Investigating the Reliability and Validity of High School Curriculum-Based Measures of Writing

by

Milaney J. Levendor

A Research Paper Submitted in Partial Fulfillment of the Requirements for the Education Specialist Degree in School Psychology

Approved: 6 Semester Credits

Thesis Advisor

Committee Member

Committee Member

The Graduate School

University of Wisconsin-Stout

May, 2008
ABSTRACT

The need for research-based, accurate, and efficient curriculum-based measures of high school student writing is increasing every year. This study was designed to identify correct punctuation marks and incorrectly spelled words as two such measures at the high school level. Writing samples were collected from 74 tenth grade students at a west-central Wisconsin high school in the fall and spring of the 2006-2007 school year. The samples were scored for the measures of correct punctuation marks and incorrectly spelled words and the scores were analyzed to determine their alternate-form reliability, criterion-related validity, and developmental validity.

Overall, alternate-form reliability results were statistically significant, although at lower levels than desirable for curriculum-based measures. The measure of correct punctuation marks generated the largest correlation coefficients for both fall and spring samples.
Results of correlation calculations computed between the measures of correct punctuation marks and incorrectly spelled words indicated some statistically significant correlations between correct punctuation marks and the WKCE Language Arts and Writing tests. However, the strength of the correlations was too low for educators to use either curriculum-based measure with confidence in high school settings.

Further, there were no significant gains found between the mean fall and spring scores of correct punctuation marks and incorrectly spelled words, indicating a lack of developmental validity. As such, it appears that neither measure is an indicator of growth over time for tenth grade students.
The Graduate School
University of Wisconsin Stout
Menomonie, WI

Acknowledgments

I would like to thank my family and friends for their encouragement and support; especially Ryan, my mom, and Jean who helped me score, copy, and sort over 250 writing samples. They, along with my dad and my in-laws, also helped keep Wes busy so that I had time to work.

I would like to thank my advisor for all of her assistance, revisions, data analysis, and encouragement along the way; and the use of her graduate students when I thought the scoring would never end. I would also like to thank the other members of my thesis committee, Dr. Michelle Hamilton and Dr. Ed Biggerstaff, for their time and input.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter I: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>3</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>4</td>
</tr>
<tr>
<td>Chapter II: Literature Review</td>
<td>7</td>
</tr>
<tr>
<td>Inter-Rater Reliability</td>
<td>7</td>
</tr>
<tr>
<td>Alternate-Form Reliability</td>
<td>9</td>
</tr>
<tr>
<td>Criterion-Related Validity</td>
<td>11</td>
</tr>
<tr>
<td>Developmental Validity</td>
<td>17</td>
</tr>
<tr>
<td>Summary of Main Findings</td>
<td>19</td>
</tr>
<tr>
<td>Conclusions</td>
<td>21</td>
</tr>
<tr>
<td>Chapter III: Methodology</td>
<td>23</td>
</tr>
<tr>
<td>Subject Selection and Description</td>
<td>23</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>24</td>
</tr>
<tr>
<td>Data Collection Procedures</td>
<td>25</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>26</td>
</tr>
<tr>
<td>Chapter IV: Results</td>
<td>28</td>
</tr>
<tr>
<td>Alternate-Form Reliability</td>
<td>28</td>
</tr>
<tr>
<td>Criterion-Related Validity</td>
<td>29</td>
</tr>
<tr>
<td>Developmental Validity</td>
<td>30</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Sample Demographic Information ................................................................. 44
Table 2: Alternate-form Reliability Results ................................................................. 44
Table 3: Criterion-related Validity Results: Correlations with WKCE
  Language Arts Scores .............................................................................................. 45
Table 4: Criterion-related Validity Results: Correlations with WKCE
  Writing Scores ......................................................................................................... 45
Table 5: Developmental Validity between Fall and Spring Mean Scores .................... 46
Chapter I: Introduction

Writing is seen as an excellent way to express personal perspectives, feelings, opinions, and serve an academic purpose (Gersten & Baker, 2001). At the secondary school level and beyond, writing is an important part of everyday life. From high school papers to college applications, standardized tests, or resumes, writing products are commonly used to express knowledge and convey information. Unfortunately, the evaluation of students' writing is seen by instructors to be the most difficult and tedious part of their job (Zinn, 1998). This difficulty in evaluation combined with the complexity of teaching written expression may partially account for the 18.5 percent of 12th grade students considered to perform below the basic proficiency level in writing on the National Assessments of Educational Proficiency (NAEP) (National Center for Education Statistics, 2007).

The assessment of writing samples can be approached subjectively or objectively (Tindal & Parker, 2001). Subjective scoring involves a broader judgment of quality, while objective scoring involves more reliable features such as tallying the number of total words written or correct word sequences in the written piece. Controversy over which scoring process is most beneficial has led educators to investigate the reliability and validity of such measures in an attempt to identify measures that have technical adequacy (Espin, Scierka, Skare, & Halverson, 1999; Espin, Shin, Deno, Skare, Robinson, & Benner, 2000; Espin, Weissenburger, & Benson, 2004; Tindal & Parker, 2001). Although subjective measures possess more apparent face validity, they often have poor reliability (Espin, Weissenburger, & Benson, 2004). Further, many objective
measures with high reliability often have “seriously doubtful validity” (Tindal & Parker, 2001, p. 170).

The movement of federal special education law toward a Response to Intervention model of education and service delivery has prompted educators to look for reliable and valid measures of academic skill that are easy and quick to administer and score repeatedly throughout the year. Curriculum-based measurement (CBM) is a technically adequate method of keeping track of the achievement of students who are learning basic skills such as reading, mathematics, and written expression (Deno, 1992; Fuchs & Fuchs, 2003). Curriculum-based measurement is designed to focus on the individual, measure small increments of progress, and promote the continuous evaluation and modification of instruction. By nature, such timed measures are easy to administer as well as score, and they are seen as efficient yet accurate in assessing the effects of instruction (Fewster & Macmillan, 2002). Curriculum-based measures (CBMs) are intended to be applied repeatedly over extended time periods, allowing for a direct link between assessment and the development of instructional plans. This method of assessment provides educators the means to identify those students in need of different or additional forms of instruction.

CBMs also are designed to improve instructional programs to ensure that students continue to make academic progress.

Curriculum-based measurement methods in reading and mathematics have been more thoroughly researched than written expression CBMs, likely due to the relative objectivity of the measures (Fuchs & Fuchs, 2002; Fuchs & Fuchs, 1991). Unfortunately, little research on CBMs of written expression exists at the high school level. Previous research at the secondary level indicates that writing samples have been scored by using
such indices as the number of words written, the number of words spelled correctly, the
number of correct word sequences, and the number of correct minus incorrect word
sequences (Espin, Scierka, Skare, & Halverson 1999; Espin, et al., 2000; Espin, et al.,
2005; Weissenburger & Espin, 2005; Weissenburger, 2003). However, no CBM of
writing has been confirmed as technically adequate for high school students. Given that
1) few students demonstrate writing proficiency at the secondary level, 2) curriculum-
based measures of written expression have been developed to address the need to
measure student progress and inform instruction, and 3) educators must know what
curriculum-based measures are technically adequate for use at the high school level, a
need exists to examine the reliability and validity of written expression curriculum-based
measures for middle and high school students.

Purpose of the Study

This study is intended to examine the reliability and validity of curriculum-based
measures in written expression. The primary purpose of this paper is to investigate the
reliability, criterion-related validity, and developmental validity of curriculum-based
measurements in written expression for high school students. The following research
questions were addressed in this study:

1. Are the number of correct punctuation marks and incorrectly spelled words
   reliable indicators of writing proficiency for tenth grade students?
2. Do the number of correct punctuation marks and incorrectly spelled words
   correlate with tenth grade students' scores on the Wisconsin Knowledge and
   Concepts Exam (WKCE)?
3. Are the number of correct punctuation marks and incorrectly spelled words developmentally valid measures of writing proficiency for tenth grade students?

**Definition of Terms**

*Correct Minus Incorrect Word Sequences (CWS-ICWS)* – A computation made by subtracting the total number of incorrect word sequences from the total number of correct word sequences in each portion of a writing sample (Scierka, Weissenburger, & Espin, 2003).

*Correct Word Sequences (CWS)* – Two adjacent, correctly spelled words acceptable within the context of the phrase to a native speaker of the English language (Videen, Deno, & Marston, 1982).

*Correct Punctuation Marks (CPM)* – The number of punctuation marks used appropriately and in the correct location in a sentence (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002).

*Criterion-Related Validity* – The extent to which a person’s score on a criterion measure can be estimated from that person’s test score on a test of unknown validity (Salvia & Ysseldyke, 1991).

*Curriculum-Based Measurement (CBM)* – A set of assessment methods used to keep track of students’ progress in learning basic skills. The assessments are brief, simple tests used on a regular basis (i.e., weekly or monthly) to assess students’ academic performance (Fuchs & Fuchs, 2003).

*Developmental Validity* – The ability of a measure to detect growth in a particular area over time.
Holistic Rating – A score applied to a writing sample after an examiner reads an essay and makes a brief, subjective judgment of quality based on their general impression of the passage (Tindal & Parker, 1989).

Incorrectly Spelled Words – The number of words spelled incorrectly.

Production-Dependent Measures – Measures that assess an individual’s ability to write fluently (i.e., number of words written, number of words written legibly, number of words spelled correctly, and number of correct word sequences) (Tindal & Parker, 1989).

Production-Independent Measures – Measures used to assess the accuracy of writing samples, (i.e., percent of legible words, percent of words spelled correctly, percent of correct word sequences, and mean length of correct word sequences) (Tindal & Parker, 1989).

Reliability – The extent to which it is possible to generalize from an observation of a specific behavior observed at a specific point in time by a specific person to observations conducted on similar behavior, at different times, or by different observers.

Total Words Written (TW) – The number of total words written in a specified amount of time. A “word” is defined as any numeral or sequence of letters with a clear separation from those numerals/sequences adjacent to it (Scierka, et al., 2003). This count also includes misspelled words (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Espin, Scierka, Skare, & Halverson, 1999).

Written Expression – The use of conventional English language to convey thoughts and experiences through a standard graphic symbol system utilizing
acceptable capitalization, punctuation, grammar, and spelling competencies
(Videen, Deno, & Marston, 1982).
Chapter II: Literature Review

This chapter is a review of literature related to curriculum-based measures of written expression. The main focus of this review is to examine the reliability and validity of measures of writing expression including total words written (TW), correct word sequences (CWS), correct minus incorrect word sequences (CWS-ICWS), and correct punctuation marks (CPM). This chapter will focus primarily on inter-rater and alternate-form reliability as well as the criterion-related and developmental validity of each measure. It is important that the measures educators employ to evaluate students provide reliable (consistent or stable) estimates of performance before they are considered to be valid (truthful) indicators (Tindal & Nolet, 1995). The primary focus of this review is to examine the existing literature related to curriculum-based measures of writing proficiency.

Inter-Rater Reliability

In an effort to establish the reliability of total words written and correct word sequences as curriculum-based measures (CBMs) of writing proficiency, Malecki and Jewell (2003) conducted a study intended to measure the reliability of total words written and correct word sequences involving 946 students from first through eighth grade. The researchers obtained three minute writing samples in response to a story starter from each participant in both the fall and the spring of the school year. Inter-rater reliability was then calculated for both measures. Results yielded extremely high reliability coefficients of .98 to .99, thus supporting the reliability of both total words written and correct word sequences.
Similar research by Parker, Tindal, and Hasbrouck (1991) examined 2,160 students in grades 2 to 11 from two school districts. The researchers divided the overall research into studies 1 and 2; the first with grades 2 through 5 and the second with grades 6, 8, and 11. Writing samples were collected for study one in both the fall and spring as a means of comparison, while samples were only available in the spring for study two. All participants in the study completed six minute creative writing samples in response to a story starter. The samples were then analyzed on five different indices including total words written and correct word sequences. Inter-rater reliability for both measures was calculated on representative groups of 30 papers, with .99 for total words written and .87 for correct word sequences.

Videen, Deno, and Marston (1982) conducted a study with the purpose of determining whether correct word sequences is a valid measure of written expression for use in a formative evaluation system for elementary students. Participants were 50 elementary students in grades three through six. Criterion measures included the Test of Written Language (TOWL), the Developmental Sentence Scoring, mean T-unit length, the Checklist of Written Expression, holistic ratings, words spelled correctly and total words written. The students were asked to write for five minutes in response to a story starter or topic sentence, the results of which were scored using each criterion measure as well as correct word sequences. The average percent of agreement between raters counting correct word sequences was 85.8%. This agreement percentage jumped to 90.5% when spelling errors were counted as incorrect word sequences. The researchers also found that the number of correct word sequences increased as the grade level...
increased. However, the number of incorrect word sequences was relatively stable across grade levels.

In a larger study involving 484 fourth, eighth, and tenth grade students (Weissenburger, 2003), inter-reliability was calculated by dividing the lower score by the higher score for total words written, correct word sequences, and correct minus incorrect word sequences for 20 randomly chosen writing samples from each grade level. All inter-scorer agreement percentages were above 89% for all eighth and tenth grade samples. These results indicated strong inter-rater reliabilities for both the middle and high school samples in this across grade level study of curriculum-based measures of writing.

A study completed by Jewell and Malecki (2005) to examine the utility of three categories of curriculum-based measures of written language used a sample of 203 second, fourth, and sixth grade students from an Illinois school district. Eight raters were taught the proper procedure for scoring, and were asked to practice scoring 20 writing samples per grade for grades one through eight (160 samples). Inter-rater reliabilities for total words written and correct word sequences were found to be above .98.

Alternate-Form Reliability

Espin, Shin, Deno, Skare, Robinson, and Benner (2000) conducted a study of 1000 students in grades 6 through 8. Eight students in the sample were labeled learning disabled, while two students were categorized as having emotional and behavioral disorders. The researchers computed alternate-form reliability coefficients on the writing samples. Results indicated that the strongest and most consistent reliability coefficients were found for total words written, correct word sequences, and correct minus incorrect word sequences. All of these curriculum-based measures had reliability coefficients
above .70. Of these, the most reliable indicator of student writing proficiency was found to be correct minus incorrect word sequences.

In a study including eighth and tenth grade students (Weissenburger, 2003), the alternate-form reliability of total words written, correct word sequences, and correct minus incorrect word sequences was investigated. Results indicated alternate-form reliabilities ranging from .55 to .82, but only the ten minute correct minus incorrect word sequences measure yielded a reliability coefficient above .80.

A study conducted by Diercks-Gransee (2005) examined writing samples from tenth grade students from two different schools in Wisconsin. This study analyzed data from 82 students who were asked to complete two writing samples each in response to two story starters. Correlation coefficients calculated between the scores from each of the two story starters resulted in significant correlations at the \( p < .001 \) level for correct punctuation marks and correct word sequences. Correct punctuation marks generated the largest correlation of .76, while correct word sequences had a correlation of .75.

Weissenburger and Espin (2005) examined the alternate-form reliability of narrative writing samples at the fourth, eighth, and tenth grade levels. Their research indicated moderate to strong alternate-form reliability at each grade for total words written (.55 to .84), correct word sequences (.59 to .84), and correct minus incorrect word sequences (.61 to .82). The researchers noted that increased duration led to higher correlations, and the correlations were weaker at higher grades.

Espin, et al.'s 2008 research revealed similar findings in their study of 183 high school students from two large Midwestern schools. Espin and colleagues utilized the measures of total words written, words written correctly, correct word sequences, and
correct minus incorrect word sequences. Their research indicated that alternate-form
reliabilities did not differ by method of scoring, but coefficients increased with the
amount of time students were allowed to write. The strongest reliability coefficients were
found for the seven and ten minute samples.

Criterion-Related Validity

Weissenburger and Espin (2005) also examined the criterion-related validity of
narrative writing prompts with Language Arts Normal Curve Equivalent (NCE) scores
from a statewide test of fourth, eighth, and tenth grade students. The results of this study
indicated that criterion validity was weak at each grade level for total words written (.04
to .45). Slightly stronger criterion-related validity coefficients were found at grades 4 and
8 for correct word sequences (.47 to .62) and correct minus incorrect word sequences (.60
to .68). Weak correlations were found for correct word sequences and correct minus
incorrect word sequences at the tenth grade level (.18 to .36).

Given some studies have found that the criterion-related validity of curriculum­
based measures vary according to the criterion measure, holistic writing scores from the
statewide test were also correlated to the curriculum-based measures in the
Weissenburger and Espin (2005) study. The researchers discovered weak correlation
coefficients for total words written at each grade level (.33 to .48), and moderate
correlations with correct word sequences and correct minus incorrect word sequences at
the fourth and eighth grade levels (.50 to .65). The researchers were also able to
determine that increased duration did not strengthen the validity of any of the three
measures under investigation.
In Espin, et al.'s (2008) study of students in two Midwestern high schools, student curriculum-based measure scores were correlated with statewide standardized test scores. Results indicated that validity coefficients were not affected by time, but were affected by method of scoring. Stronger correlation coefficients were found for the measures of correct word sequences and correct minus incorrect word sequences than for total words written or words written correctly. Similar results have been documented in middle school studies (Espin et al., 2000; Espin et al., 2005; Parker, et al., 1991; Tindal & Parker, 1989).

Scierka, Weissenburger, and Espin (2003) completed two studies with the purpose of investigating the criterion-related validity of a variety of curriculum-based measures of writing proficiency. The first study involved 137 eighth-grade students from both suburban and rural areas of the Midwest who completed two timed writing samples from story starters over a seven day period of time. The writing samples were then scored for total words written, correct word sequences, and correct minus incorrect word sequences. Standard scores on Wisconsin's statewide assessment comprised the criterion measures for comparison. Results of the data analysis demonstrated that correct word sequences and correct minus incorrect word sequences had moderate to large correlations (.47 to .63) with the standard scores on the statewide assessment's language arts test. However, total words written yielded lower correlations (.24 to .34) with the criterion measures for the middle school students.

The second Scierka et al. study (2003) involved 83 eighth-grade students from one junior high school in the rural Midwest. As in the first study, two samples of narrative writing were collected from each student over a ten day period. The writing
samples were scored for total words written, correct word sequences, and correct minus incorrect word sequences. These measures were then correlated with the criterion measure of text coherence (defined as the number of events on the causal chain through the story). Results of the data analysis indicated that total words written, correct word sequences, and correct minus incorrect word sequences had strong correlations with the number of events on the causal chain: from .70 to .97 for total words written, .72 to .92 for correct word sequences, and .66 to .82 for correct minus incorrect word sequences.

Overall, these two studies by Scierka and colleagues (2003) provided consistent support for the criterion-related validity of correct word sequences and correct minus incorrect word sequences as moderate to strong indicators of writing performance when correlated with such criterion measures as a statewide assessment and a measure of text coherence. Results of the Scierka et al. studies, however, were mixed as far as total words written was concerned. One study found low correlations between total words written and statewide assessments, the second study generated strong and significant correlations with text coherence, indicating that the results for total words written varied with the criterion measure applied.

Espin, et al.’s (2000) study of written expression involved writing samples from 1000 students in grades 6 through 8. In this study, criterion variables included classroom teachers’ ratings of student writing proficiency and students’ performance on a district writing test. Results of the study indicated that correct minus incorrect word sequences was the single measure of written expression that related well to teachers’ ratings of middle school writing samples. Not only was this indicator strongly correlated to the teachers’ ratings, it also correlated strongly with the students’ performance on a district
writing test. This study supported the criterion-related validity of correct minus incorrect word sequences. Results of this study also indicated that when measuring growth, correct minus incorrect word sequences avoids the pitfalls of a percentage measure and is a stronger indicator of students' writing performance than correct word sequences alone. A drawback of correct minus incorrect word sequences is that it is a more time-consuming measure than total words written and other measures of writing performance, thus leaving practitioners to choose between efficiency and accuracy. In short, the researchers found correct minus incorrect word sequences to be the most valid predictor of student writing proficiency for the middle school students.

Tindal and Parker (1989) conducted a study of 172 students in special education and remedial programs from grades 6 through 8. Based on correlations with holistic ratings of the students' writing samples, results indicate that total words written did not predict holistic ratings of students' writing. Total words written also could not significantly discriminate between those students in special education and those in remedial programs.

To also examine the criterion-related validity of correct word sequences and correct minus incorrect word sequences at the secondary level, Espin, Scierka, Skare, and Halverson (1999) assessed 1,776 students grades 10 through 12 enrolled in three different levels of English class: basic, regular, and enriched. A group of students with learning disabilities were also included in the sample. Criterion measures included English grades, holistic ratings of student writing, and the language subtest of the California Achievement Test (CAT; CTB/McGraw-Hill, 1985). This study's results revealed significant correlations between the criterion measures and the number of correct word
sequences. Correct minus incorrect word sequences also were shown to distinguish students within the English groups: learning disabilities; basic English; regular English; and enriched English.

Malecki and Jewell’s (2003) research of curriculum-based writing measures with students in first through eighth grades indicated that fluency, accuracy, and accurate fluency writing measures are all significantly related for early elementary and elementary students. At the middle school level, however, no significant correlations were found between the measures of writing accuracy and how much students wrote. Thus, the authors concluded that secondary-level students’ levels of writing fluency and accuracy were not closely related. The study by Malecki and Jewell suggests that students at all grade levels, especially secondary-level students, are most appropriately evaluated in written expression with accuracy measures (production-independent) or the accurate-production measure of correct minus incorrect word sequences. The researchers also suggested that correct word sequences is a very important index of students’ writing skills as well as being a necessary factor in computing other valid indices, such as correct minus incorrect word sequences.

Jewell and Malecki (2005) furthered their research by examining 203 samples of second, fourth and sixth grade students from an Illinois school district. Results of this study suggest that at all grade levels measures of writing accuracy, such as correct word sequences and correct minus incorrect word sequences, may be more strongly related to students’ performance on other types of writing criteria than measures of writing fluency such as total words written.
The Diercks-Gransee (2005) study mentioned earlier examined writing samples from tenth grade students from two different schools in Wisconsin. This study analyzed data from 82 students who were asked to complete two writing samples each in response to two story starters. The criterion-related validity of the measures of incorrect word sequences and correct punctuation marks were evaluated in comparison to holistic scores and scores from the Wisconsin Knowledge and Concepts Exam (WKCE) that is administered to all tenth grade students in Wisconsin. Results indicated that correct punctuation marks and incorrect word sequences were moderately to strongly correlated with holistic scores at correlations of .62 and -.71, respectively. Number of incorrect word sequences correlated moderately (.51) with the standard scores from the WKCE Language Arts subtest, while correct punctuation marks were weakly correlated with the same statewide assessment scores (.28).

A study completed by Espin, De La Paz, Scierka, and Roelofs (2005) examined writing samples from 22 students in the seventh and eighth grades in two suburban middle schools in the southeastern part of the United States. The students were asked to write expository essays and given 35 minutes to write. The essays were scored for correct word sequences, correct minus incorrect word sequences, and the criterion variables of functional elements and quality ratings. Numbers of words written were also scored via computer. Results of the data analysis for this study revealed that correct word sequences and correct minus incorrect word sequences were both strongly correlated with the criterion measures of the number of functional elements and quality ratings of the essays.
Developmental Validity

The study by Espin, De La Paz, Scierka, and Roelofs (2005) also examined writing samples from 22 students in the seventh and eighth grades in two suburban middle schools in the southeastern part of the United States to identify whether correct word sequences, correct minus incorrect word sequences, and number of words written could be used to measure growth in writing proficiency over time. The students were given 35 minutes to write expository essays, completing a minimum of six essays at the beginning of the study. The students were then instructed in composition strategies in writing for four weeks and were asked to write expository essays again within one week of the instruction ending. The essays were scored for correct word sequences, correct minus incorrect word sequences, and numbers of words written (via computer). Results of the data analysis for this study revealed that both correct word sequences and correct minus incorrect word sequences were sensitive to change in student performance over time.

Malecki and Jewell’s (2003) research of curriculum-based writing measures on students in first through eighth grades demonstrated significant differences between students’ fall and spring total words written, correct word sequences, and correct minus incorrect word sequences scores in early elementary, elementary, and middle school. The researchers obtained three-minute writing samples in response to a story starter from each participant in both the fall and the spring of the school year. All three fluency measures were found to measure change over time, with all production-dependent indices demonstrated to be higher in the spring than in the fall. Also, correct minus incorrect
word sequences scores were significantly higher in the spring for all levels, although the developmental validity evidence was not conclusive for the eighth grade students.

Parker, Tindal, and Hasbrouck (1991) examined 2,160 students in grades 2 to 11 from two school districts. Their findings indicated that by examining mean score increases across grade levels and across assessment periods within a school year, both total words written and correct word sequences appeared to measure growth at all grade levels, with one exception. Results showed a temporary regression in the spring total words written and correct word sequences scores from grade 4 to grade 5 before progressing further in the next grade.

Tindal took part in another study with Parker (1991), in which they examined all students from three grade levels attending an elementary school in a small, Pacific Northwestern town. These students participated in a group-administered writing assessment task twice over the course of the year. The authors calculated total words written and correct word sequences for each sample, finding a significant main effect for time. All students improved in total words written and correct word sequences over the course of the year, although the authors cautioned that a change in total words written or correct word sequences scores may not coincide with a change in the overall quality of writing.

As previously mentioned, Espin, Scierka, Skare, and Halverson (1999) assessed 1,776 students from grades 10 through 12 enrolled in three different levels of English classes. Results of the study indicated that the measure of correct word sequences was sensitive to small differences in performance, demonstrating that correct word sequences may be a developmentally valid measure of writing proficiency for high school students.
Summary of Main Findings

The basis of this literature review was to ascertain what is currently known about the reliability and validity of measures of writing proficiency used to assess secondary-level students. A great deal of research has been completed on total words written, correct word sequences, and correct minus incorrect word sequences, but the question remains as to whether the same measures can be applied at the high school level. Preliminary results indicate that these measures are reliable and valid at the elementary and middle school levels, but their technical adequacy appears to diminish as students get older and become more proficient (Parker, et al., 1991; Weissenburger & Espin, 2005). Converging evidences suggests that the measures must gain complexity as students become more educated and their writing becomes more skilled.

Overall, research on the reliability of measures of writing proficiency indicate that total words written and correct word sequences both have very high inter-rater reliability coefficients for both elementary and secondary students (Gansle, et al., 2002; Malecki & Jewell, 2003; Parker, et al., 1991; and Videen, et al., 1982). Alternate-form reliability research (Diercks-Gransee, 2005; Espin, et al., 2000; Espin, et al., 2008; Weissenburger, 2003; and Weissenburger & Espin, 2005) indicates that total words written, correct word sequences, and correct minus incorrect word sequences all have adequate alternate-form reliability coefficients, with the highest coefficients derived from correct minus incorrect word sequences for 10-minute samples.

In examining the research on the criterion-related validity of total words written, correct word sequences, and correct minus incorrect word sequences, it can be said that correct word sequences, correct minus incorrect word sequences, and incorrect word
sequences generally have the strongest correlations with standardized measures of writing proficiency at the secondary level.

Overall, results of research on criterion-related validity of measures of writing proficiency indicate that correct word sequences, incorrect word sequences, and correct minus incorrect word sequences are valid measures of writing proficiency when correlated with criterion measures (Diercks-Gransee, 2005; Espin, et al., 1999; Espin, et al., 2000; Espin, et al., 2005; Espin, et al., 2008; Gansle, et al., 2002; Malecki & Jewell, 2003; Marston, 1989; Scierka, et al., 2003; Weissenburger, 2003; and Weissenburger & Espin, 2005). The measure of total words written, however, shows variability dependent upon the particular criterion measure used for correlation computation (Scierka, et al., 2003) and has been determined to be inappropriate for use with students in higher grade levels (Jewell & Malecki, 2005). Studies have also shown that measures of writing accuracy, such as correct word sequences or correct punctuation marks, may be more strongly related to students’ performance on other types of writing assessments than measures of writing fluency, such as total words written (Jewell & Malecki, 2005).

In examining the research on developmental validity of measures of writing proficiency at the secondary school level, Malecki and Jewell (2003) found that all three fluency measures, total words written, correct word sequences, and correct minus incorrect word sequences were found to measure change over time. The developmental validity of correct word sequences and correct minus incorrect word sequences is also supported by Espin, et al.’s 2005 study of seventh and eighth grade students. Furthermore, the results of Malecki and Jewell’s (2005) research indicate that all production-dependent indices showed students receiving higher scores in the spring than
the fall. The number of correct minus incorrect word sequences was also significantly higher from fall to spring for elementary and middle school samples. Additional support for these findings comes from Parker, Tindal, and Hasbrouck's results (1991), which indicate that by examining mean score increases across grade levels and across assessment periods within a school year, both total words written and correct word sequences appear to be developmentally valid measures.

Overall, with regard to developmental validity, research conducted on total words written, correct word sequences, and correct minus incorrect word sequences indicate that all three measures can measure growth over time in elementary and middle school samples (Espin, et al., 1999; Espin, et al., 2005; Malecki & Jewell, 2003; Parker, et al., 1991; and Tindal & Parker, 1991). However, validity research of the same measures at the high school level has not sufficiently proven total words written, correct word sequences, and correct minus incorrect word sequences to be developmentally valid for this older population.

Conclusions

Curriculum-based measures of writing proficiency are a promising alternative to subjective holistic ratings and lengthy standardized tests. More research, however, is needed to determine the technical adequacy of these measures before they can be used at the high school level.

This literature review examined the relevant literature on the reliability and validity of total words written, correct word sequences, correct minus incorrect word sequences, and correct punctuation marks as measures of writing proficiency. Results reviewed here indicate that the total words written, correct word sequences, and correct
minus incorrect word sequences measures have high inter-rater reliability (Gansle, et al., 2002; Malecki & Jewell, 2003; Parker, et al., 1991; and Videen, et al., 1982). Those same measures, as well as incorrect word sequences and correct punctuation marks, also have high alternate-form reliability (Diercks-Gransee, 2005; Espin, et al., 2000; Espin, et al., 2008; Weissenburger, 2003; and Weissenburger & Espin, 2005). In looking at criterion-related validity, correct word sequences and correct minus incorrect word sequences have been found to be valid measures of writing proficiency when correlated with criterion measures for middle school students (Espin, et al., 1999; Espin, et al., 2000; Espin, et al., 2008; Gansle, et al., 2002; Malecki & Jewell, 2003; Marston, 1989; Scierka, et al., 2003). However, total words written shows variability, depending on the particular criterion measure used for correlational analyses (Scierka et al., 2003). With regard to developmental validity, total words written, correct word sequences, and correct minus incorrect word sequences have been shown to be developmentally valid at the elementary and middle school levels, with further research needed at the secondary level (Espin, et al., 1999; Malecki & Jewell, 2003; Parker, et al., 1991; and Tindal & Parker, 1991).
Chapter III: Methodology

This study was intended to examine the reliability, criterion-related validity, and developmental validity of curriculum-based measurements in written expression for high school students. This study involved an analysis of data collected over the 2006-2007 school year from 74 tenth grade students attending a Wisconsin high school. This chapter will be focused on a description of the sample, and data collection means, methods, and procedures. This chapter will also address the procedures employed to analyze the data.

Subject Selection and Description

The participating school district was a public school district in a city of approximately 63,000 people. The district had a total student enrollment of approximately 11,000; 1,751 of which attended the high school targeted in this study. At the targeted high school, 12% of the student population received special education services; 2% of the population was considered English Language Learners; 21% of the population was identified as Gifted and Talented; and 18% of the student population received subsidized lunches. The population of the identified high school was comprised of 91% White/not Hispanic; 1% Black/not Hispanic; 1% Hispanic; 5% Asian/Pacific; and 1% American Indian/Alaskan students. Students were involved in extra or co-curricular activities at the following rates: Academic 36%; Athletic 34%; and Music 26%. The average ACT composite score at the targeted high school in the 2006-2007 school year was 23.3, which was higher than both the state and national average composite scores.

A total of 130 tenth grade students from the targeted high school participated in this study. Of those, 74 students (57%) produced complete, readable, data sets. The participant sample consisted of students receiving instruction through both general
education and special education classrooms. Students were selected for participation based on their enrollment in tenth grade English classes. Specific classes were identified based on teachers’ responses to the examiner’s request to use instructional time on six days throughout the school year. Three teachers responded that they would be willing to participate, and six of their classes were chosen for data collection. More specific participant demographic data are presented in Table 1.

Instrumentation

Curriculum-based measures

Curriculum-based measures (CBMs) were derived from the students’ narrative responses to story starters. A list of story starters was obtained from www.aimsweb.com, and each class period was provided with a different story starter for each of the six data collection sessions. (For examples of story starters used, see Appendix A.)

The number of correct punctuation marks and incorrectly spelled words were scored for each 10 minute writing sample. Correct punctuation marks were derived by counting the number of punctuation marks used appropriately and in the correct location in a sentence (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). Incorrectly spelled words were derived by counting the number of misspelled words throughout the sample.

Criterion measure

The criterion measure for this study was the students’ scores on the Language Arts and Writing portions of the statewide Wisconsin Knowledge and Concepts Exam (WKCE). This exam is administered to tenth grade students in the fall of every school year. The WKCE is a standardized test designed to measure Wisconsin academic
standards. The test measures achievement in reading, language applications, mathematics, science, and social studies using multiple-choice and short-answer questions. Students also provide a rough draft writing sample.

WKCE Language Arts scaled scores were used for data analysis. Possible scores for tenth grade students in the fall of 2006 ranged from 290 to 630, with four categories ranging from Minimal to Advanced. Writing scores were based on a 9 point scale, with 6 points possible for composition and 3 points possible for conventions.

Data Collection Procedures

The targeted school district was contacted in the spring and fall of 2006 and permission was granted at both the building and district level to administer the curriculum-based writing measures to a select population of tenth grade students. The tenth grade English teachers were then contacted, and three of those teachers volunteered their classes for the study. In all, six English classes were identified for participation in the study. Classes included enriched, general, and special education students. Six data collection sessions were scheduled for each of the six English classes throughout the course of the school year. Two sessions were scheduled in the fall, two in the winter, and two in the spring. Each set of two data collection sessions were separated by at least a day. For the purposes of this study, only the samples collected in the fall and spring data collection sessions were analyzed.

The investigator collected six samples of narrative writing from each class, with only one sample collected per day. The narrative writing samples were composed in response to story starters obtained from www.aimsweb.com. After receiving a story
starter, students were given 30 seconds to think and 10 minutes to write. (Refer to Appendix B for directions.)

At the end of each data collection session, the investigator collected the writing samples. To protect anonymity, student names were removed from the samples and each sample was assigned a code.

Data Analysis

In order to determine whether or not the number of correct punctuation marks and incorrectly spelled words are reliable indicators of writing proficiency for tenth grade students, Pearson product-moment correlation coefficients were calculated between the two CBM scores obtained during each of the two testing sessions in the fall and in the spring. A p value of .001 was used to determine if the correlations were statistically significant.

To answer the second research question of whether or not the number of correct punctuation marks and incorrectly spelled words correlate with tenth grade students’ scores on the Wisconsin Knowledge and Concepts Exam (WKCE), Pearson product-moment correlation coefficients were calculated between the mean scores for each testing session and the students’ scaled scores on the WKCE Language Arts test and Writing test score. A p value of .05 was used to determine whether or not the criterion-related validity coefficients were statistically significant.

In order to determine whether or not the number of correct punctuation marks and incorrectly spelled words are developmentally valid measures of writing proficiency for tenth grade students, repeated measures t test analyses were computed between the mean
scores of the fall sample and the mean scores of the spring sample. As above, a $p$ value of .05 was used to determine if the mean differences were statistically significant.
Chapter IV: Results

The purpose of this study was to investigate the reliability and validity of curriculum-based measures of written expression for high school students. Specifically, the measures of correct punctuation marks and incorrectly spelled words were utilized. Alternate-form reliability was calculated using the two curriculum-based measure scores obtained in each season from each student. Criterion-referenced validity was calculated between mean scores of each fall and spring curriculum-based measure and student scores on the Wisconsin Knowledge and Concepts Exam (WKCE) in the areas of Language Arts and Writing. Finally, developmental validity was determined through comparing the mean scores derived from the fall and spring curriculum-based measures to determine whether growth was indicated.

Alternate-form Reliability

Alternate-form reliability for correct punctuation marks and incorrectly spelled words was calculated using Pearson product-moment correlation coefficients for both the fall and spring measures. Overall results of were statistically significant, although at lower levels than expected for curriculum-based measures. Coefficients ranged from .503 to .601 at the $p < .001$ level. Correct punctuation marks generated the largest correlation coefficients with .56 for the fall samples and .60 for the spring samples. The alternate-form correlations for incorrectly spelled words for fall and spring were .55 and .50, respectively. Refer to Table 2 for more details.
Criterion-related Validity

**WKCE Language Arts Scores**

The criterion-related validity of curriculum-based measures of writing were examined by correlating the measures of correct punctuation marks and incorrectly spelled words with the scaled scores on the Language Arts portion of the WKCE test. Correlations between the curriculum-based measurement scores and the WKCE Language Arts scores were calculated using Pearson product-moment correlation coefficients and are presented in Table 3. Results of these calculations indicate statistically significant relationships between the correct punctuation mark mean scores and the WKCE Language Arts test for both the fall and spring samples \( r = .256 \) and \( .208, p < .05 \). Further, statistically significant negative relationships were identified between WKCE Language Arts scores and the mean scores of incorrectly spelled words for the fall and spring at the \( p < .05 \) level \( r = -.260 \) and \( -.309 \).

**WKCE Writing Scores**

The criterion-related validity of curriculum-based measures of writing were also examined by correlating the mean scores for correct punctuation marks and incorrectly spelled words with the scores on the Writing portion of the WKCE test. Correlations between the curriculum-based measurement scores and the WKCE Writing scores were calculated using Pearson product-moment correlation coefficients and are presented in Table 4. The data indicated that there is not a statistically significant relationship between the mean scores of correct punctuation marks for both the fall and spring samples at the \( p < .05 \) level \( r = .176 \) and \( .198 \). However, statistically significant and negative
relationships were identified between WKCE Writing scores and the mean scores of incorrectly spelled words for the fall and spring at the $p < .05$ level ($r = -.231$ and -.271).

*Developmental Validity*

Developmental validity was assessed by conducting matched-pair t test analyses between the fall and spring mean scores for correct punctuation marks and incorrectly spelled words. There were no significant gains found between the mean fall and spring scores of correct punctuation marks and incorrectly spelled words, indicating a lack of developmental validity. However, a significant difference was evident between the mean scores of correct punctuation marks in the fall and spring. Unfortunately, the data indicated a higher rate of correct punctuation marks in the fall samples than in the spring samples. Overall, neither correct punctuation marks nor incorrect word sequences appear to be a good measure of growth over time for tenth grade students. Refer to table 5 for more details.
Chapter V: Discussion

This purpose of this study was to examine the reliability, criterion-related validity, and developmental validity of curriculum-based measurements in written expression for high school students. Data was collected over the 2006-2007 school year from 74 tenth grade students attending a Wisconsin high school.

Previous research revealed that the measures of total words written, correct word sequences, correct minus incorrect word sequences, and correct punctuation marks have all been found to have high alternate-form reliability (Diercks-Gransee, 2005; Espin, et al., 2000; Espin, et al., 2008; Gansle, et al., 2002; Malecki & Jewell, 2003; Parker, et al., 1991; Videen, et al., 1982; Weissenburger, 2003; and Weissenburger & Espin, 2005). Further, correct word sequences and correct minus incorrect word sequences have been found to be valid measures of writing proficiency when correlated with criterion measures for middle school students (Espin, et al., 1999; Espin, et al., 2000; Espin, et al., 2008; Gansle, et al., 2002; Malecki & Jewell, 2003; Marston, 1989; Scierka, et al., 2003). However, the criterion-related validity of total words written varies, depending on the proficiency level of the students and the particular criterion measure used for correlational analyses (Scierka et al., 2003; Weissenburger & Espin, 2005). Total words written, correct word sequences, and correct minus incorrect word sequences have been shown to be developmentally valid at the elementary and middle school levels, with further research needed at the secondary level (Espin, et al., 1999; Malecki & Jewell, 2003; Parker, et al., 1991; and Tindal & Parker, 1991).

Due to the relative lack of research at the secondary level, this study focused on identifying simple and technically adequate curriculum-based measures that could be
used for screening and progress monitoring purposes with high school students. As such, the number of correct punctuation marks and incorrectly spelled words in 10-minute samples were examined. The data was analyzed to determine the alternate-form reliability, criterion-related validity, and developmental validity coefficients of both measures.

Limitations

This study was somewhat limited in the scope due to the characteristics of the participant sample. Although the sample size was large, it is important to note that the data was collected at one school in west central Wisconsin where the majority of the population was categorized as Caucasian. As a result, the findings of this study may not be applicable to other student populations. Also, within the sample, there was an over-representation of students from the Enriched English classes (51.4%). As such, the advanced proficiency levels of the students may have served to decrease the variance of the curriculum-based measurement and criterion scores.

Another possible limitation of this study is the use of WKCE Writing test scores as criterion measures. Although these scores had a possible range from 0 to 9, the majority of students did not receive a score below 6, leaving little variance within the majority of student scores. As such, more substantive correlation coefficients were difficult to achieve. Further, due to the restricted range of the WKCE writing scores in the district and throughout the state, critics may argue that the WKCE Writing test is not a valid measure of writing proficiency and should not be used as a criterion measure.

The main focus on this study was to examine the reliability and validity of curriculum-based measures of writing at the high school level; however, data was
obtained from tenth grade students only. As a result, the findings of this study may not necessarily apply to 9th, 11th, or 12th grade students.

Conclusions

Alternate-form Reliability

Overall, results of alternate-form reliability calculations were statistically significant, although at lower levels than desired for curriculum-based measures. Coefficients ranged from .503 to .601, and were significant at the $p < .001$ level. Correct punctuation marks generated the largest correlation coefficients with .56 for the fall samples and .60 for spring samples. The correlations for incorrectly spelled words for the fall samples were .55, and the alternate-form correlations were .50 for the spring samples. These findings are similar to those found by Diercks-Gransee in a 2005 study of tenth grade writing samples, in which the strongest and most consistently reliable coefficients were found for incorrect word sequences ($r = .75$) and correct punctuation marks ($r = .76$). Regardless, the alternate-form correlations in this study were lower compared to most others using curriculum-based measures of writing at lower grade levels (i.e., .80 or greater), indicating practitioners will need to use two or more writing samples to establish acceptable reliability if educators choose to use these measures with high school students.

Criterion-Related Validity: WKCE Language Arts Scores

The criterion-related validity of curriculum-based measures of writing was examined by comparing measures of correct punctuation marks and incorrectly spelled words with the scaled scores on the Language Arts portion of the WKCE test. Results indicated statistically significant relations between the mean scores of correct punctuation marks and the WKCE Language Arts test for both the fall and spring samples.
Statistically significant negative relationships were also identified between WKCE Language Arts scores and the mean scores of incorrectly spelled words for the fall and spring. Although statistically significant criterion-related correlation coefficients were found for these measures, the correlation coefficients were not robust or meaningful for most assessment purposes.

Research of curriculum-based measures of writing is still in the beginning stages. Thus, the measures of correct punctuation marks and incorrectly spelled words have not been investigated thoroughly at this point. The Diercks-Gransee (2005) study mentioned earlier is, in fact, the only research reviewed that included the measure of correct punctuation marks for secondary students. Results of the 2005 Diercks-Gransee study and a companion study (Diercks-Gransee, Weissenburger, Johnson, & Christensen, in press) indicated a weak correlation between correct punctuation marks and standard scores from the WKCE Language Arts subtest, but a higher correlation between correct punctuation marks and researcher-applied holistic scores. It appears the criterion-related validity of the number of correct punctuation marks varies according to the criterion measure for high school students.

Criterion-Related Validity: WKCE Writing Scores

The criterion-related validity of curriculum-based measures of writing were also examined by comparing measures of correct punctuation marks and incorrectly spelled words with scaled scores on the Writing portion of the WKCE test. Correlation calculations between the curriculum-based measurement scores and the WKCE Writing scores indicated that there was not a statistically significant correlation between the mean scores of correct punctuation marks for either the fall or spring samples. However,
statistically significant negative relationships were identified between WKCE Writing scores and the mean scores of incorrectly spelled words for the fall and spring. As was true when using the WKCE Language Arts score as a criterion measure, however, all criterion-related validity coefficients for incorrectly spelled words and the WKCE Writing scores were too weak to be considered meaningful. As mentioned previously, the research on the curriculum-based measures of writing identified in this study is scarce, which signifies a need for further research to substantiate to the findings presented here.

**Developmental Validity**

Finally, the developmental validity of correct punctuation marks and incorrectly spelled words was calculated between the mean scores of the fall and spring administrations. There were no significant gains found between the mean fall and spring scores of correct punctuation marks and incorrectly spelled words, indicating a lack of developmental validity. However, a significant difference was evident between the mean scores of correct punctuation marks in the fall and spring; unfortunately, the data indicated a higher rate of correct punctuation marks in the fall samples than in the spring samples. Overall, neither correct punctuation marks nor incorrect word sequences appear to be a good measure of growth over time for tenth grade students. As such, these measures should not be used to monitor the writing progress of students at the high school level.

**Recommendations**

**Implications for Further Research**

It is evident that further research is necessary to identify technically adequate curriculum-based measures of writing for high school students. Since Response to
Intervention models of identifying students with disabilities are becoming more prevalent and the Nation’s Report Card has found 75 percent of our high school students scored below the proficient level in writing (National Center for Education Statistics, 2007), it is imperative now more than ever that an accurate and efficient measure of student writing is identified at the high school level. The measures of correct punctuation marks and incorrectly spelled words show some promise in the results presented here, however, further investigation needs to occur with other criterion indices prior to implementing these measures in a practical setting. Given other measures found to have some promise in earlier investigations (i.e., incorrect word sequences and correct minus incorrect word sequences), further exploration is needed to determine the utility of those measures or to identify new alternative curriculum-based measures for high school students.

Implications for Practice

Although this study did demonstrate some statistically significant results, it would be premature to recommend the use of correct punctuation marks and incorrectly spelled words as reliable and valid measures of high school student writing. Professionals in the field need to increase awareness of the components of proficient writing and convey that knowledge to the research community for further exploration.

Summary

The need for research-based, accurate, and efficient curriculum-based measures of high school student writing is increasing every year. This study was designed to identify correct punctuation marks and incorrectly spelled words as two such measures at the secondary level. Writing samples were collected from 74 tenth grade students at a west-central Wisconsin high school in the fall and spring of the 2006-2007 school year. The
samples were scored for the measures of correct punctuation marks and incorrectly
spelled words and the scores were analyzed to determine their alternate-form reliability,
criterion-related validity, and developmental validity.

Overall, alternate-form reliability results were statistically significant, although at
lower levels than desirable for curriculum-based measures. The measure of correct
punctuation marks generated the largest correlation coefficients for both fall and spring
samples.

Results of correlation calculations computed between the measures of correct
punctuation marks and incorrectly spelled words indicated some statistically significant
correlations between correct punctuation marks and the WKCE Language Arts and
Writing tests. However, the strength of the correlations was too low for educators to use
either curriculum-based measure with confidence in high school settings.

Further, there were no significant gains found between the mean fall and spring
scores of correct punctuation marks and incorrectly spelled words, indicating a lack of
developmental validity. As such, it appears that neither measure is an indicator of growth
over time for tenth grade students.

In conclusion, while results of this study demonstrate some promise for the
reliability and criterion-related validity of correct punctuation marks and incorrectly
spelled words, neither measure was found to accurately predict growth over time. Further
research is necessary in order to substantiate the findings presented here and to identify
alternative curriculum-based measures of writing that are reliable, valid, and easily
administered and scored.
References


Appendix A: Sample Curriculum-Based Writing Prompts

AIMSweb W-CBM Story Starters: Cross-Age Suitable for All Benchmark Grades

1. I couldn’t fall asleep in my tent. I heard this noise outside and...

2. A police officer stopped the driver for speeding and...

3. The noise was getting louder and louder...

4. As the jet flew over the mountains...

5. The rocket ship landed on the moon and...

6. It was a hot, dry day and I had been walking for hours without food or water when...

7. The day was warm and sunny and we were the only ones to see...

8. I waved out the window at my family as...

9. Maybe animals aren’t supposed to talk, but...

10. The other day my father took me with him when...

11. Just as we got into our seats, the...

12. Strange footprints were seen in the sand and...

13. The phone call was mysterious and...

14. Due to the serious nature of this mission, you...

15. The airport control tower was busy when...

16. The roaring snow storm howled and...

17. I was shipwrecked on a deserted island when...

18. The river was raging and it was clear that...

19. He crossed his fingers and opened the box. Suddenly...

20. She woke from a sound sleep when something...
21. As he opened the door the...
22. My heart seemed to stop beating as I opened the door...
23. The day was dark and misty as...
24. If I were to make a TV show, it would be about...
25. The car drove off the road and...
26. The bear attacked my dog and...
27. When the boat went out of control, I...
28. I was in the middle of the lake when...
29. I was riding on an elevator when...
30. My friend fell off the horse and...
31. I was sleeping soundly when...
32. I would like to be invisible because...
33. My friend and I were walking by an old deserted house and...
34. On Tuesday, a big wind came up and...
35. One night in the abandoned graveyard...
36. Yesterday, a monkey climbed through the window at school and...
WRITTEN EXPRESSION (Form A)

Say to the students: I want you to write a story. I am going to read part of a sentence to you first – and then you can write a short story about what will happen.

Before you write, I want you to think about the story. First you will think, then you will write. You will have 30 seconds to think, then you will write for 10 minutes. Do your best work. If you do not know how to spell a word, you should guess.

Keep your pencils down until I tell you to start. Listen, the story begins:

I stepped into a time machine . . .

After 30 seconds say: Listen: I stepped into a time machine . . . You have 10 minutes to write. Keep writing until I tell you to stop. You may begin.

Start the stopwatch immediately.

After 1 minute, give the following prompt: Remember to keep writing until I tell you to stop.

After 10 minutes, say: Stop. Thank you. Put your pencils down.
Table 1

*Sample Demographic Information*

<table>
<thead>
<tr>
<th>Demographics</th>
<th>n (Total = 74)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>40</td>
<td>54.1%</td>
</tr>
<tr>
<td>Females</td>
<td>34</td>
<td>45.9%</td>
</tr>
<tr>
<td>English Class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enriched</td>
<td>38</td>
<td>51.4%</td>
</tr>
<tr>
<td>General</td>
<td>33</td>
<td>44.6%</td>
</tr>
<tr>
<td>Special Education</td>
<td>3</td>
<td>4.1%</td>
</tr>
</tbody>
</table>

Table 2

*Alternate-Form Reliability Results (N = 74)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Punctuation Marks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>23.43</td>
<td>8.49</td>
<td>.563*</td>
</tr>
<tr>
<td>Spring</td>
<td>20.45</td>
<td>9.73</td>
<td>.601*</td>
</tr>
<tr>
<td>Incorrectly Spelled Words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>2.98</td>
<td>3.33</td>
<td>.548*</td>
</tr>
<tr>
<td>Spring</td>
<td>2.64</td>
<td>2.54</td>
<td>.503*</td>
</tr>
</tbody>
</table>

*p < .001*
Table 3

Criterion-related Validity Results: Correlations with WKCE Language Arts Scores

<table>
<thead>
<tr>
<th>Measure</th>
<th>Fall</th>
<th>Spring</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Punctuation Marks</td>
<td>.256</td>
<td>.208</td>
<td>.028</td>
</tr>
<tr>
<td>Incorrectly Spelled Words</td>
<td>-.260</td>
<td>-.309</td>
<td>.025</td>
</tr>
</tbody>
</table>

Table 4

Criterion-related Validity Results: Correlations with WKCE Writing Scores

<table>
<thead>
<tr>
<th>Measure</th>
<th>Fall</th>
<th>Spring</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Punctuation Marks</td>
<td>.176</td>
<td>.198</td>
<td>.136</td>
</tr>
<tr>
<td>Incorrectly Spelled Words</td>
<td>-.231</td>
<td>-.271</td>
<td>.049</td>
</tr>
</tbody>
</table>


Table 5

*Developmental Validity between Fall and Spring Mean Scores*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Fall Mean, SD</th>
<th>Spring Mean, SD</th>
<th>df</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Punctuation Marks</td>
<td>23.43, 8.49</td>
<td>20.45, 9.73</td>
<td>73</td>
<td>3.373</td>
</tr>
<tr>
<td>Incorrectly Spelled Words</td>
<td>2.98, 3.33</td>
<td>2.65, 2.54</td>
<td>73</td>
<td>1.097</td>
</tr>
</tbody>
</table>

*p < .05*