWS-IWS. These results do not necessarily support previous findings by Espin et al. (1999), except to determine that traditional CBM measures used with elementary aged students in the area of writing do not appear to be adequate with secondary students.

Weissenburger and Espin (2005) studied the technical adequacy of CBM in writing across three different grade levels, fourth, eighth, and tenth grades. CBM examined included TWW, CWS, and CWS-IWS. Measures were taken for three different durations: 3, 5, and 10 minutes. Alternate form reliability was examined, as well as criterion-related validity to a state standards test (i.e., Wisconsin Knowledge and Concept Examinations). Results indicated that at all grade levels, CWS and CWS-IWS had stronger criterion-related validity than TWW, and at all grade levels, CWS-IWS was the strongest predictor of performance. Results also indicated that the technical adequacy of CBM in written expression decreases with age, but this finding was less pronounced for the more complex measure of CWS-IWS. Finally, data showed decreasing alternate form reliability coefficients with shorter sample lengths for older students.

A recent study by Espin et al. (2008) examined the predictive validity of CBM in writing to a state standards test for high school students. This study was unique in that it examined CBM with high school students, and most previous studies with secondary level students were conducted with middle school students. The results of this study indicate that the measures that had the strongest validity and reliability coefficients were CWS-IWS and CWS (in comparison to TWW and TWC). Additionally, this study examined curriculum based measures written for four different durations (i.e., 3, 5, 7, and 10 minutes). Similar to the Weissenburger and Espin (2005) study, results indicated that alternate form reliability increased steadily with writing time up to seven minutes, with strongest coefficients found for 7 and 10 minute writing probes, but with little difference between these two.

In summary, the research completed at the secondary level indicates that measures used at the elementary level may not be appropriate for use with secondary students, and using multiple measures may be a better predictor of student performance. Results from studies conducted with secondary students have produced some conflicting results, pointing to a need for future research in this area. Specifically, the variables which need to be examined further with this population of students include number of sentences, characters per words, MLCWS and CWS-IWS. Additionally, at the secondary level, longer duration of writing samples may produce increased reliability and validity of the measures.

Alternate Methods for Evaluating Writing with Curriculum-Based Measures

A small number of research studies have examined alternative CBM in writing, including previously discussed studies by Tindal and Parker (1989a), Watkinson and Lee

(1992), and Espin et al. (1999). Gansle et al. (2002) sought to expand on the existing research base on CBM in writing by comparing a wide range of variables to criterion variables, including standardized tests and teacher ranking of student writing in order to assess the validity of these measures. Predictor variables examined included TWW, parts of speech (i.e., nouns, verbs, and adjectives), long words, TWC, total punctuation marks, correct punctuation marks, correct capitalization, complete sentences, CWS, sentence fragments, simple sentences, and a number of computer-scored variables (i.e., MS Word Flesch Reading Ease, MS Word Flesch-Kincaid Grade Level, WP Flesch-Kincaid Grade Level, WP Sentence Complxity, and WP Vocabulary Complexity). This study was the first to examine many of these variables, including parts of speech, total punctuation marks, correct punctuation marks, correct capitalization, complete sentences, sentence fragments, simple sentences, and the computer-scored variables. In addition to calculating correlations between each of the predictor variables and the criterion variables, the researchers calculated interscorer reliability, alternate forms reliability, and multiple regressions for each of the criterion variables.

Participants included 179 third and fourth grade students who completed two 3-minute writing probes on two consecutive days. Interscorer reliability was calculated, and all but four variables were above .80. The four variables which fell below had agreements between .71 and .80 and included complete sentences, words in complete sentences, sentence fragments, and simple sentences. Despite lower agreements, these variables were still included in the analyses due to the exploratory nature of the study. Pearson correlations were calculated between the criterion measures and the CBM in writing. With the Iowa Test of Basic Skills (ITBS) Language Usage score, the statistically

significant correlations were found with correct punctuation marks (r =.36), words in complete sentences (r =.34), CWS (r =.36), and simple sentences (r =.32). For the Total Subscale score of the ITBS, statistically significant correlations were found with long words (r =.33), total punctuation marks (r =.43), correct punctuation marks (r =.44), words in complete sentences (r =.34), CWS (r=.43), and simple sentences (r =.38). For the Louisiana Educational Assessment Program (LEAP) write competently subscale, statistically significant correlations were found with number of verbs (r =.33), TWC (r =.29), and CWS (r =.28). For the LEAP conventions of language subscale, a statistically significant correlation was found only with CWS (r =.41). In order to further examine the relationship between variables and criterion test scores, a series of stepwise forward multiple-regression analyses were carried out. For third graders, the variables entering the regression equation for the LEAP language usage subscale included CWS, verbs, and correct punctuation marks. The variables entering the equation for the LEAP language total score included correct punctuation marks, CWS, and long words. For the fourth graders, the variables entering the equation for the LEAP write competently scale included verbs and the WP vocabulary complexity score. For the use of conventions subscale, the variables included CWS, TWW, and nouns. When examining the correlations with teacher rankings of student writing, the variables significantly related included CWS, TWW, and correct punctuation marks.

The results of this study present a number of interesting and important findings that point to the need for future research in this area. Based on correlations with criterion test scores as well as teacher rankings, two new variables presented themselves as possible choices as an index of writing skill, including correct punctuation marks and

words in complete sentences. In addition, results supported the validity of CWS as a CBM writing measure. Future research should examine these variables more closely, and should develop procedures to quickly train scorers to score words in complete sentences more reliably. The results of this study also support previous research that demonstrates that TWW may not be the best predictor of writing skill. Because this study was the first to examine many of the variables, the social validity of these measures requires replication.

Rationale for the Current Study

An obvious strength of CBM is its cost effectiveness to school districts. As a formative evaluation tool, CBM is relatively inexpensive to reproduce, requires minimal administration time, can be given frequently to assess student progress over short periods of time, and is sensitive to small improvements in student performance. Effective educators need formative evaluation tools to guide their decision making when designing instruction. Over the past 20 years, much of the CBM research has been conducted in the areas of reading and math, whereas very little has focused on writing (Gansle et al., 2002; Jewell & Malecki, 2005). This may be due, in part, to the complexity of the construct of writing. Since the development of CBM, researchers have been making gains in the area of writing CBM, including the development of new measures such as production-independent measures (e.g., %CWS, %TWC) (e.g., Parker et al., 1991). Despite these gains, CBM in writing continues to be the least researched area of CBM (Gansle et al.). Gansle et al. provided the field with a seminal study that has not been replicated to date. In order to assess the validity of the measures that appeared to be promising in terms of written expression, the measures must be examined more closely, which this study does.

The current study replicated and extended the work of Gansle and colleagues, using a substantially larger sample of participants in order to further extend the research in the area of CBM in writing.

The participating school district has evaluated the practical utility of CBM in reading by examining the relationship between CBM in reading and students' performance on the Michigan Educational Assessment Program's (MEAP) fourth grade reading assessment (McGlinchey & Hixson, 2004). The school district was pleased with the degree to which CBM reading probes predicted later performance on the MEAP and has begun using CBM reading probes as a formative evaluation measure in its K-6 schools. The participating school district wanted to evaluate the practical utility and predictive validity of CBM in the area of writing. Results may be used to employ CBM writing measures to screen students more frequently in order to identify those students in need of more intensive intervention. The school district requested our assistance in this evaluation process and we were interested in providing such assistance for two main reasons. First, this evaluation process could improve the school district's capacity to improve students' written expression in a proactive and preventative manner. Second, data concerning the utility of writing CBM that are collected by the school district may extend the current base of research in the area, and may be a topic of interest to many other researchers, as well as school staff.

CHAPTER II

METHOD

Participants

Participants in this study included 700 fourth grade students in a large urban school district in a Midwestern city located within a North Central state in the United States. All 14 elementary schools within the school district participated in the study. All students in the fourth grade at the elementary schools participated in the study if they were present when the writing probe was administered. Three hundred fifty five (50.7%) students were female and 345 (49.3%) students were male. Three hundred sixty one (51.6%) students were African American, 274 (39.1%) were Caucasian, 45 (6.4%) were Hispanic, and for 20 students (2.9%) ethnicity information was unavailable. Four hundred fifty six (65.1%) of the participants were identified as being economically disadvantaged (i.e., received free or reduced lunch).

For the purpose of this study, the primary investigator requested archival data from the school district in order to extend the research in the area of CBM in writing, as well as provide the school district with useful information concerning the use of writing CBM probes as formative evaluation tools. Archival data requested included information regarding students in the fourth grade during the 2002-2003 school year, including the students' performance on the Michigan Educational Assessment Program (MEAP) English Language Arts (ELA) assessments administered in February 2003 and October 2005, their performance on the Iowa Test of Basic Skills (ITBS) Reading assessment administered in April 2005, and copies of the CBM writing probes administered during the spring of 2003.

Scorers

One individual with a specialist degree in education in school psychology, three individuals with a master's degree in speech pathology, and one individual with a bachelor's degree in education, who are all employed full-time within public school districts, scored the writing probes.

Independent Variables

This study examined the relationship between the CBM in writing and two standardized assessments (i.e., MEAP, ITBS). To further explore variables that might have a stronger predictive relationship than TWW, nineteen independent variables were calculated for each writing probe. Independent variables were chosen based on measures used in previous research (Gansle et al., 2002). This study used the same definitions employed by previous researchers (Shinn, 1989; Gansle et al.). The definitions are as follows:

Total Words Written

The total number of words written during the 3-minute period were recorded, including words spelled incorrectly. Numbers that were not spelled out were not counted as words (Shinn, 1989).

Total Words Correct

The total number of words spelled correctly on each probe were counted. For a word to be included, it was not necessary that it be spelled correctly in context. If the word was spelled correctly in isolation, it was included (Shinn, 1989).

Parts of Speech

The total number of nouns, verbs, and adjectives written were recorded for each writing probe. Those parts of speech were defined based on Howell and Memering (1993).

Long Words

The total number of words on each probe that were spelled correctly and contained eight or more letters were counted.

Punctuation Marks

The total number of punctuation marks were counted, regardless of whether or not they were used correctly. When a set of double quotation marks were used (e.g., "I like it"), two punctuation marks were counted.

Correct Punctuation Marks

Punctuation marks that were applied correctly were counted. "If a punctuation mark appeared in the passage, a determination was made whether it was in the correct location in the sentence (e.g., a period, question mark, or exclamation mark appeared at the end of the sentence, and after a subject/verb combination) and whether it was appropriate for that sentence in that location (e.g., a question mark followed a question word such as 'what' or 'how')" (Gansle et al., 2002, p. 482).

Correct Capitalization

Correct uses of capital letters were counted. This included words at the beginning of a sentence, proper nouns, and proper nouns within quotation marks (Gansle et al., 2002).

Correct Word Sequences

Correct word sequences were counted according to Shinn's (1989) definition: "Count as a word sequence the joining of two words together that are spelled correctly and are grammatically correct" (p. 241). Each two-word sequence was considered in isolation. If two words in sequence could be correct as written in the context of any sentence, they were counted as a correct word sequence. For example, "He go to the store" was counted as three correct word sequences. He-go: incorrect, go-to: correct 1, to-the: correct 2, the-store: correct 3. Punctuation was not considered in the correct word sequence count. Incorrect word sequences were also counted, using the same definition as above. In addition, three variables derived from these measures were calculated, including correct minus incorrect word sequences (CWS-IWS), percentage of correct word sequences (%CWS), and percentage of incorrect word sequences (%IWS).

Complete Sentences

Complete sentences were counted using the rules outlined in Gansle et al. (2002). It was required that they (a) start with a capital letter, (b) have a recognizable subject, (c) have a verb, and (d) have ending punctuation.

Words in Complete Sentences

The total number of words in all sentences that are counted as complete were counted. Percentage of words in complete sentences was also calculated.

Fragments

Sentence fragments were counted based on the definition by Howell and Memering (1993, p. 76): "A sentence fragment is incomplete. Part of the sentence is missing."

Simple Sentences

Simple sentences were counted according to the definition found in Howell and Memering (1993): "A simple sentence is one independent clause... Each contains only one subject and one main verb" (p.73). Only sentences that were counted as complete sentences as defined above could be scored as simple sentences.

Not Simple Sentences

Not simple sentences were counted based on definitions for compound, complex, and compound-complete sentences found in Howell & Memering (1993). "A compound sentence contains at least two independent clauses but no dependent clauses" (p.73). "A complex sentence contains one independent clause and at least one dependent clause" (p. 74). "A compound-complex sentence contains two (or more) independent clauses and one (or more) dependent clauses" (p.74).

Dependent Variables

Michigan Educational Assessment Program

In Michigan, students are required to take the Michigan Educational Assessment Program (MEAP), or an alternate test, at various points in their educational career, assessing several different academic domains. The MEAP is a criterion-referenced testing program initiated by the State of Michigan during the 1969-1970 school year for the purpose of determining what students can do in comparison to the standards set by the State Board of Education. During the 2002-2003 and the 2005-2006 school year, students in the fourth and seventh grades respectively, were required to take a Language Arts portion of the assessment, including both a reading and a writing section. The writing score was based upon two writing samples, writing from knowledge and experience, for

which students were given a theme to write about, and writing in response to reading, for which students first read a passage and then wrote answers to specific questions regarding the passage. The total ELA score was based on a combination of the students' performance on the reading and writing portions of the MEAP. For the edition of the MEAP administered during the 2002-2003 and 2005-2006 school years, there were four levels of scores on the writing and ELA portions (i.e., Level 1, or "exceeded standards," Level 2, or "met standards," Level 3, or "basic performance," and Level 4, "apprentice"). Students who achieved a level 1 or level 2 were considered to have "passed" the MEAP. Both the writing assessment and the ELA assessment scores were included as dependent measures in this study.

Iowa Test of Basic Skills

The Iowa Test of Basic Skills (ITBS) is a standardized, norm-referenced assessment initially developed at The University of Iowa College of Education (2007). The assessments are designed to provide achievement scores to monitor a student's progress from year to year. The ITBS is used as an annual assessment of all students in Iowa, but many other school districts across the country use this assessment to monitor their students' academic achievement. The students in the current study took the reading portion of the ITBS during their fifth grade year, and the results from that assessment were used for the current study. The reading portion of the administered ITBS included the following subtests: Vocabulary, Reading Comprehension, Spelling, Capitalization, Punctuation, and Usage and Expression.

Time Variables

For 21% of the writing probes, the scorer recorded the time it took her to score each hand-scored variable, using a hand-held stopwatch or timer.

Procedures

Probe Administration

CBM writing probes were administered at 14 elementary schools, which included 39 fourth grade classrooms, over a period of three weeks in the spring of 2003, using story starters that were approved by curriculum specialists at the school district. Within two weeks of the first administration, a second administration with a different story starter was completed in 11 classrooms in order to assess the stability of the scores over a short period of time. All writing probes were administered in whole-class format, following procedures outlined by Shinn (1989).

Procedural Integrity

Thirty-nine classrooms participated in this study, and eleven were administered the writing probe on two occasions. Of those 50 administration sessions, 10 (20%) were assessed for procedural integrity of administration of the writing probes by a trained observer. In all ten sessions assessed for procedural integrity, 100% of steps were completed correctly.

Scorer Training

Scoring instructions for each predictor variable were developed and refined by Gansle et al. (2002). These scoring instructions were used with slight alterations in scoring procedures in the present study. In order to train scorers on each of the predictor variable definitions, as a group, the scorers reviewed each of the definitions, including

examples, with the first author. Scorers scored each of the variables using practice writing probes and discussed the scoring definitions to explain discrepancies. After all scorers agreed that the definitions were sufficiently clear, each scorer received copies of five randomly selected practice writing probes to compute each of the measures. When completed, the probes were scored for reliability. On all measures that had less than 80% reliability, the definitions were again reviewed and practiced as a group with practice writing probes. Following the second training session, scorers were provided with a packet of 15 - 20 writing probes that they were to score independently for interscorer reliability. When they were completed, their probes were scored for accuracy. The scorer was then provided with a second packet of 15 – 20 probes to score independently.

Scoring

All probes were independently scored by one of the three scorers. Ninety-seven of the probes (13.9%) were scored a second time by a second independent scorer to assess interscorer reliability.

Analyses

Multiple stepwise regressions were completed to determine the best predictors of the dependent variables (i.e., MEAP writing assessment, MEAP ELA assessment, ITBS Reading assessment). A stepwise method (Afifi & Clark, 1990) was used, which is appropriate for exploratory analyses and allowed for adding variables to the predictive equation (Pedhazur, 1982). Stepwise regression starts with only the predicted variable, after which the computed *F*-to-enter is calculated for each variable and the variables entered one at a time (Afifi & Clark). The *F*-to-enter is a measure of the statistical association between each predictor and the criterion. After each step, the *F*-to-enter is

recalculated for the residual variance in the criterion, and the predictor with the largest *F* is entered. After each predictor has been added to the regression equation, all the predictors already in the equation are reexamined to determine whether they should be removed. A partial F test is performed on the predictor already in the equation that produced the smallest increment in R^2. If the predictor no longer satisfies the criteria for inclusion, it is removed from the equation. This process continues until no remaining predictor's *F* ratio is statistically significant based on the probability value set as the criterion to enter, which was $p \leq .05$ for these analyses.

CHAPTER III

RESULTS

Preliminary Data Analyses

Descriptive Statistics and General Assumptions of Analyses

As a first step in the data analyses, descriptive statistics were calculated on all independent and dependent variables to determine the mean, standard deviation, skewness and kurtosis. Please see Table 1 for the mean and standard deviation for all independent variables. Data were checked for accuracy, linearity, normality, and homoscedasticity. Bi-variate scatterplots were examined in order to assess linearity. Normality was assessed by examining bi-variate scatterplots as well as examining variable statistics for skewness and kurtosis, and by the Kolomogorov-Smirnov statistic. Homoscedasticity was also assessed by examining bi-variate scatterplots. Outliers were identified using box plots, and subsequently these outliers were checked for accuracy. No cases were dropped from the analyses due to data error.

Evaluations of normality led to transformation of several independent and dependent variables. A square root transformation was applied to the following variables: total words written (TWW), total words correct (TWC), nouns, verbs, adjectives, correct word sequences (CWS), not-simple sentences, ITBS in Reading, and 2005 MEAP writing assessment. A logarithm formula was used to transform punctuation marks, correct punctuation marks, correct capitalization, complete sentences, and incorrect word sequences (IWS). An inverse transformation formula was used to transform long words, fragments, and simple sentences. A reflect and square root formula was used to transform

%CWS and %WdsSent. These transformations resulted in variables that appeared to be more normally distributed, based on the skewness statistics. The variables that were normally distributed and did not require data transformations included words in complete sentences, correct minus incorrect word sequences (CWS-IWS), 2003 MEAP writing assessment, 2003 MEAP ELA assessment. The completed transformations resulted in normal distributions for each individual variable; however, multivariate normality was not achieved across the data set, therefore, the transformations were not successful. Based on this information, and the idea that regression formulas are robust to violations of normality, it was decided that only the untransformed data would be utilized for the analyses.

Pearson Correlations Between Variables

Correlations between independent variables. Pearson correlations were run between the independent variables in order to assess multicollinearity. Please see Table 2 for these statistics. High correlations existed for TWW with TWC, nouns, verbs, and CWS; TWC with nouns, verbs, and CWS; CWS with nouns and verbs; and correct punctuation marks with punctuation marks and complete sentences. High correlations between these variables suggest that in each regression equation they may share a large amount of common variance within the dependent variable, and subsequently regression equations that include highly correlated variables should be interpreted with caution due to expected instability of the results. After running the regressions, tolerance values were examined in order to better identify independent variables most impacted by multicollinearity. Across the 10 regression formulas, only three variables were identified

as having tolerance values below 0.10: TWW, TWC, and punctuation marks, and these were within the excluded variables in the regression equations.

Correlations between independent variables and MEAP assessments. Pearson correlations were calculated between each independent variable and MEAP assessment scores (please see Table 3 for these statistics). Correlations were small to moderate, and the 2003 MEAP assessments had the highest correlations with independent variables. For the 2005 MEAP writing assessment, correlations with independent variables ranged from .09 (fragments) to -.32 (CWS-IWS). For the 2003 MEAP writing assessment, the correlations ranged from .14 (IWS) to -.40 (words in complete sentences). Correlations with the 2005 ELA assessment ranged from .15 (IWS) to -.37 (words in complete sentences), and correlations with the 2003 ELA assessment ranged from -.20 (long words) and .20 (IWS) to -.42 (CWS-IWS). For both the writing and ELA assessments in 2005 and 2003, all correlations were negative, with the exceptions of fragments and IWS. This was expected, as the lowest MEAP score (i.e., 1) actually represents the highest achievement.

Correlations between independent variables and ITBS Reading assessment. Pearson correlations were calculated between each independent variable and the ITBS score (please see Table 3 for these statistics). Correlations were positive, with the exception of fragments and IWS, and ranged from -.16 (IWS) to .35 (punctuation marks).

Reliability

Interscorer agreement. Mean agreement and standard deviation of agreement for each variable are presented in Table 4. All variables have average interscorer reliability scores above 80%. Ninety-seven probes (13.9%) were scored by a second scorer to assess

Table 1

Mean, standard deviation, and alternate form reliability for independent variables.

Independent Variable	M (n = 700)	SD (n = 700)	r (n = 199)
Total words written	35.14	15.07	.68
Total words correct	32.12	14.98	.68
Nouns	11.63	5.04	.60
Verbs	10.63	5.25	.50
Adjectives	4.72	2.89	.42
Long words	.72	1.14	.19
Punctuation marks	2.63	2.21	.41
Correct punctuation marks	2.46	2.16	.42
Correct capitalization	3.91	2.92	.43
Complete sentences	1.69	1.48	.53
Words in complete sentences	24.20	18.50	.52
Simple sentences	.65	1.07	.40
Not-simple sentences	1.06	.94	.47
Fragments	1.03	1.29	.35
Correct word sequences	27.92	14.00	.68
Incorrect word sequences	6.32	5.25	.60
Percentage correct word sequences	79.09	17.21	.54
Percentage words in sentences	65.15	40.23	.41
Correct minus incorrect word sequences	21.59	15.46	.66

Table 2

Pearson correlations between independent variables (n=700).

	TWW	TWC	Long	Nouns	Verbs	Adj	Tot Punc	Cor Punct	Cor Cap
TWC	.98**	-							
Long	.18**	.23**	-						
Nouns	.89**	.87**	.22**	-					
Verbs	.87**	.86**	.16**	.72**	-				
Adj	.70**	.69**	.08*	.62**	.45**	-			
TotPunct	.45**	.48**	.23**	.50**	.39**	.33**	-		
CorrPunct	.45**	.48**	.23**	.49**	.41**	.33**	.98**	-	
CorrCap	.63**	.64**	.21**	.67**	.58**	.40**	.56**	.57**	-
TotSent	.44**	.46**	.19**	.46**	.43**	.32**	.77**	.80**	.59**
WdsSent	.64**	.66**	.26**	.61**	.59**	.46**	.59**	.61**	.58**
CWS	.92**	.96**	.27**	.82**	.83**	.65**	.45**	.46**	.61**
Frag	.19**	.14**	-.11**	.13**	.11**	.14**	-.18**	-.21**	-.01
S.Sent	.19**	.20**	.07	.21**	.19**	.16**	.60**	.61**	.43**
N.S.Sent	.48**	.51**	.23**	.49**	.46**	.33**	.54**	.57**	.46**
IWS	.24**	.07	-.23**	.19**	.13**	.16**	-.15**	-.17**	-.02
%CWS	.34**	.48**	.31**	.33**	.37**	.25**	.34**	.35**	.32**
%Wds	.21**	.24**	.20**	.23**	.23**	.15**	.45**	.47**	.30**
CWS-IWS	.75**	.84**	.32**	.68**	.71**	.53**	.46**	.47**	.56**

* $p < .05$. ** $p < .01$.

Table 2 – Continued

	Tot Sent	Wds Sent	CWS	Frag	S. Sent	N.S. Sent	IWS	%CWS	%Wds
WdsSent	.74**	-							
CWS	.44**	.66**	-						
Frag	-.37**	-.48**	.09*	-					
S.Sent	.77**	.34**	.16**	-.14**	-				
N.S.Sent	.69**	.79**	.51**	-.42**	.08*	-			
IWS	-.14**	-.09*	-.09*	.32**	-.07	-.16**	-		
%CWS	.32**	.39**	.39**	-.15**	.11**	.38**	-.70**	-	
%Wds	.65**	.81**	.81**	-.64**	.31**	.68**	-.20**	.31**	-
CWS-IWS	.45**	.63**	.94**	-.03	.17**	.52**	.43**	.79**	.31**

Note. TWW = total words written, TWC = total words correct, Long = long words, Adj = adjectives, TotPunct = punctuation marks, CorrPunct = correct punctuation marks, CorrCap = correct capitalization, TotSent = complete sentences, WdsSent = words in complete sentences, CWS = correct word sequences, Frag = sentence fragments, S.Sent = simple sentences, N.S.Sent = not simple sentences, IWS = incorrect word sequences, %CWS = percentage correct word sequences, %Wds = percentage words in complete sentences, CWS-IWS = correct minus incorrect word sequences.

* $p < .05$. ** $p < .01$

interscorer reliability. In order to calculate interscorer reliability for the pair of scores for each variable, the smaller number was divided by the larger number, and the result was multiplied by 100. Scorers completed approximately 15-20 probes at a time. Scorers were not asked to re-score any probes, even when the reliability was below .80 for a particular independent variable in the group of probes. This scoring procedure is thought to be a more accurate representation of how writing probes would be utilized in a school setting. If, after the extensive training and practice that was completed, reliability among scorers was still low, that particular measure may not be the most appropriate to use within a school setting.

Delayed alternate forms reliability. In order to assess the stability of the measures over a short period of time, a second writing probe was administered to 199 (28.4%) students within two weeks of the initial measure. A different story starter was used, as it was thought if the same story starter was used, students' performance could be enhanced due to practice effects. Reliability was calculated as a Pearson correlation between the two variables. Pearson correlations ranged from .19 (long words) to .68 (TWW) for the variables, and all correlations were positive. The data on the alternate forms reliability are presented in Table 1.

Time to Score Independent Variables

Due to the exploratory nature of this study, the duration to score each independent variable was collected in seconds. Timed data were collected on 21.7% of writing probes. The first author as well as two other scorers completed timings for each measure. Please see Table 4 for these data. Sentence fragments, simple sentences, not-simple sentences and correct punctuation marks took the least amount of time to score at

approximately 3-4 seconds to score per probe. CWS took the longest time to score, at approximately 43 seconds per probe.

Variables to Be Entered in the Regression Equations

After examining initial descriptive and correlational analyses, one concern was the high levels of multicollinearity between several of the independent variables. When examining these correlation statistics, as well as the concept that each variable purports to measure, it was determined that %IWS would be removed from the analyses due to the very high correlation with %CWS ($r = -1.0$) and because the two variables measure the inverse of the same concept. Despite the presence of high correlations between other variables, no other variables were removed from the analyses, as it was felt that it is the nature of the measures to be highly correlated, and they were in fact measuring different concepts. Therefore, the complete set of stepwise regression analyses included TWW, TWC, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, words in complete sentences, simple sentences, not-simple sentences, fragments, CWS, IWS and %CWS (variables not included were percentage of words in sentences and CWS-IWS). Stepwise regressions were run with these independent variables for each of the dependent variables.

In addition to the regression analyses including the complete set of independent variables, the author felt that it may be appropriate to single out production-independent, as well as more quality-based measures (vs. quantity based measures) and run a second set of stepwise regression analyses. Previous research had demonstrated the validity of production-independent measures (Tindal & Parker, 1989a; Watkinson & Lee, 1992; Espin et al., 2000). Due to the need for additional research on the validity and

Table 3

Pearson correlations between dependent and independent variables.

	MEAP 2003		MEAP 2005		ITBS
	Writing (n = 622)	ELA (n = 584)	Writing (n = 517)	ELA (n = 517)	Reading (n = 532)
TWW	-.32**	-.30**	-.27**	-.26**	.23**
TWC	-.35**	-.33**	-.29**	-.29**	.26**
Long	-.19**	-.20**	-.13**	-.18**	.21**
Nouns	-.29**	-.29**	-.28**	-.27**	.22**
Verbs	-.31**	-.31**	-.25**	-.26**	.25**
Adj	-.25**	-.24**	-.23**	-.22**	.17**
TotPunct	-.28**	-.37**	-.28**	-.31**	.33**
CorrPunct	-.30**	-.39**	-.28**	-.32**	.35**
CorrCap	-.28**	-.32**	-.29**	-.31**	.24**
TotSent	-.32**	-.37**	-.27**	-.35**	.33**
WdsSent	-.40**	-.40**	-.30**	-.37**	.33**
Frag	.17**	.20**	.09*	.18**	-.17**
S.Sent	-.16**	-.21**	-.16**	-.19**	.18**
N.S.Sent	-.32**	-.35**	-.24**	-.33**	.30**
CWS	-.38**	-.38**	-.31**	-.32**	.29**
IWS	.14**	.20**	.10**	.15**	-.16**
CWS-IWS	-.40**	-.42**	-.32**	-.34**	.32**

* $p < .05$. ** $p < .01$

Table 3 – Continued

	MEAP 2003		MEAP 2005		ITBS
	Writing (n = 622)	ELA (n = 584)	Writing (n = 517)	ELA (n = 517)	Reading (n = 532)
% CWS	-.32**	-.38**	-.30**	-.34**	.29**
% WdsSent	-.29**	-.29**	-.21**	-.30**	.26**

Note. TWW = total words written, TWC = total words correct, Long = long words, Adj = adjectives, TotPunct = punctuation marks, CorrPunct = correct punctuation marks, CorrCap = correct capitalization, TotSent = complete sentences, WdsSent = words in complete sentences, CWS = correct word sequences, Frag = sentence fragments, S.Sent = simple sentences, N.S.Sent = not simple sentences, IWS = incorrect word sequences, %CWS = percentage correct word sequences, %Wds = percentage words in complete sentences, CWS-IWS = correct minus incorrect word sequences.

* $p < .05$. ** $p < .01$

Table 4

Interscorer agreement and time to score in seconds for independent variables.

	Interscorer Agreement		Scoring Time (in seconds)	
	M (n = 93)	SD (n = 93)	M (n = 152)	SD (n = 152)
TWW	99.40	1.37	13.90	6.16
TWC	97.95	2.91	20.55	9.90
Nouns	93.26	9.18	6.60	4.30
Verbs	93.93	8.01	22.40	10.90
Adj	81.92	22.05	21.40	10.01
Long	87.29	30.60	13.25	7.87
TotPunct	93.53	15.77	8.25	6.16
CorrPunct	92.97	17.63	4.60	9.73
CorrCap	89.39	21.97	8.40	5.00
TotSent	95.67	16.26	7.25	6.80
WdsSent	94.99	14.90	3.75	2.75
S.Sent	82.73	32.64	43.20	21.79
N.S.Sent	83.59	28.30	3.50	4.72
Frag	89.25	27.14	3.20	2.98
CWS	93.76	11.33	3.75	3.69
IWS	84.89	19.50	37.70	23.94

Table 4 – Continued

	Interscorer Agreement		Scoring Time (in seconds)	
	M (n = 93)	SD (n = 93)	M (n = 152)	SD (n = 152)
% CWS	95.37	2.60	3.75	3.69
% WdsSent	96.56	3.84	3.20	2.98
CWS-IWS	85.81	9.78	41.45	

Note. Scoring for percentage correct word sequences, percentage words in complete sentences and correct minus incorrect word sequences was computerized; therefore, scoring time reported is for the production-dependent counterpart of percentage of correct word sequences and percentage of words in sentences. For correct minus incorrect word sequences, the total scoring time of correct word sequences and incorrect word sequences were added together. TWW = total words written, TWC = total words correct, Long = long words, Adj = adjectives, TotPunct = punctuation marks, CorrPunct = correct punctuation marks, CorrCap = correct capitalization, TotSent = complete sentences, WdsSent = words in complete sentences, CWS = correct word sequences, Frag = sentence fragments, S.Sent = simple sentences, N.S.Sent = not simple sentences, IWS = incorrect word sequences, %CWS = percentage correct word sequences, %Wds = percentage words in complete sentences, CWS-IWS = correct minus incorrect word sequences.

effectiveness of production-independent variables, as well as the high multicollinearity among many independent variables in the current study, a second set of regressions were run with each dependent variable which included the following independent variables: TWC, long words, nouns, verbs, adjectives, correct punctuation marks, correct capitalization, complete sentences, fragments, simple sentences, not-simple sentences, and %CWS. These regressions also allowed for the inclusion of two less frequently researched variables in the equation: CWS-IWS and %WdsSent. Previous studies have demonstrated that CWS-IWS and %WdsSent may be valid predictors (e.g., Espin et al., 2000; Gansle et al., 2002), and additional data is needed to further examine the utility of these variables.

Regression Models

Complete Set of Independent Variables

2005 MEAP writing assessment. Please refer to Table 5 for the statistics for this regression equation. The first variable to enter the regression equation for the 2005 MEAP writing assessment, when examining the complete set of independent variables was CWS, accounting for 9.9% of the variance. The second variable to enter the regression equation was correct punctuation marks, accounting for an additional 2.6% of the variance, for a total of 12.5%. The third variable to enter the regression equation was %CWS, which accounted for an additional 1.7% of the variance, for a total of 14.2%. The fourth variable to enter the regression equation was IWS, which accounted for an additional 0.8% of the variance, for a total of 15.0%. In the fifth step, CWS was removed from the regression equation, removing less than 1% of the variance, for a total of 14.8%. In the sixth and final step, correct capitalization entered the equation, accounting for an

additional 0.9% of the variance, for a total of 15.7% ($F_{1, 512} = 23.91$, $p < .01$). The Beta coefficients in the final step of the regression equation were small: correct punctuation -.14, %CWS -.33, IWS -.15, and correct capitalization -.12. CWS, correct punctuation marks, %CWS and correct capitalization were negatively correlated with the MEAP assessment, meaning that as the particular variable increased, MEAP performance also increased. In contrast, IWS was positively correlated with the assessment meaning that as this variable increased, MEAP performance decreased.

2003 MEAP writing assessment. Please refer to Table 6 for the statistics for this regression equation. The first variable to enter the regression equation for the 2003 MEAP writing assessment, when examining the complete set of independent variables was words in complete sentences, which accounted for 15.8% of variance. The second variable to enter the equation was %CWS, which accounted for an additional 3.3% of the variance, for a total of 19.1%. In the third and last step of the model, CWS entered the equation, accounting for an additional 0.7% of variance, for a total of 19.8% of variance accounted for by this equation ($F_{1, 618} = 50.91$, $p < .01$). The Beta coefficients in the final step of the regression equation were small: words in complete sentences -.26, percentage of correct word sequences -.14, and correct word sequences -.13. All variables were negatively related to the dependent variable, meaning that as one independent variable increased, MEAP performance also increased.

2005 MEAP ELA assessment. Please refer to Table 7 for the statistics for this regression equation. The first variable to enter the regression equation for the 2005 MEAP ELA assessment, when examining the complete set of independent variables, was words in complete sentences, accounting for 13.8% of variance. In the second step,

%CWS was entered, accounting for an additional 4.9% of variance, for a total of 18.7%. In the third and final step of the model, complete sentences was entered, accounting for an additional 1.1% of variance, for a total of 19.3% of variance accounted for by this model ($F_{1, 513} = 42.25$, $p < .00$). The Beta coefficients in the final step of the regression equation were small: words in complete sentences -.17, %CWS -.23, and complete sentences -.16. All variables were negatively related to the MEAP ELA assessment meaning that as the particular variable increased, MEAP performance also increased.

2003 MEAP ELA assessment. Please refer to Table 8 for the statistics for this regression equation. The first variable to enter the regression equation for the 2003 MEAP ELA assessment, when examining the complete set of independent variables was words in complete sentences, accounting for a total of 15.9% of variance. In the second step, %CWS was entered, accounting for an additional 5.9% of variance, for a total of 21.8%. In the third and final step, correct punctuation marks was entered, accounting for an additional 2.3% of variance, for a total of 24.1% of variance accounted for by this model ($F_{1, 580} = 61.30$, $p < .00$). The Beta coefficients in the final step of the regression equation were small: words in complete sentences -.19, %CWS -.24, and correct punctuation -.19. All variables were negatively correlated with the MEAP, meaning that as the particular variable increased, MEAP performance also increased.

ITBS reading assessment. Please refer to Table 9 for the statistics for this regression equation. The first variable to enter the regression equation for the ITBS reading assessment, when examining the complete set of independent variables, was correct punctuation marks, which accounted for 12.2% of the variance. Percentage of CWS entered the equation in the second step, accounting for an additional 3.4% of the

variance for a total of 15.6%. In the third step, words in complete sentences entered the regression equation, accounting for an additional 1.3% of the variance, for a total of 16.9%. In the fourth and final step, long words entered the equation, accounting for an additional 0.7% of the variance. The variables entering this equation accounted for a total of 17.6% of the variance ($F_{1, 527} = 28.18$, $p < .00$). The Beta coefficients in the final step of the equation were small: correct punctuation marks .20, %CWS .15, words in complete sentences .13, and long words .09. All variables entering the regression equation were positively related to the dependent variable. On the Iowa Test of Basic Skills, a higher score indicates higher achievement.

Production-Independent and Quality-Based Independent Variables

2005 MEAP writing assessment. Please refer to Table 10 for the statistics for this regression equation. When examining only the quality-based and production-independent set of independent variables, the first variable to enter the regression equation for the 2005 MEAP writing assessment was CWS-IWS, which accounted for 10.2% of the variance. In the second step of the model, correct punctuation was entered, accounting for an additional 2.4% of the variance, for a total of 12.6%. In the third step, %CWS was entered into the equation, accounting for an additional 0.8% of the variance, for a total of 13.4%. In the fourth step of the model, correct capitalization was entered accounting for an additional 1.3% of the variance, for a total of 14.7%. Finally, in the fifth step, CWS-IWS was removed from the model, removing .01% of the variance, for a total of 14.6% of the variance accounted for by the model ($F_{1, 513} = 29.32$, $p < .00$). The Beta coefficients in the final step of the regression equation were small: correct punctuation -.13, %CWS -.21, and correct capitalization -.16. All variables were negatively related to

the dependent variable meaning that as the particular variable increased, MEAP performance also increased.

2003 MEAP writing assessment. Please refer to Table 11 for the statistics for this regression equation. When examining only the quality-based and production-independent set of independent variables, the first variable to enter the regression equation for the 2003 MEAP writing assessment was CWS-IWS, accounting for 15.6% of the variance. In the second and final step of the model, %WdsSent entered the equation, accounting for an additional 3.2% of the variance, for a total of 18.8% ($F_{1, 619} = 71.82$, $p < .00$). The Beta coefficients in the final step of the regression equation were as follows: CWS-IWS had a moderate coefficient of -.34, and %WdsSent had a small coefficient of -.19. Both variables were negatively related to the dependent variable, meaning that as the particular variable increased, MEAP performance also increased.

2005 MEAP ELA assessment. Please refer to Table 12 for the statistics for this regression equation. When examining only the quality-based and production-independent set of independent variables, the first variable to enter the regression equation for the 2005 MEAP ELA assessment was complete sentences, accounting for 12.2% of variance. In the second step, %CWS was entered, accounting for an additional 6.3% of variance, for a total of 18.5%. In the third step, nouns was entered, accounting for an additional 0.8% of variance, for a total of 19.3%. In the fourth and final step, fragments was entered, accounting for an additional 0.7% of variance, for a total of 20.0% of variance accounted for by this model ($F_{1, 512} = 32.04$, $p < .00$). The Beta coefficients in the final step of the regression equation were small: complete sentences -.19, %CWS -.23, nouns -.14, and

Table 5

Summary of stepwise regression analyses for predicting the 2005 MEAP writing assessment with the complete set[a] of independent variables (n = 517).

	Variable	B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CWS	-.02	.00	-.31	-7.51	515	.10	.10
Step 2	CWS	-.01	.00	-.24	-5.11			
	CorrPunct	-.06	.02	-.18	-3.93	514	.13	.02
Step 3	CWS	-.01	.00	-.15	-2.72			
	CorrPunct	-.05	.02	-.17	-3.66			
	%CWS	-.01	.00	-.16	-3.13	513	.14	.02
Step 4	CWS	-.00	.00	-.06	-.99			
	CorrPunct	-.06	.02	-.18	-3.90			
	%CWS	-.01	.00	-.31	-3.67			
	IWS	-.02	.01	-.16	-2.24	512	.15	.01
Step 5	CorrPunct	-.06	.01	-.19	-4.47			
	%CWS	-.01	.00	-.37	-6.24			
	IWS	-.02	.01	-.20	-3.40	513	.15	.00
Step 6	CorrPunct	-.05	.02	-.14	-2.81			
	%CWS	-.01	.00	-.33	-5.22			
	IWS	-.02	.01	-.15	-2.59			
	CorrCap	-.03	.01	-.12	-2.36	512	.16	.01

Table 5 – Continued

Note. CWS = correct word sequences, CorrPunct = correct punctuation marks, %CWS = percentage of correct word sequences, IWS = incorrect word sequences, CorrCap = correct capitalization.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences, words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 6

Summary of stepwise regression analyses for predicting the 2003 MEAP writing assessment with the complete set[a] of independent variables (n = 622).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	WdsSent	-.02	.00	-.40	-10.78	620	.16	.16
Step 2	WdsSent	-.01	.00	-.32	-8.16			
	%CWS	-.01	.00	-.20	-5.03	619	.19	.03
Step 3	WdsSent	-.01	.00	-.26	-5.37			
	%CWS	-.01	.00	-.14	-3.21			
	CWS	-.01	.00	-.13	-2.37	618	.20	.01

Note. WdsSent = Words in complete sentences; %CWS = percentage of correct word sequences; CWS = correct word sequences.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences, words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 7

Summary of stepwise regression analyses for predicting the 2005 MEAP ELA assessment with the complete set[a] of independent variables (n = 517).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	WdsSent	-.02	.00	-.37	-9.09	515	.14	.14
Step 2	WdsSent	-.01	.00	-.29	-6.67			
	%CWS	-.01	.00	-.24	-5.56	514	.19	.05
Step 3	WdsSent	-.01	.00	-.17	-2.87			
	%CWS	-.01	.00	-.23	-5.51			
	TotSent	-.09	.03	-.16	-2.66	513	.20	.01

Note. WdsSent = words in complete sentences; %CWS = percentage of correct word sequences, TotSent = complete sentences.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences, words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 8

Summary of stepwise regression analyses for predicting the 2003 MEAP ELA assessment with the complete set[a] of independent variables (n = 584).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	WdsSent	-.02	.00	-.40	-10.47	582	.16	.16
Step 2	WdsSent	-.01	.00	-.30	-7.43			
	%CWS	-.01	.00	-.26	-6.64	581	.22	.06
Step 3	WdsSent	-.01	.00	-.19	-4.09			
	%CWS	-.01	.00	-.24	-5.97			
	CorrPunct	-.07	.02	-.19	-4.18	580	.24	.02

Note. WdsSent = words in complete sentences; %CWS = percentage of correct word sequences; CorrPunct = correct punctuation marks.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences, words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 9

Summary of stepwise regression analyses for predicting the ITBS reading assessment with the complete set[a] of independent variables (n = 532).

	Variable	B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CorrPunct	4.58	.53	.35	8.59	530	.12	.12
Step 2	CorrPunct	3.78	.55	.29	6.84			
	%CWS	.32	.07	.20	4.63	529	.16	.03
Step 3	CorrPunct	2.72	.66	.21	4.17			
	%CWS	.27	.07	.17	3.86			
	Wds in Sent	.22	.08	.14	2.79	528	.17	.01
Step 4	CorrPunct	2.65	.66	.20	4.03			
	%CWS	.24	.07	.15	3.36			
	WdsSent	.20	.08	.13	2.56			
	Long Words	2.32	1.04	.09	2.22	527	.18	.01

Note. CorrPunct = correct punctuation marks; %CWS = percentage of correct word sequences; WdsSent = words in complete sentences.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences, words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] In order for steps to be added to the formula ΔR^2 had to be stat. significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

fragments .10. All variables, with the exception of fragments, were negatively correlated to the assessment, meaning that as the particular variable increased, MEAP performance also increased. The positive correlation between fragments and MEAP performance indicates that as scores on this variable increases, MEAP performance decreases.

2003 MEAP ELA assessment. Please refer to Table 13 for the statistics for this regression equation. When examining only the quality-based and production-independent set of independent variables, the first variable to enter the regression equation for the 2003 MEAP ELA assessment was CWS-IWS, accounting for a total of 17.3% of variance. In the second step, correct punctuation marks was entered, accounting for an additional 5.0% of variance, for a total of 22.3%. In the third and final step, fragments was entered, accounting for an additional 2.0% of variance. A total of 24.3% of variance was accounted for with this equation ($F_{1, 580} = 62.12$, $p < .00$). The Beta coefficients in the final step of the regression equation were as follows: CWS-IWS had a moderate coefficient of -.31, correct punctuation had a small coefficient of -.21, and fragments had a small coefficient of .15. CWS-IWS and correct punctuation were negatively correlated with the MEAP assessment, meaning that as the particular variable increased, MEAP performance also increased. In contrast, fragments was positively correlated with the assessment, meaning that as this variable increased, MEAP performance decreased.

ITBS reading assessment. Please refer to Table 14 for the statistics for this regression equation. When including only the production independent and quality-based set of independent variables, the first variable to enter the equation for the ITBS Reading assessment was correct punctuation marks, accounting for 12.2% of the variance. The second variable to enter the equation was %CWS, which accounted for an additional

3.4% of the variance, for a total of 15.6%. The third variable to enter the equation included long words, which accounted for an additional 1.0% of the variance, for a total of 16.6% of the variance. The fourth variable to enter the regression equation was verbs, which accounted for an additional 0.7% of the variance, for a total of 17.3% of variance. Finally, in the fifth step, fragments entered the equation, which accounted for 0.9% of the variance, for a total of 18.2% ($F_{1, 526} = 23.36$, $p < .00$). The Beta coefficients in the final step of the regression equation were small: correct punctuation marks .22, %CWS .14, long words .10, verbs .12, and fragments -.10. Correct punctuation, %CWS, long words, and verbs were negatively correlated with the MEAP assessment, meaning that as the particular variable increased, MEAP performance also increased. In contrast, fragments was positively correlated with the assessment, meaning that as this variable increased, MEAP performance decreased. Please refer to Table 15 for a visual summary of the independent variables entering all regression equations for both sets of independent variables. All independent variables that entered one or more of the regression equations are included in the table.

Table 10

Summary of stepwise regression analyses for predicting the 2005 MEAP writing assessment with the quality based/production-independent set[a] of independent variables (n = 517).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CWS-IWS	-.01	.00	-.32	-7.65	515	.10	.10
Step 2	CWS-IWS	-.01	.00	-.24	-5.16			
	CorrPunct	-.06	.01	-.18	-3.77	514	.13	.02
Step 3	CWS-IWS	-.01	.00	-.12	-1.77			
	CorrPunct	-.06	.02	-.18	-3.89			
	%CWS	-.01	.00	-.15	-2.21	513	.13	.01
Step 4	CWS-IWS	-.00	.00	-.04	-.52			
	CorrPunct	-.04	.02	-.13	-2.55			
	%CWS	-.01	.00	-.18	-2.74			
	CorrCap	-.03	.01	-.15	-2.74	512	.15	.01
Step 5	CorrPunct	-.04	.02	-.13	-2.64			
	%CWS	-.01	.00	-.21	-4.80			
	CorrCap	-.04	.01	-.16	-3.22	513	.15	.00

Note. CWS-IWS = correct minus incorrect word sequences; CorrPunct = correct punctuation; %CWS = percentage of correct word sequences; CorrCap = correct capitalization.

[a] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete

Table 10 – Continued

sentences, fragments, simple sentences, not-simple sentences, correct minus incorrect word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 11

Summary of stepwise regression analyses for predicting the 2003 MEAP writing assessment with the quality-based/production-independent set[a] of independent variables (n = 622).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CWS-IWS	-.02	.00	-.40	-10.72	620	.16	.16
Step 2	CWS-IWS	-.02	.00	-.34	-8.81			
	%WdsSent	-.00	.00	-.19	-4.95	619	.19	.03

Note. CWS-IWS = correct minus incorrect word sequences; %WdsSent = percentage of words in sentences.

[a] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete sentences, fragments, simple sentences, not-simple sentences, correct minus incorrect word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 12

Summary of stepwise regression analyses for predicting the 2005 MEAP ELA assessment with the quality based/production-independent set[a] of independent variables (n = 517).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	TotSent	-.20	.02	-.35	-8.45	515	.12	.12
Step 2	TotSent	-.15	.02	-.27	-6.58			
	%CWS	-.01	.00	-.26	-6.33	514	.19	.06
Step 3	TotSent	-.13	.03	-.24	-5.41			
	%CWS	-.01	.00	-.24	-5.74			
	Nouns	-.02	.01	-.10	-2.23	513	.19	.01
Step 4	TotSent	-.11	.03	-.19	-3.73			
	%CWS	-.01	.00	-.23	-5.47			
	Nouns	-.02	.01	-.14	-2.90			
	Fragments	-.07	.03	.10	2.14	512	.20	.01

Note. TotSent = complete sentences; %CWS = percentage of correct word sequences.

[a] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete sentences, fragments, simple sentences, not-simple sentences, correct minus incorrect word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 13

Summary of stepwise regression analyses for predicting the 2003 MEAP ELA assessment with the quality based and production independent set[a] of independent variables (n = 584).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CWS-IWS	-.02	.00	-.42	-11.03	582	.17	.17
Step 2	CWS-IWS	-.02	.00	-.30	-7.26			
	CorrPunct	-.09	.02	-.25	-6.12	581	.22	.05
Step 3	CWS-IWS	-.02	.00	-.31	-7.64			
	CorrPunct	-.08	.02	-.21	-5.11			
	Fragments	.08	.02	.15	3.94	580	.24	.02

Note. CWS-IWS = correct minus incorrect word sequences; CorrPunct = correct punctuation marks.

[a] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete sentences, fragments, simple sentences, not-simple sentences, correct minus incorrect word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 14

Summary of stepwise regression analyses for predicting the ITBS reading assessment with the quality based and production independent set[a] of independent variables (n = 532).

Variable		B	SE B	β	t	df	R^2	$\Delta R^{2\,bc}$
Step 1	CorrPunct	4.58	.53	.35	8.57	530	.12	.12
Step 2	CorrPunct	3.78	.55	.29	6.84			
	%CWS	.32	.07	.20	4.63	529	.16	.03
Step 3	CorrPunct	3.57	.56	.27	6.43			
	%CWS	.28	.07	.17	3.97			
	Long Words	2.59	1.05	.10	2.47	528	.17	.01
Step 4	CorrPunct	3.16	.59	.24	5.39			
	%CWS	.24	.07	.15	3.35			
	Long Words	2.66	1.04	.11	2.55			
	Verbs	.52	.25	.09	2.12	527	.17	.01
Step 4	CorrPunct	2.82	.60	.22	4.70			
	%CWS	.22	.07	.14	3.08			
	Long Words	2.49	1.04	.10	2.40			
	Verbs	.66	.25	.12	2.61			
	Fragments	-2.21	.93	-.10	-2.37	526	.18	.01

Note. CorrPunct = correct punctuation marks; %CWS = percentage of correct word sequences.

[a] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete

Table 14 – Continued

sentences, fragments, simple sentences, not-simple sentences, correct minus incorrect word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

[b] In order for steps to be added to the formula ΔR^2 had to be statistically significant at $p < .05$.

[c] Table figures differ slightly from narrative text due to rounding error.

Table 15

Summary of independent variables entering the regression equations for each dependent variable.

	Complete Set of IVs[a]					Quality-Based/Production-Independent Set of IVs[b]					
	MEAP 2003		MEAP 2005		ITBS	MEAP 2003		MEAP 2005		ITBS	Total
	Writing	ELA	Writing	ELA		Writing	ELA	Writing	ELA		Entered
CWS	✓					-	-	-	-	-	1
%CWS	✓	✓	✓	✓	✓			✓	✓	✓	8
CorrPunct		✓	✓		✓		✓	✓		✓	6
IWS			✓			-	-	-	-	-	1
CorrCap			✓					✓			2
WdsSent	✓	✓		✓	✓	-	-	-	-	-	4
TotSent				✓					✓		2
Long					✓					✓	2

Table 15 – Continued

	Complete Set of IVs					Quality-Based/Production-Independent Set of IVs					
	MEAP 2003		MEAP 2005		ITBS	MEAP 2003		MEAP 2005		ITBS	Total
	Writing	ELA	Writing	ELA		Writing	ELA	Writing	ELA		Entered
CWS-IWS	-	-	-	-	-	✓	✓				2
%WdsSent	-	-	-	-	-	✓					1
Nouns									✓		1
Fragments							✓		✓	✓	3
Verbs										✓	1

Note. CWS = correct word sequences, %CWS = percentage of correct word sequences, CorrPunct = correct punctuation, IWS = incorrect word sequences, CorrCap = correct capitalization, WdsSent = words in complete sentences, TotSent = complete sentences, Long = long words, CWS-IWS = correct minus incorrect word sequences, %WdsSent = percentage of words in complete sentences.

[a] Complete set of independent variables includes total words written, total words correct, long words, nouns, verbs, adjectives, punctuation marks, correct punctuation marks, correct capitalization, complete sentences, simple sentences, not simple sentences,

Table 15 – Continued

words in complete sentences, fragments, correct word sequences, incorrect word sequences, and percentage of correct word sequences.

[b] Quality-based/production independent set of independent variables includes total words correct, long words, nouns, verbs, adjectives, correct punctuation marks, complete sentences, fragments, simple sentences, not-simple sentences, correct minus in correct word sequences, percentage of correct word sequences, and percentage of words in complete sentences.

CHAPTER IV

DISCUSSION

The purpose of this study was twofold: to replicate elements of a previous study on the predictive validity of CBM in the area of writing to schoolwide criterion measures (i.e., Gansle et al., 2002) and to provide the participating school district with information that would build capacity to improve students' writing skills in a proactive manner. Overall, based on the results of the statistical analyses performed, independent variables which appeared to be most promising predictors for the MEAP writing and ELA assessments and the ITBS reading assessment included *percentage of correct word sequences*, *correct punctuation marks*, and *words in complete sentences*. These variables occurred relatively more frequently than other independent variables within the regression formulas, had high inter-scorer reliability, and high to moderate correlations with dependent measures. In this section, factors influencing data analyses will be discussed, followed by a discussion of independent variables identified through the regression analyses as being stronger predictors of writing performance, and finally, limitations and implications for future research will be outlined.

Discussion of Data Analysis Procedures Utilized

Prior to running regression analyses, correlations among independent variables were examined, and one concern that arose was the high multicollinearity among several variables, which violates one assumption of regression analyses. Highly correlated independent variables may cause potentially useful variables to be excluded from the regression formulas. To correct for this, some independent variables needed to be excluded from the analyses. A second consideration when determining which

independent variables would be included in the regression analyses was that literature in the field supports the theory that production-independent measures may be stronger predictors than traditionally studied measures (e.g., Tindal & Parker, 1989a; Watkinson & Lee, 1992), and this study could potentially further that line of findings. It was determined that the current study would also include production-independent measures in the regression analyses. In light of this, two separate sets of regression analyses were performed with each dependent variable. The first set of analyses was designed to replicate elements of the Gansle et al. (2002) study, and the second, to allow inclusion of production-independent variables and the exclusion of independent variables with high multicollinearity. More specifically, the first set of regression analyses included the entire set of independent variables, with the exception of *percentage of words in complete sentences* and *correct minus incorrect word sequences*. These two variables were excluded from the first set of analyses to replicate portions of the Gansle et al. study, which did not include these variables.

The second set of regression analyses was designed to correct for the high multicollinearity between independent variables, excluding some independent variables (i.e., *total words written*, *punctuation marks*, *words in complete sentences*, *correct word sequences*, *incorrect word sequences*, and *percentage of incorrect word sequences*) and to allow the inclusion of two production-independent variables (i.e., *percentage of correct word sequences*, *percentage of words in complete sentences*) and one accurate-production variable (Jewell & Malecki, 2005) (i.e., *correct minus incorrect word sequences*). Therefore, the second set of regression analyses focused more on independent variables that examined accurately applied writing concepts, rather than the

quantity of words the student wrote. This is evidenced by the inclusion of the production-independent and accurate-production variables, as well as excluding two of the independent variables that were scored solely on quantity of writing (i.e., *TWW*, *punctuation marks*).

It was expected that there would be high correlations among independent variables. Statistically significant correlations were found between all independent variables except three (i.e., *incorrect word sequences* with *TWC*, *correct capitalization*, and *simple sentences*). Correlations between independent variables and dependent variables ranged from .09 to .42. Independent variables with the highest correlation with dependent variables included *words in complete sentences* (ranged from .30 to .40 for all of the dependent measures) and *correct minus incorrect word sequences* (ranged from .32 to .42 for all dependent measures). Independent variables that had the lowest correlations with other independent variables included *fragments* (range from -.01 to -.48), *simple sentences* (range of .07 to .77), and *incorrect word sequences* (range of -.02 to .24). Other variables that had consistently moderate to high correlations across the group of dependent measures included *correct word sequences* (range .29 to .38), *percentage of correct word sequences* (range .29 to .38), *percentage of words in complete sentences* (range .21 to .30), and *correct punctuation marks* (range .28 to .39). Specifically, for the 2003 MEAP writing assessment, correlations were highest with *words in complete sentences* (.40), *correct minus incorrect word sequences* (.40), *correct word sequences* (.38), and *total words correct* (.35). For the 2005 MEAP writing assessment, the variables with the highest correlations included *correct minus incorrect word sequences* (.32),

correct word sequences (.31), *percentage of correct word sequences* (.30), and *words in complete sentences* (.30).

Independent measures appeared to be scored reliably, with interscorer reliability above 80% for all independent variables, with an average agreement of 90.91%. The measures with the highest inter-scorer reliability included *total words written* (99.40), *total words correct* (97.95), *complete sentences* (95.67), and *number of words in complete sentences* (94.99). Measures with the lowest interscorer reliability included *adjectives* (81.92), *simple sentences* (82.73), *not-simple sentences* (83.59), and *incorrect word sequences* (84.89).

Independent Variables Frequently Occurring in Regression Equations

In the present study, in addition to including dependent measures administered in the same year as the writing probe, dependent measures that were taken years following the administration of the writing probe were included. Both the writing probe and the 2003 MEAP assessments were administered during the participants' fourth grade year; whereas, the ITBS was administered in 2005 during the participants' sixth grade year, and the 2005 MEAP assessments were administered during the participants' seventh grade year. Although six different dependent variables were examined, perhaps the most important results are those related to the MEAP writing assessments, as those were the only dependent measures which strictly measured writing skills.

Despite the two and a half year gap between the MEAP assessments, the variables entering the regression equations for the 2003 and 2005 MEAP writing assessments have some important similarities, and independent variables occurring frequently in these regression formulas also occur frequently across regression formulas for the other

dependent variables as well. There were also some noteworthy differences between the regression formulas for the complete set of independent variables as well as for the second set of independent variables. Perhaps most notably, *percentage of correct word sequences* appeared in three of the four regression formulas for the MEAP writing assessments. This, along with the high inter-scorer reliability of *percentage of correct word sequences* (95.37%), high alternate forms reliability (r = .68), and moderate correlations with the dependent measures (range of .29 to -.38), supports previous research which indicates that *percentage of correct word sequences* is a reliable and valid predictor when using CBM in writing (e.g., Espin et al., 1999; Watkinson & Lee, 1992, Tindal & Parker, 1989a). *Percentage of correct word sequences* also appeared frequently in regression formulas for other dependent variables (i.e., with the complete set of independent variables: 2003 and 2005 MEAP ELA assessments, and ITBS Reading assessment; with the second set of independent variables: 2005 MEAP ELA assessment and ITBS Reading assessment).

Words in complete sentences was another variable that frequently entered regression formulas, appearing in four of the five regression formulas when using the complete set of independent variables (*words in complete sentences* was excluded from the second set of analyses). The only dependent variable for which it did not enter into the regression equation was the 2005 MEAP writing assessment. In both MEAP ELA assessments (i.e., 2003 and 2005), and the MEAP 2003 writing assessment, *words in complete sentences* entered the regression equations first, accounting for the largest amount of variance. This variable also demonstrated moderate correlations with dependent measures (range r = -.30 to -.40), and strong alternate form reliability (r = .52).

When examining the production-independent counterpart for this measure (i.e. *percentage of words in complete sentences*), it occurred in only the MEAP 2003 writing assessment when using the second set of independent variables; however, *fragments*, which conceptually may be a complementary variable to *words in complete sentences*, entered three of the five equations with the second set of independent variables (in which *words in complete sentences* was excluded). Despite frequently entering regression equations, *fragments* does not appear to be a strong predictor variable because it has very low correlations with dependent variables (range .09 to .20), low alternate forms reliability (.35) and also had low to moderate correlations with the other independent variables (range -.01 to -.64). Its highest correlations with independent variables were *percentage of words in complete sentences* (-.64), *words in complete sentences* (-.48), and *not simple sentences* (-.42). Because fragments is potentially a competing variable, it may make sense to exclude it from analysis in the future, to determine if more variance could be explained by *words in complete sentences* or *percentage of words in complete sentences.*

Correct punctuation marks entered four out of eight regression formulas for the MEAP assessments, and both of the regression formulas for the ITBS. *Correct punctuation marks* entered both regression formulas for the 2005 MEAP writing assessment, as well as both regression formulas for the 2003 MEAP ELA assessment and both regression formulas for the ITBS Reading assessment. *Correct punctuation marks* had moderate alternate form reliability (r = .42) and correlations with the MEAP writing assessments were also low to moderate (r = -.28 to -.30). These findings may speak to the possibility that *correct punctuation marks* may be a strong predictor of writing

performance. Gansle et al. (2002) also found similar results in regards to correct punctuation marks, as this was a variable cited as having high correlations with dependent measures.

The regression equations for the MEAP ELA assessments are somewhat similar to those for the MEAP writing assessments. These similarities were expected, as the writing assessment contributes to the ELA assessment score. *Percentage of correct word sequences* entered three of the four equations for the MEAP ELA assessments. *Percentage of correct word sequences* was moderately correlated with the MEAP ELA Assessments (r = -.32, -.34). *Correct punctuation marks* entered both regression formulas for the 2003 MEAP ELA assessment, and this measure was also moderately correlated with the MEAP ELA assessments (r = -.30, -.32).

Previous studies have demonstrated that *correct minus incorrect word sequences* was highly correlated with dependent measures (e.g., Espin et al., 2000; Weissenburger & Espin, 2005; Espin et al., 2008). In the current study, this particular variable was only included in the regression analyses conducted with the second set of independent variables, and of the five possible regression formulas, it entered only the 2003 MEAP writing and ELA assessments, despite having moderate correlations with dependent measures (range -.32 to -.42). In both of these regression formulas, it was the first variable to enter the equation, which indicates that it accounted for the highest amount of variance. Given the time frame that the assessments were administered, the 2003 MEAP assessments are thought to be most highly related to the writing probe performance. It is important to note, however, that *correct minus incorrect word sequences* is conceptually similar to *percentage of correct word sequences*, as both measure the degree to which

correct versus incorrect writing is occurring. The correlation between these two measures is high (.94), and when examining both the complete set of independent variables, as well as the second set, which focused mainly on the production-independent variables, the pattern of significant contributions are complementary. Either *percentage of correct word sequences* or *correct minus incorrect word sequences* entered the equations, and at least one of the two entered all ten regression formulas.

To date, this is the only study that has examined the predictive validity of *percentage of words in complete sentences*. In the current study this variable appeared in only one of the five possible regression formulas, despite having moderate correlations with dependent measures (range -.21 to -.30). *Percentage of words in complete sentences* entered the regression formula for the 2003 MEAP Writing Assessment, accounting for only 3% of the variance in the model. It is surprising that *percentage of words in complete sentences* did not appear more frequently in the regression formulas, given how frequently *words in complete sentences* appeared in the regression formulas for the dependent measures run with the complete set of independent variables, but this may be related to the inclusion of *fragments*, as discussed above.

Current Results in Regard to Using Production-Independent Measures with Elementary Aged Children

Previous research in the area of CBM in writing has demonstrated the predictive validity and strong correlational relationships between production-independent measures and both criterion variables and teacher ratings of students' writing skills (e.g., Tindal & Parker, 1989a; Watkinson & Lee, 1992). Additionally, researchers have suggested that TWW and TWC may not be the best predictors of student writing skill level at a secondary level (e.g., Espin et al., 1999; Tindal & Parker; Watkinson & Lee; Espin et al.,

2008) or at an elementary level (Gansle et al., 2002; Weissenburger & Espin, 2005). The current study supports the premise that non-traditional CBM in writing (e.g., production-independent measures such as %CWS) may be stronger predictors of student writing skill than traditional curriculum based measures in writing (e.g., TWW, TWC) for elementary students. It has been suggested that as the students' grade level increases, writing complexity increases, and therefore, more simplistic measures such as TWW and TWC may be weaker predictors than more complex variables that focused on accuracy and quality (i.e., production-independent measures such as percentage of words in complete sentences) (Espin, Scierka et al.; Gansle et al.; Weissenburger & Espin; Espin, Wallace et al.). However, other researchers have suggested that production-independent variables were more strongly related to teacher holistic ratings (Tindal & Parker), teacher grades and standardized assessments (i.e., Jewell & Malecki, 2005) at an elementary as well as a secondary level. The results of the current study suggest that *percentage of correct word sequences*, a production-independent measure, is a strong predictor, when factored into regression formulas for assessments administered at both the elementary and secondary level. Because only one writing probe was administered in spring of 2003, and those scores were factored into regression formulas for assessments administered years later, it is not clear whether the same variables would be identified as strong predictors if the writing probe had been administered in the same year as the dependent variable assessment.

Similarities and Differences Between the Current Study and Gansle et al. (2002)

Although this study was designed to replicate a previous study by Gansle et al. (2002), there were several differences in methodology. Specifically, the computer-

generated variables were excluded, as they were not identified by Gansle et al. as being useful measures, and because re-typing the students' writing probes appears to be an extremely time-consuming step. As a screening tool, time-consuming variables are less desirable to teachers than variables that can be scored more quickly, especially for measures that are administered to all students frequently over the course of a school year. Second, additional variables not utilized by Gansle et al. were included in the regression analyses, including *not simple sentences*, *incorrect word sequences*, *correct minus incorrect word sequences*, and production-independent counterparts of two of the variables (i.e., *percentage of correct word sequences* and *percentage of words in complete sentences*). *Not simple sentences* was an independent variable initially included by Gansle et al., but dropped from the analyses due to low interscorer reliability. Third, teacher rank of student writing was not included as a dependent measure as this information was not available to the researcher in the present study. Finally, because of the breadth of variables included in the total set of independent variables, two separate sets of regression analyses were run, which included a complete set of independent variables and a set of independent variables that focused on quality-based and production-independent measures.

Despite these differences in methodology, when comparing the results obtained in the current study to those obtained by Gansle et al. (2002) there are several important similarities. Specific regression formulas cannot be compared across studies, as dependent measures were different; however, independent variables identified by Gansle et al. as frequently occurring in regression formulas, having high correlations with dependent variables, and having high alternate forms reliability were very similar to those

identified in the present study, including *correct punctuation marks* and *words in complete sentences*. In the present study, the independent variable *percentage of correct word sequences* frequently appeared in regression formulas for the dependent variables, while Gansle et al. found the production-dependent counterpart to occur more frequently (i.e., CWS).

In terms of technical adequacy of measures and procedural differences, Gansle et al. (2002) noted several measures with interscorer reliability between 70% and 76%, including complete sentences, words in complete sentences, simple sentences, and fragments. The current study obtained interscorer agreements above 80% for all independent variables. Measures which had the lowest interscorer reliability in the current study included *adjectives*, *simple sentences*, *not simple sentences*, and *incorrect word sequences*, two of which (i.e., *not simple sentences*, and *incorrect word sequences*) were not included in the Gansle et al. study. This increased level of interscorer agreement may be due to the slightly different training and scoring method utilized in the current study. In this study, fewer probes were scored by each person (with the exception of the first author), and probes were checked for reliability before scorers were given another set. In the Gansle et al. study, scorers re-scored a group of writing probes for specific independent variables that fell below 80% agreement. In the current study, although the scorers did not re-score any probes to improve inter-score reliability, specific variables with interscorer reliability below 80% were discussed and questions clarified before scoring additional probes. This may also explain the larger standard deviation for interscorer agreement for several independent variables in the current study. When examining the alternate forms reliability between the Gansle et al. study and the current

study, the hand-scored variables that were noted as being low for Gansle et al. (i.e., long words and fragments), were the independent variables that had the lowest reliability in the current study. No other variables in the current study had alternate form reliability below .40, which may suggest that they are somewhat stable measures over the course of the two administrations.

Limitations

One possible limitation of the study includes the generalizability of the results. All students who participated in the study attended school within the same school district and were in the same grade. This may limit the generalizability to populations of students which are different in terms of student make-up or grade. Additionally, all of the writing probes were administered within the same time of the school year. Results may have been different if collected during different time periods of the same school year. Finally, only one writing probe administered to students was used to score the independent variables used in the regression equations, even for standardized tests that were administered years later. Independent variables identified by the current regression analyses may have been different had writing probes been administered every year that a standardized assessment was administered, or repeatedly throughout each year.

A second potential limitation is the time necessary to train the scorers to become reliable with the measures. Training took 5 to 10 hours (i.e., some scorers took longer to learn the definitions and to become reliable on the practice probes) before the scorers began scoring probes independently. The current study took a somewhat different approach to training scorers than the Gansle et al. (2002) study. This may be one factor

that led to an increased amount of time needed to train scorers; however, it did result in an increased interscorer reliability on several measures.

Within the statistical analyses, one concern was the high multicollinearity of the independent measures. The high correlations between independent variables may have caused potential variables to be excluded from the regression formulas, despite accounting for a relatively large portion of the variance in the model. Because this study was considered exploratory, and because it was designed to replicate portions of the Gansle et al. (2002) study, no variables were excluded from analyses based on multicollinearity. In part to correct for this concern, the current study included two separate sets of analyses, but despite this, multicollinearity among independent variables was still a concern. Gansle and colleagues have further examined a smaller subset of curriculum based measures identified as being strong predictors for the purpose of further studying the criterion validity of those measures (i.e., TWW, total punctuation marks, correct punctuation marks, words in complete sentences, CWS, and simple sentences) (Gansle et al., 2004); however, additional studies that include a much smaller subset of independent variables are needed. Based on the results of this study, variables recommended for further exploration may include correct punctuation, words in complete sentences, percentage of words in complete sentences, correct word sequences, and percentage of correct word sequences.

Another potential weakness of the current study is the lack of writing-specific dependent variables. Only one of the standardized assessments that served as a dependent variable was a direct writing measure. The other two measures included a reading score, which was based only on students' reading performance, and an ELA measure, which

was a combination of reading and writing. Results of the regression equations may have been more useful if they were predictions of only writing measures. This may further limit the generalizability of the results. Previous researchers have included teacher rank of student writing (e.g., Gansle et al., 2002; Tindal & Parker, 1989a; Espin et al., 2000) as well as grades awarded in ELA classes (e.g., Fewster & MacMillan, 2002) as dependent measures. These types of dependent measures were not available in the current study.

Finally, one limitation of the current study is directly related to the MEAP assessment. There have been no published validity studies on the MEAP assessments. Therefore, there are no data to demonstrate that this assessment is a valid measure of student writing skills. Reports from the State of Michigan Department of Education indicate that the assessment is valid based on data collected regarding the number of students who answer the item correctly, a committee review on possible bias in the test item, an item's ability to discriminate between high and low scoring students, comparison of difficulty of the items to previous assessments, and the number of students achieving score ranges at each level (Michigan Department of Education, 2006). Additionally, data are not collected on the treatment integrity of the assessment, and although the administration guidelines are highly regulated by the state of Michigan, this is an unknown variable. The MEAP is an assessment administered only in the state of Michigan, and this also limits the generalizability of the results, as the correlation of the MEAP with other standardized assessments is unknown.

Implications for Practice and Future Research

There are several important implications for practice that have come out of this study. Variables that appeared to have the strongest predictive validity included

percentage of correct word sequences, *words in complete sentences*, and *correct punctuation marks*. Analyses indicated that *total words written* and *total words correct* do not appear to be the most valid predictors of performance on standardized tests for the students in the current sample at either the elementary or secondary level. These results support the results found by Gansle et al. (2002) while utilizing a much larger sample of students. This was one of only a few studies that have examined the predictive validity of production-independent measures with elementary-aged children. Additionally, this study extends the research base by including dependent measures taken in school years following the administration of the writing probe. Results indicated that the variables identified as being the most valid predictors of future success on standardized measures were very similar to those identified as being valid predictors of success on the standardized test administered in the same year. Additionally, this is the first study that examined the use of *percentage of words in complete sentences* as an independent variable, and results suggested that it was highly correlated with dependent measures despite its infrequent occurrence in regression formulas. Future researchers may want to further examine the utility of this measure.

One area of future research that is needed includes the link between these results and intervention development for teachers. At this time, there is very little information available concerning the use of curriculum-based measures in writing for intervention development. We do not have data that indicate the link between specific independent variables and how to improve the students' performance. Further research could examine the possible intervention recommendations that could be made based on the results of curriculum based measures in writing.

REFERENCES

Afifi, A. A., & Clark, V. (1990). *Computer-aided multi-variate analysis* (2nd ed.). New York: Van Nostrand Reinhold.

CTB/McGraw-Hill. (1985). *California Achievement Test.* Monterey: CA: Author.

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

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

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

Deno, S. L. (2003b). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*, 3-12.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48,* 368-371.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). *Relationships among simple measures of spelling and performance on standardized achievement tests* (Research Report No. 22). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading and Writing Quarterly, 14*, 5-27.

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

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

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

Finn, P. (1977). Computer-aided description of mature word choices in writing. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging.* Urbana, IL: National Council of Teachers of English.

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 alternative measures for curriculum-based measurement in writing. *School Psychology Review, 31*, 477-497.

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

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of

written expression for elementary school students. *School Psychology Review, 35*, 435-450.

Good, R. H., III, & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11*, 326-336.

Good, R. H., III, Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcome driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 699-720). Bethesda, MD: National Association of School Psychologists.

Hammill, D. D., & Larsen, S. C. (1978). *The Test of Written Language.* Austin, TX: PRO-ED.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34*, 9-26.

Howell, J. F., & Memering, D. (1993). *Brief handbook for writers* (3rd ed.). Prentice Hall: Englewood Cliffs, NJ.

Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English, 42,* 732-739.

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

Lee, L., & Canter, S. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*, 315-340.

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

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, T. C. (1978). *Stanford Achievement Test.* New York: Harcourt Brace Jovanovich.

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

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193-203.

Michigan Department of Education Office of Educational Assessment and Accountability, Pearson Educational Measurement, & Harcourt Assessments, Inc. (2006). Michigan educational assessment program technical report 2005-2006. Retrived July 6, 2007 from http://www.michigan.gov/mde/0,1607,7-140-22709_31168_40135-172665--,00.html

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

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). Fort Worth, TX: Holt, Rhinehart, and Winston.

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

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 671-697). Bethesda, MD: National Association of School Psychologists.

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.* New York: The Guilford Press.

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

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

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled LD and low achieving on measures of classroom performance. *Journal of Learning Disabilities, 19*, 545-552.

Tindal, G., & Marston, D. (1990). *Classroom-based assessment: Evaluating instructional outcomes.* Columbus, OH: Merrill.

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

Tindel, G., & Parker, R. (1989b). Development of written retell as a curriculum-based measure in secondary programs. *School Psychology Review, 18*, 328-343.

The University of Iowa College of Education (2007). Description of Iowa tests of basic skills tests, levels 9-14. Retrieved July 6, 2007 from http://www.education.uiowa.edu/itp/index.htm.

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

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

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

Woodcock, R. M. (1987). *Woodcock Reading Mastery Test – Revised.* Circle Pines, MN: American Guidance Corp.

Appendix

HISRB Protocol Approval

WESTERN MICHIGAN UNIVERSITY

Human Subjects Institutional Review Board

Date: September 11, 2006

To: Galen Alessi, Principal Investigator
Christina Terenzi, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair Amy Naugle

Re: HSIRB Project Number: 06-10-22

This letter will serve as confirmation that your research project entitled "Curriculum-based Measures in Writing: A School-based Evaluation of Its Predictive Value—Data Analysis" has been **approved** under the **exempt** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may **only** conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: September 11, 2006

Walwood Hall, Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276