

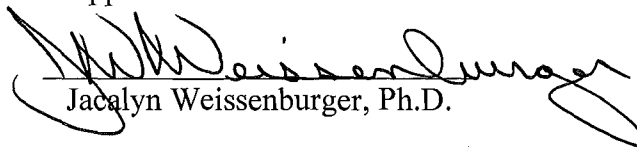
Curriculum-Based Measurement of Written Expression
at the Secondary Level

by

Shanna Dawson

A Research Paper
Submitted in Partial Fulfillment of the
Requirements for the
Master of Science Degree
In
School Psychology

Approved: 2 Semester Credits



Jacalyn Weissenburger, Ph.D.

The Graduate School
University of Wisconsin-Stout

May, 2009

The Graduate School
University of Wisconsin-Stout
Menomonie, WI

Author: Dawson, Shanna L.
Title: *Curriculum-based Measurement of Written Expression at the Secondary Level*
Graduate Degree: M.S.Ed. in School Psychology
Research Adviser: Jacalyn Weissenburger, Ph.D.
Month/Year: May, 2009
Number of Pages: 37
Style Manual Used: American Psychological Association, 5th edition

ABSTRACT

A literature review of all research conducted on CBM of written expression at the secondary level was completed. Findings indicate that CWS and CWS-ICWS have the best criterion-related validity for this population, and these measures can be used with accuracy for screening purposes. Results also indicate that seven minute writing samples meet reliability and validity standards, and seven minutes may be the best administration time for CBM purposes, but more research needs to be completed. Further, findings are very limited regarding the use of CBM measures of written expression with students receiving special education. Further research is needed to examine CBM measures of written expression at the secondary level to determine their technical adequacy for students receiving special education.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
Chapter I: Introduction.....	1
<i>Statement of the Purpose</i>	4
<i>Research Questions</i>	5
<i>Assumptions of the Study</i>	5
<i>Limitations of the Study</i>	5
<i>Definition of Terms</i>	6
Chapter II: Literature Review	9
<i>Criterion-Related Validity of CBMs for Students in General Education</i>	9
<i>Criterion-Related Validity of CBMs for Students in Special Education</i>	20
<i>Technical Adequacy of Administration Time for CBM in Written Expression</i>	21
<i>Discriminate Validity of CBM Measures in Written Expression</i>	27
Chapter III: Summary and Discussion.....	30
<i>Noteworthy Results</i>	30
<i>Limitations of This Literature Review</i>	32
<i>Implications for Future Research</i>	32
<i>Implications for Practice</i>	33
<i>Summary</i>	33
References.....	35

Chapter I: Introduction

Curriculum-based measurement (CBM) is an assessment tool used in the educational system to assess if students are achieving academic competence in reading, writing, spelling, and mathematics (Hosp, Hosp, & Howell, 2007). CBM employs short, simple, standardized measures to quickly screen students for adequate academic performance. CBM is used to monitor and track students' academic progress within these basic skill areas and screen for students who are at risk for future failure. CBM is unique because it can be utilized in any school to monitor the overall academic progress of students regardless of the specific curriculum being used by educators in the classroom.

CBM was first created in the late 1970s at the University of Minnesota Institute for Research on Learning Disabilities by Deno and colleagues for use by special education teachers (Deno, 1985). The objective of their research was to develop an easy and efficient way for special education teachers to assess the effectiveness of their instruction. Deno and colleagues determined that monitoring their students' academic gains through CBM was effective. By assessing the effectiveness of instruction through monitoring gains of students, special education teachers were able to receive immediate feedback on whether their instruction was working for each child. If gains were not visible, it would signal the teacher to change the method of teaching so progress could be made.

Since its inception, CBM has been researched, validated, and expanded to be used in the general education system. Curriculum-based measures (CBMs) can be utilized in the education system in four primary ways: screening/benchmarking, progress-monitoring, diagnostic decisions, and outcome decisions (Hosp et al., 2007). CBM is

primarily used for screening/benchmarking purposes to determine if students are at risk for future failure, and for progress-monitoring purposes to ensure students are making sufficient progress towards academic goals. Diagnostic decisions, in which CBM are used to create an alternative instructional plan when a significant problem arises with a student, and outcome decisions, which verify an educational program's effectiveness, are other uses of CBM, but these uses are secondary to its first two functions.

CBM is different than many other methods for measuring academic performance because it employs criterion-referenced measures instead of norm-referenced measures. While norm-referenced measures simply compare how a student performs to others, criterion-referenced measures are used to determine a student's proficiency at a task by determining if the student meets or will reach a specific level of performance over time. The benchmarks are pre-determined, and the level of performance can be monitored because the student is compared only to the specific benchmark. A benchmark level of performance has been determined at each grade level. The level of performance is a criterion-based score; therefore, more than 50% of students can meet the requirement (Hosp et al., 2007). Furthermore, curriculum-based measures were designed to be sensitive enough to measure minor academic performance gains, thus students are able to be measured frequently to determine if gains and goals are obtained.

In our current education system, CBM is ideal for response to intervention (RTI) models of service delivery. RTI is a multi-level model aimed to maximize student achievement by utilizing early prevention and intervention; therefore, a goal of RTI is to identify students early who are at risk for future academic failure. RTI does not identify a specific system to use to monitor academic achievement, but the assessment system

needs to be reliable, valid, and able to monitor small gains. CBM is an excellent complement to RTI because it is able to meet its screening and progress monitoring needs. When a student does not meet a certain CBM benchmark, they are monitored more closely for academic progress. If academic growth is not visible during the subsequent CBM administrations, educators are able to identify possible reasons for the lack of growth and implement various changes to the instruction or curriculum accordingly. Thus, CBM is an effective way to meet the goals set forth by RTI.

The ability to write clearly and effectively is an important skill in today's society. Writing proficiently is fundamental for a student to convey information and express thoughts and ideas on paper. The importance of having adequate skills in written expression is evident by its inclusion in compulsory state tests, college entrance exams, and The National Report Card (Scierka, Weissenburger, & Espin, 2003). In 41 states, students are required to complete testing which includes a writing component, and 20 of these states have a high school graduation requirement of passing a test in writing (Espin et al., 2008). Furthermore, effective, well-developed writing skills are an important aspect of not only quality academic work, but also of effective later job-related performance (Kellogg & Raulerson, 2007). It is important to ensure students develop effective writing skills in school; however, statistics gathered from the National Assessment for Educational progress showed that 14-26% of all United States students are unable to write at the basic level (cited in Dierkes-Gransee, 2006). Identifying these students is crucial as they will need to pass academic requirements and develop need proficiencies to be successful in the future.

Currently, most research on CBM of written expression has been completed at the elementary and middle school levels. Multiple studies have established strong criterion-related validity correlations between CBMs of written expression and criterion measures for elementary school students and moderately strong correlations for middle school students (Scierka, Weissenburger, & Espin, 2003). The few, but growing, number of studies concerning CBMs for secondary students have revealed the need for more research to determine accurate measures of written expression (Leverson, 2008). Scoring methods, such as Total Words Written (TW) and Correct Word Sequences (CWS), have been found to be effective measures for young students, but these methods have been found to be technically inadequate for measuring written expression of secondary students (Hartquist, 2006). There is a clear need in the field of CBMs to be able to screen and progress monitor students in general and special education in the secondary setting.

Statement of Purpose

Most research to date on curriculum-based measurement has focused on elementary and middle school students. Studies have validated various methods for measuring writing proficiency of elementary and middle school students, including indicators such as number of correct writing sequences (CWS), incorrect writing sequences (ICWS), and total words written (TW) to assess writing samples. These methods of evaluating CBMs of written expression have been used to identify students struggling with writing and to measure their progress in developing writing skills. However, the little research completed has shown little validity in utilizing the same CBMs of written expression to identify and measure student progress at the secondary level.

The purpose of this literature review is to examine the technical adequacy of different methods of curriculum-based measures in written expression for secondary students in special education. Currently, little research exists on CBMs of writing at the secondary level. In this review, research on the criterion-related validity of different CBM scoring methods for secondary students will be explored.

Research Questions

The following research questions are addressed in this literature review:

1. What is the criterion-related validity of different scoring methods used for CBMs of written expression with secondary students in special and general education?
2. What is known about how the administration time affects the technical adequacy of CBMs of written expression secondary students in special and general education?
3. Do CBM measures of writing differentiate the performance of secondary students receiving special education from students in general education?

Assumptions

All published literature pertaining to secondary CBM is available to the author and it covers the most important literature to date.

Limitations

This paper is only a literature review. As such, it is not contributing new knowledge to the field. Also, this paper is limited to the investigation of CBMs of written expression at the secondary level. Thus, it is not an exhaustive literature review across grade levels.

Definition of Terms

The following terms are commonly used when discussing CBM, and will be used throughout this paper.

Accurate-production measures - A group classification of CBM written expression scoring measures that depends on the amount the students writes accurately. Production-dependent measures include CWS and CWS-ICWS (Espin et al., 2000; Jewell & Malecki, 2005).

Adjectives (ADJ) - A method of scoring in which the total number of correctly used adjectives in a writing sample are counted. Predicate adjectives (e.g., bright, big, blue) and proper adjectives (e.g., Mexican, Shakespearian, Australian) are counted towards the total number of correctly used adjectives, but possessive adjectives (e.g., their, his, her), articles (e.g., the, a, an), and demonstrative adjectives (e.g., these, that, those) are not (Diercks-Gransee, Weissenburger, Johnson, & Christensen, 2008).

Adverbs (ADV) - A method of scoring a writing sample in which the total number of correctly used adverbs, or words that modify a word in a sentence, are counted. Adverbs indicate when, where how, how much, and to what extent in a sentence (e.g., suddenly, lots, tomorrow, often, above, slowly) (Diercks-Gransee et al., 2008).

Correct Punctuation Marks (CPM) - A method of scoring a writing sample in which the total number of correctly used punctuation marks are counted (Diercks-Gransee et al., 2008; Leverson, 2008).

Correct Word Sequences (CWS) - A method of scoring a writing sample which indicates two correctly spelled words are adjacent to each other and are contextually acceptable to a native English language speaker. A correct word sequence is scored as a correct word

sequence when two adjacent words are grammatically and syntactically correct (Leverson, 2008; Weissenburger & Espin, 2005).

Correct Word Sequences minus Incorrect Word Sequences (CWS-ICWS) - A method of scoring a writing sample in which the total number of incorrect word sequences are subtracted from the total number of correct word sequences (Weissenburger & Espin, 2005).

Curriculum-based measurement (CBM) - An assessment tool used in the educational system to evaluate whether students are achieving academic competence in reading, writing, spelling, and mathematics. CBM functions primarily as a quick screening/benchmarking tool for academic performance and as a system for progress monitoring (Hosp, Hosp, & Howell, 2007).

Incorrect Word Sequences (ICWS) - Two adjacent words in which either one or both words are incorrectly spelled or not contextually acceptable to a native English language speaker (Espin & Tindal, 1998).

Production-dependent measures - A group classification of CBM written expression scoring measures which means the measure depends on the amount the students writes because the score of the measure varies with the length of the writing sample.

Production-dependent measures include: TWW, WSC, CWS, and words written legibly (Espin, Weissenburger, & Benson, 2004; Parker, Tindal, & Hasbrouck, 1991a, 1991b).

Production-independent measures - A group classification of CBM written expression scoring measures that depends on the amount the students writes because the score of the measure does not vary with the length of the writing sample. Production-dependent measures include: percentage of WSC, percentage of CWS, percentage of legible words,

and mean length of CWS (Espin, Weissenburger, & Benson, 2004; Parker, Tindal, & Hasbrouck, 1991a, 1991b).

Total Words Written (TWW) – The total number of words written in a writing sample. A word is defined as any sequence of letters or numerals clearly separated from an adjacent sequence or numeral. TWW includes all identifiable words whether spelled correctly or not (Weissenburger & Espin, 2005).

Words Spelled Correctly (WSC) – The total number of words spelled correctly in a writing sample (Parker, Tindal, & Hasbrouck, 1991a). Is the same measure as words written correctly (WWC).

Words Written Correctly (WWC) – The total number of words written correctly in a writing sample (Espin et al., 2008). WWC is the same measure as words spelled correctly (WSC).

Chapter II: Literature Review

Introduction

The criterion-related validity of different curriculum-based measurement (CBM) scoring methods to assess written expression for secondary students in special and general education will first be discussed. This literature review will then examine what is known about how administration time affects the technical adequacy of CBMs of written expression for secondary students in special and general education. Finally, the discriminate validity of CBM measures of writing between students receiving special education from students in general education will be explored.

Criterion-Related Validity of CBMs for Students in General Education

Most studies to date concerning the criterion-related validity of CBM scoring methods for written expression have been completed using elementary and middle school students. Relatively few studies have focused on the technical adequacy of CBM written expression methods at the high school level. The first major research to examine written expression CBMs for students at the secondary level was completed by Parker, Tindal, and Hasbrouck (1991a, 1991b). Participants of the first study (1991a) included students in 2nd, 5th, 6th, 8th, and 11th grade, and participants in the second study (1991b) included middle school students in grades 6-8. In both studies, students were given a story starter, 30 seconds to think, and then 6 minutes to write their responses. Writing samples were scored using both production-dependent measures and production-independent measures. Production-dependent measures, defined by how much the student wrote, were TWW, WSC, CWS, and words written legibly. Production-independent measures, those free

from how much the student wrote, were percentage of WSC, percentage of CWS, percentage of legible words, and mean length of CWS.

Findings from both studies (Parker et al., 1991a, 1991b) indicated that production-independent variables generally were more strongly correlated with the criterion measures than the production-dependent scores. Because of the differences in correlations across grade levels, an analysis of the data was completed to see if there was a difference in the ability to discriminate students across grades using production-dependent variables or production-independent variables. The analysis revealed that the percentage of CWS was able to discriminate students in lower grade levels and students with lower scores better than CWS. However, CWS was able to discriminate between students in different grade levels and between students with different levels of proficiency better than percentage of CWS.

Through their studies, Parker et al. developed the basis for future research on CBM of written expression at the secondary level (1991a, 1991b). The correlational scores between the various measures and grade levels suggested that simpler measures of written performance, such as TWW and WSC, were adequate, reliable, and valid at the elementary level; however, these measures were not found to be valid at the secondary level. Parker et al. suggested that production-independent measures, such as percentage of CWS, was a more valid indicator than production-dependent measures of individual performance in written expression. The authors noted the need for more research to determine valid measures of writing at the secondary level.

Although Parker et al. (1991a, 1991b) found production-independent measures to be better indicators of written expression performance, using production-independent

CBM measures of written expression is problematic as they do not adequately fit the requirements of a CBM (Espin, Weissenburger, & Benson, 2004). Percentage measures could stay consistent over time or vary greatly, even though the amount of writing could increase, decrease, or stay the same. Although percentage measures may be adequate for identifying low-performing students, because of its variability, it would not be a reliable way to monitor progress over time, which is a crucial, fundamental requirement of CBM.

The majority of the subsequent research on CBMs focused on identifying technically adequate production-dependent measures to identify and to monitor progress (Espin, Weissenburger, & Benson, 2004). One of the first studies to explicitly focus on CBM written expression at the high school level was conducted by Espin et al. (1999). Espin and colleagues collected writing samples and data from 147 students in 10th grade. All students were randomly chosen from four English class placements: Learning Disabled, Basic, Regular, and Enriched English. Samples were scored using TWW, WSC, CWS, characters per word, total sentences written, and mean length of CWS strings. Criterion measures included the Language Arts subtest from the California Achievement Test (CAT), English class placement, English class semester grades, and holistic ratings of the writing sample.

In the Espin et al. (1999) study, criterion correlations indicated that CWS, the mean length of CWS, total number of sentences written, and number of characters per word had the strongest correlations, although they were in the low to moderate range ($r = .34 - .45; p < .001$). The researchers conducted a regression analysis and found that using a combination of measures predicted writing proficiency better than one measure alone. A moderately high correlation was found with the measure combination of mean length

of CWS, number of characters per word, and total number of sentences written with the criterion measure CAT Language Arts subtest ($R = .62$). The results from this study indicated that using only one measure was inadequate to assess writing proficiency at the 10th grade level, and a combination of measures proved to be a better predictor of writing proficiency at the high school level. However, it was noted that using a combination of measures, although a better predictor, may be too complicated for use as a CBM measure. Also, further research would be necessary to determine how to calculate and accurately graph combination scores over time for progress monitoring purposes (Espin et al., 2000).

Armed with the knowledge that CWS produced only moderately strong correlations, Espin et al. (2000) investigated a new, more complex measuring method for CBMs of written expression. In Espin and colleagues' study, they included CWS-ICWS, an accurate-production measure, as a method for scoring samples of written expression. They hypothesized this novel scoring method may more accurately measure written expression; and, as the authors noted, this method would not have the same progress-monitoring difficulty as production-independent measures. In the Espin et al. study, a total of 112 students in 7th and 8th grade were asked to produce four writing samples: two descriptive and two story writing samples. Students composed their writing samples by typing on a computer with editing features for a total of 5 minutes, with an identification mark at the end of 3 minutes to be used for scoring purposes. Teacher ratings and scores obtained from a district writing test were used as the criterion measures.

In the Espin et al. study (2000), CWS-ICWS produced the strongest correlations with the teachers' ratings and the district writing test scores. Moderately strong

correlations were found with CWS-ICWS for the 3 and 5 minute samples of both the story and descriptive writing samples. Statistical analysis also revealed that the reliability and validity of both the descriptive and story writing samples, across administration times, were very similar. The results of their study suggested CWS-ICWS may be a better indicator of written expression achievement for secondary students than simpler forms of measurement, and different styles of writing may be used for CBMs of written expression. A potential limitation identified by the authors was the use of computers for collecting students' writing samples because of potential differences in performance based on their word processing skills.

A longitudinal study (Fewster & Macmillan, 2002) was then conducted to determine the predictive validity of written expression and oral reading fluency CBM of 6th and 7th graders using teacher-awarded grades earned their 8th, 9th, and 10th grade years as the criterion measures. Four hundred sixty-five 6th and 7th graders in the 1995-1996 school year were given CBM oral reading fluency probes and a 3 minute written expression probe. The reading CBM was scored by the number of words read correctly (WRC), and writing was scored using the number of words spelled correctly (WSC). For three subsequent years, teacher-awarded grades in both English and Social Studies classes were recorded for the students' 8th, 9th, and 10th grade years. Data analysis of the teacher-awarded grades verified a high degree of consistency for within-course correlations and high internal consistency for all grades and courses, thus indicating the teacher-awarded grades had a strong degree of validity and would be an acceptable criterion measure. A positive correlation between initial reading and writing CBM scores was found to be significant at the $p < .005$ level for both English and Social Studies

grades and over time; however, these correlations were small. Further, WRC was more highly correlated than WSC at all grade levels, and both measures were more highly correlated with English grades than the Social Studies grades. This study suggested that using school-based evidence as criteria to establish the validity of a CBM measure was sufficient for future use.

The criterion-related validity of three different CBM measures of written expression for secondary students was examined by Scierka, Weissenburger, and Espin (2003). The study obtained writing samples from 137 eighth grade students in the Midwest and used the scoring measures TWW, CWS, and CWS-ICWS. The Wisconsin Knowledge and Concept Examinations (WKCE), a statewide assessment of achievement, was used as the criterion-referenced measure. Normal curve equivalent (NCE) scores from the WKCE Language Arts subtest were used as the criterion score. Writing samples were scored at the 3 minute, 5 minute, and 10 minute portions of the writing session. The results indicated that only the CWS and CWS-ICWS correlations were statistically significant at the $p < .001$ level for CBMs of written expression at the 8th grade level, and both had moderate to strong correlations (.47 - .63). Concerning sample length, no reliable differences were found between shorter and longer samples. Overall, CWS-ICWS was found to have statistically stronger criterion-related correlation coefficients than CWS, suggesting that more complex CBM scoring measures of written expression were better indicators of writing achievement for students in 8th grade.

A comparison study conducted by Weissenburger and Espin (2005) investigated the alternative-form reliability and criterion-related validity of writing CBM across grade levels. In their study, the same three CBM measures, TWW, CWS, and CWS-ICWS,

were used, and writing samples were scored at the 3, 5, and 10 minute portions of the writing session. The NCE scores from the Language Arts subtest of the WKCE and holistic writing scores from a direct writing assessment were used as the criterion scores. The Language Arts subtest was administered to all 4th, 8th, and 10th graders, but due to a pilot test, the Writing Assessment was only given to 4th and 8th graders that year. Thus, no 10th grade holistic scores were available for use as a criterion score.

When correlating scores with the WKCE Language Arts subtest, the researchers found that the criterion-related validity was stronger for CWS and CWS-ICWS than TWW across all grades (Weissenburger & Espin, 2005). TWW was found to be statistically significant only at the 4th grade level. CWS was found to be a valid indicator of performance at the 4th and 8th grade level (.59 & .50; $p < .001$), but not at the 10th grade level (.18 - .26; $p < .001$). CWS-ICWS was found to be statistically significant at all grade levels; however, at the 10th grade level, the criterion-related correlation coefficients were in the very low range (.29 - .36; $p < .001$), while the 4th and 8th grade CWS-ICWS scores produced correlations in the moderate to strong range. When correlating the 4th and 8th grade scores with the WKCE Writing Assessment, most CBM scoring methods produced correlations in the moderate to strong range. Generally, for all CBM measures, sample duration did not affect the correlation coefficients, as little differences were seen. The results of this study contributed to reference that the technical adequacy of CBM measures in written expression decreased as the age of the writer increased. However, it was noted that the trend was less prominent for the more complex CBM measure of CWS-ICWS. This study's findings indicated that CWS-ICWS was the strongest predictor

of written expression performance, CWS was the second strongest predictor, and TWW was the weakest performance predictor across all grade levels.

A study which focused on the 7th and 8th grade population also substantiated the validity of the CWS and CWS-ICWS scoring methods (Espin, La Paz, Scierka, & Roelofs, 2005). In this study, a different genre of writing was explored as the basis for writing samples: expository writing. Expository writing was chosen because students were required to pass a state's competency tests in which they needed to write an expository essay. A total of 22 students participated in the study. Six students were identified as having a learning disability with difficulties in written expression, 6 students had low written expression achievement, 6 had average written expression achievement, and 4 had high written expression achievement as measured by their scores on the written expression subtest of the *Wechsler Individual Achievement Test*. The 6 students in the learning disability group had been previously identified as having a learning disability through the district's criteria.

The Espin et al. (2005) research used a pre-test, treatment, post-test design, and 35 minute writing samples were collected each week for a total of 6 weeks for all student groups. After collecting the pre-test writing samples the first week, an intensive 4 week long expository instruction was implemented, and then a writing sample was taken on the last week. Samples were scored for CWS, CWS-ICWS, and TWW. Criterion scores were quality ratings and functional elements. Functional elements were quantified by counting the number of units in the essay, such as premises, reasons, elaborations, and conclusions. Quality ratings based on the holistic rating system were applied by trained raters who were unaware of the purpose of the study. Before the essays were given to the raters for

scoring, the writing samples were typed. The writing samples were also corrected for spelling, capitalization, and punctuation. The researchers justified correcting the essays by indicating these factors would particularly penalize the students with learning disabilities' writing samples.

Espin et al. (2005) found that CWS and CWS-ICWS had strong correlations with the two criterion measures, functional elements and quality ratings ($r = .66 - .83$). Surprisingly, TWW was also found to have moderately strong to strong correlations with both criterion measures ($r = .58 - .90$). This finding was particularly unusual given the amount of previous research concerning secondary level written expression that found very low correlations with this measure. However, over time, CWS and CWS-ICWS were much better indicators of student performance.

Espin et al.'s (2005) conclusion about CWS and CWS-ICWS supported prior research that these measures may be valid and reliable indicators of the 7th and 8th grade students' writing achievement by using different criterion measures, functional elements and quality ratings to analyze its validity and ability to measure change in performance over time. The Espin et al. study also added to the CBM field of research by finding expository writing was an alternative method for assessing written expression proficiency. Lastly, the unusual finding of TWW having a moderately strong to strong correlation suggested further research should be completed with this measure. The researchers did recognize this effect may have been due to having an exceptionally long administration time (i.e., 35 minutes).

Other more recent studies supported the idea that scoring longer writing samples using CWS-ICWS has produced the highest reliability and validity coefficients for older

students (Espin et al., 2008; Hartquist, 2006). Espin et al. (2008) found that, for 10th grade students, CWS-ICWS was more reliable and valid than TWW, WWC, and CWS. This study used holistic scores from two state assessments of written expression, Minnesota Basic Standards Test (MBST) and Minnesota Comprehensive Assessments (MCA) as the criterion variables. Correlation coefficients indicated CWS-ICWS was statistically significant at the $p < .001$ at 7 minutes ($r = .58$) and 10 minutes ($r = .60$). CWS was statistically significant, but had lower coefficients than CWS-ICWS ($r = .46 - .48$). In the Hartquist (2008) study, CWS-ICWS was also found to be the most reliable and valid measure for 10th grade students when correlated against the Language Arts score from the WKCE ($r = .62$). Again, TWW did not produce statistically significant results, and CWS was statistically significant, but the correlation was smaller than CWS-ICWS ($r = .52$).

Although CWS-ICWS has emerged as a potentially valid and reliable measure of secondary students' written expression abilities, much more research must be completed to determine if a more technically adequate measure can be found for use at the secondary level. Further, more investigation is needed to determine what measure is the most useful measure for progress monitoring at the secondary level (McMaster & Espin, 2007). Recently, alternative methods for scoring secondary written expression samples have been explored. These studies have used alternative measures including correct punctuation marks (CPM), adjectives (ADJ), and adverbs (ADV).

Diercks-Gransee (2006) investigated the criterion-related validity of CPM, ADJ, and ADV of 85 tenth grade students using 10 minute writing samples. The criterion measures used in the study were the NCE scores from the WKCE Language Arts test and

holistic ratings. Statistical analysis revealed that both ADJ and ADV did not produce significant correlation coefficients. CPM did reveal a significant correlation at the $p < .012$ level; however, the correlation was very low ($r = .275$).

Using similar criterion measures, Levenson (2008) examined the validity of CPM to measure tenth grade writing samples in both the fall and spring of a school year. NCE scores from the WKCE Language Arts test were used as the criterion measure. Results from Levenson's study were similar to Diercks-Gransee's (2006) findings. Correlation coefficients between CPM and WKCE scores indicated that statistically significant relationships existed at the $p < .05$ level for both the fall and spring samples, but the relationships were low ($r = .256$ and $.208$).

Diercks-Gransee, Weissenburger, Johnson, and Christensen (2008) conducted a reanalysis of Diercks-Gransee (2006) data, and they investigated CPM, ADJ, and ADV from 82 data sets. Again, the criterion measures were the NCE scores from the WKCE Language Arts test and holistic ratings. The ADJ and ADV correlation results were consistent with prior findings. That is, they were not statistically significant. When correlated with the WKCE scores, CPM had similar coefficients as prior studies ($r = .28$, $p < .05$); however, the correlation between CPM and holistic ratings was moderately strong ($r = .62$, $p < .001$). Based on their findings, Diercks-Gransee et al. (2008) suggested ADJ and ADV should not be used as measures for scoring secondary written expression samples, and further research was needed to determine CPM's effectiveness in identifying students with learning disabilities.

Criterion-Related Validity of CBMs for Students in Special Education

To date, little research has been completed that specifically examined the technical adequacy of CBMs of written expression scoring methods for secondary students in special education (Hartquist, 2006). Most studies have grouped all students, both general and special education, together for statistical analysis. Only one study by Hartquist (2006) specifically examined the criterion-related validity of written expression measures for secondary students in special education.

Hartquist (2006) investigated the technical adequacy of CBM measures in written expression of students in 4th, 8th, and 10th grade. A total of 484 writing samples from students in 4th, 8th, and 10th grade were used in the study, with 55 of those students identified as receiving special education services. Of the 55 students receiving special education, 44 were eligible for special education services on the basis of having a learning disability. Writing samples were collected by using two forms of a story starter, and students were given 30 seconds to think, and then 10 minutes to write. Criterion measures used in this study were the NCE scores from the WKCE Language Arts test and holistic ratings of the writing sample scored by an experienced high school English teacher. The scoring methods included TWW, CWS, and CWS-ICWS.

In the Hartquist (2006) study, the criterion-related validity of the three CBM measures in written expression was calculated using the scores of students receiving special education. Findings from this study indicated the correlations between the WKCE Language Arts test score and CWS-ICWS were significant at the $p < .05$ level only for 4th and 10th graders in special education, with the 10th graders correlation at .62. CWS was also found to be statistically significant for students receiving special education in 10th

grade ($r = .52$). No significant findings were found at the 8th grade level. This result is dissimilar from other research which has demonstrated the technical adequacy of CWS and CWS-ICWS of students in 8th grade. However, the author noted that the majority of prior research analyzed the criterion-related validity of all students and did not directly analyze just students in special education. The author suggested more research with larger samples of students receiving special education was needed.

Technical Adequacy of Administration Time for CBM in Written Expression

Most research concerning the technical adequacy of administration time for CBM in written expression has been completed at the primary level to date, and little research has been completed at the secondary level (Weissenburger & Espin, 2005). At the elementary level, CBM research indicates that 3 minute writing samples are valid and reliable indicators of writing proficiency (Watkinson & Lee, 1992). However, current findings with a focus on students at the secondary level suggests students need to write for longer periods of time than 3 minutes to obtain valid and reliable evidence of writing performance (Watkinson & Lee, 1992; Weissenburger & Espin, 2005). When Parker and colleagues researched the criterion-related validity of CBM across grade levels using a 6 minute writing time, they found a decrease in correlations as students increased with age (Parker et al., 1991a). Subsequent studies have revealed that as students get older, the validity of CBM measures in written expression decrease (Espin et al., 2000; Espin et al., 2005). Therefore, it has been hypothesized that as students become older, more complex methods of scoring and longer samples of writing may be needed (Espin et al., 2000; Espin et al., 2005; Weissenburger & Espin, 2005). Many of the studies that investigated validity of various CBM written expression scoring methods have used 10 minute

administration times to collect their data and analyze the methods' criterion-related validity (Diercks-Gransee, 2006; Diercks-Gransee et al., 2008; Hartquist, 2006; Leverson, 2008). A few studies, presented here, have examined the validity and reliability of different written expression administration times to determine what length of sample duration is the most technically adequate.

Research conducted by Scierka, Weissenburger, and Espin (2003) examined the criterion-related validity of different CBM measures in written expression of secondary students using different lengths of administration time. In their study, two writing samples from 137 eighth grade students were collected during a seven day period. Two different story starters were used, and order effects were controlled by counter-balancing the story starters. The procedures for data collection were students were told their story starter, given 30 seconds to think, and then 10 minutes to write. During the 10 minutes, students were instructed to make a slash mark on their paper at the 3 and 5 minute time marks. Samples were scored using TWW, CWS, and CWS-ICWS, and scored at the 3, 5, and 10 minute mark. NCE scores from the WKCE Language Arts test were used at the criterion measure.

The criterion-related coefficients were calculated for the 3, 5 and 10 minute sample lengths, and the differences were analyzed (Scierka, Weissenburger, & Espin, 2003). For each of the three measures, no significant differences were found according to sample length. Therefore, this study's findings suggested that for 8th grade students' writing samples, the criterion-related validity of the scoring measures TWW, CWS, and CWS-ICWS did not change with an increase in sample duration.

In a second study conducted by the same authors, two samples from 83 eighth graders were collected over a ten-day period (Scierka, Weissenburger, & Espin, 2003). After the students received their story starter, they were given 30 seconds to think and then asked to write for 30 minutes. At the 5, 10, and 15 minute time intervals, students were directed to make slash marks. The same story starters from the first study were used and the order was counterbalanced. Writing samples were scored using TWW, CWS, and CWS-ICWS for all sample lengths. Text coherence was used as the criterion measure. Text coherence was calculated by counting the number of causally connected events in the writing sample.

The analysis showed that as the length of writing time increased, the correlation coefficients increased (Scierka, Weissenburger, & Espin, 2003). However, differences in the correlations between the 5, 10, and 15 minute writing samples were not significant, and each measure only differed by a maximum of .06 between the 5 minute and 15 minute sample. The greatest increase in correlation was seen in the 30 minute samples, and only between the 15 and 30 minute sample, a statistically significant difference in the correlations was found. For 30 minute samples using a $p < .001$ significance level, the correlation between text coherence and TWW was .97, CWS was .92, and CWS-ICWS was .82. Although there was not a significant difference between the 5, 10, and 15 minute samples, the correlation between text coherence and all three scoring methods indicated TWW, CWS, and CWS-ICWS were moderate to moderately strong predictors of text coherence, as correlations ranged from .66 to .78 ($p < .001$). Overall, these studies found that that 3, 5, 10, and 15 minute samples produced similar correlations within each measure; however, the 30 minute sample produced the strongest correlations.

Another study looked at the technical adequacy of 35 minute writing samples (Espin et al., 2005). In this study, all 22 seventh and eighth graders were statistically pre-grouped into writing ability level based on achievement test scores and whether there was a diagnosis of a learning disability. Students were asked to write for 35 minutes for each sample. Between the pre- and post-test, all students participated in a 4 week long, 4 days per week writing instruction class. Writing samples were scored using CWS and CWS-ICWS, and the criterion measures used were holistic ratings and the number of functional essay elements. The number of functional essay elements was counted by identifying the number of units in the writing sample which supported the development of the essay.

Espin et al. (2005) study's results indicated that both measures, CWS and CWS-ICWS, showed a significant difference between pre- and post-test, and both demonstrated a correlation with both criterion measures ($r = .66 - .83, p < .01$) using a 35 minute administration time. To expand their statistical analysis, the researchers calculated the magnitude of correlations of using the CWS and CWS-ICWS scoring methods using only the first 50 words. This was completed to see if not using a specific administration time, but using a certain number of words, would have any technical adequacy. All subjects, except the students with learning disabilities, showed little change from pre-test to post-test. Students with learning disabilities did show a marked increase; however, the increase did not reach statistical significance.

Concerning the administration time in this study, the researchers commented that the administration time in this study was probably too long for CBM purposes (Espin et al. 2005). Although the researchers did find significant findings using the 35 minute administration time, this timeframe would be too lengthy for progress monitoring

purposes. They noted that one of the fundamental notions of CBM is to be quick, and this administration time would probably not meet the efficiency standard. Not only is the 35 minute administration time lengthy, the time it takes to score long writing samples is also time-consuming for educators.

Two other researchers investigated the technical adequacy of different CBM measures in written expression across grade levels and analyzed the effect of administration time on its technical adequacy (Weissenburger & Espin, 2005). Specifically addressed in their study were the alternate-form reliability and criterion-related validity of the measures. The researchers questioned if there were differences between measures across grade levels and if it was influenced by sample duration or scoring procedure. Two different writing prompts, “I stepped into a time machine” (Form A), and “It was a dark and stormy night” (Form B) were used. Two samples were collected from a total of 484 students in 4th, 8th, and 10th grade over a two week period, and the order of story-starters were counterbalanced to control for order effects. The NCE scores from all WKCE subject areas were used as the criterion measures, although the main criterion-related validity score was Language Arts. Scoring methods included TW, CWS, and CWS-ICWS, and samples were scored at the 3, 5, and 10 minute intervals of the writing sample.

Findings indicated there was an increase in the alternative-form reliability coefficients with an increase in sample duration across all grade levels and scoring methods (Weissenburger & Espin, 2005). For all grade levels, the alternative-form correlation coefficients for all three scoring methods were significant at the $p < .001$ level (.55 to .84). The alternative-form reliability between Form A and B increased with age

and had the strongest correlations at the 8th and 10th grade levels. Therefore, for all scoring methods at the 4th, 8th, and 10th grade levels, an increase in sample duration increased the strength of the alternative-form correlation, especially at the 8th and 10th grade levels.

Results of criterion-related validity analyses revealed that the correlation coefficients with the WKCE Language Arts subtest scores were generally stable across sample duration (Weissenburger & Espin, 2005). Across all three grades and scoring methods, only small differences in the strength of the correlations were seen with an increase in sample duration. For secondary students' samples (i.e., 8th and 10th grade), a small increase in criterion-related validity coefficients occurred with an increase in sample duration, but the increase was not meaningful. Therefore, the Weissenburger and Espin study found that although the criterion-related validity coefficients did not increase with longer sample duration, the alternative-form reliability did increase when longer samples were written by secondary level students.

One other study examined the effect of administration time on the validity and reliability of secondary students' writing samples (Espin et al., 2008). Two writing samples were collected from 183 tenth grade students, and writing samples were scored at 3, 5, 7, and 10 minutes. Samples were scored using TWW, WWC, CWS, and CWS-ICWS. The criterion-related measures used were the students scores obtained from the MBST and MCA writing tests.

In the Espin et al. (2008) study, statistical analysis showed that alternative-form reliability progressively increased with an increase in administration time from 3 to 10 minutes for all scoring procedures. The strongest reliability coefficient was found for 7

and 10 minute sample lengths, and the differences in reliability for these sample lengths were very small. Criterion-related validity correlations indicated very little change in the validity coefficients with an increase in sample duration. The measure with the strongest coefficients for secondary students, CWS-ICWS, varied between .56 and .60 ($p < .001$) on the 3, 5, 7, and 10 minute time samples. Based on these findings, the researchers recommended a 7 minute administration time if the writing CBM is collected for screening purposes three times per year. However, for more frequent use, such as progress monitoring purposes, the researchers suggested that educators can use the more efficient 5 minute writing samples.

Discriminate Validity of CBM Measures in Written Expression

Limited research has examined the technical adequacy of CBM measures for students receiving special education (Hartquist, 2006). Furthermore, an insufficient amount of research has been conducted to determine if the current production-independent CBM measures of written expression, such as CWS and CWS-ICWS, are technically adequate to differentiate between the performance of students with writing disabilities or in special education from students who receive general education students. The few studies which have examined the discriminate validity of written expression CBM measures for secondary students receiving special and general education will be discussed next.

Espin et al. (2005) looked at 35 minute CBM writing samples of seventh and eighth graders with varying levels of writing proficiency. Results indicated there was a difference between students with learning disabilities and low, average, and high achieving writers. Students were pre-grouped into learning disability, low achieving,

average achieving, and high achieving writers for statistical analysis. When using CWS and CWS-ICWS measures with only the first 50 words written, only students with learning disabilities showed a marked increase in performance over time. When using scores derived from the 35 minute writing sample, all groups demonstrated similar gains after a 4 week intensive writing intervention. Although students in the learning disability population produced CWS and CWS-ICWS scores substantially lower on both the pre- and post-test probes, the students with disabilities made equivalent progress when compared with their peers.

A study by Diercks-Gransee and colleagues (2008) indicated students in the general education sample produced significantly more CPMs than students in the learning disability sample. Furthermore, students in the learning disability sample produced more ICWS than those in the general education sample. The predictive accuracy of CPM and ICWS was calculated in this study, and results showed there using these measures for differentiating students in general education from students in special education. Findings revealed that a 20th percentile cutoff score accurately identified 7 out of the 8 students diagnosed with a learning disability using ICWS. When using CPM, 6 out of the 8 students diagnosed with a learning disability were accurately identified.

A recent study by Hartquist (2006) investigated the technical adequacy and discriminate validity of CBM measures in written expression of 484 students in 4th, 8th, and 10th grade. Of the 484 students, 55 were identified as receiving special education, and 44 were categorized as having a learning disability. Ten minute writing samples were collected and scored using TWW, CWS, and CWS-ICWS methods. The criterion

measures were the NCE scores from the WKCE Language Arts test and holistic ratings of the writing sample scored by an experienced high school English teacher.

In the Hartquist (2006) study, a one-way analysis of variance (ANOVA) was used to find out if these measures were able to differentiate students receiving special education from students receiving general education. Results revealed that across grade levels, both CWS and CWS-ICWS were able to differentiate students receiving general education from students receiving special education ($p < .01$). Furthermore, CWS-ICWS was able to differentiate students across grade levels at the $p < .001$ significance level. The author stated these findings suggest that, because this study did directly measure the criterion-related validity of CBM measures in written expression for students receiving special education, both CWS and CWS-ICWS were technically adequate for students in special education and could be used to help identify students with potential learning disabilities.

Chapter III: Summary and Discussion

Introduction

In this section, a summary of the literature review, highlighting all noteworthy results, will be presented. The limitations of the literature review, implications for future research, and implications for practice will then be discussed.

Noteworthy Results

Throughout the literature available on CBM in written expression at the secondary level, it was clear that CWS and CWS-ICWS have emerged as the currently most valid and reliable indicators of written expression (Espin et al., 2000, 2005, 2008; Hartquist, 2006; Scierka, Weissenburger, & Espin, 2003; Weissenburger & Espin, 2005). In the majority of the research conducted, CWS and CWS-ICWS have been found to have moderately strong criterion-related correlations, with CWS-ICWS having slightly stronger correlations than CWS. Furthermore, generally speaking, somewhat stronger correlations have been found for both measures at the 8th grade level than at the 10th grade level.

Most research to date on the technical adequacy of administration time for CBM in written expression has been completed at the primary level, and relatively little research has been completed at the secondary level (Weissenburger & Espin, 2005). Recent findings with a focus on students at the secondary level suggested that to obtain valid and reliable evidence of writing performance, students needed to write for longer periods of time than 3 minutes (Watkinson & Lee, 1992; Weissenburger & Espin, 2005). While many studies of secondary students' CBM in written expression have used 10 minute administration times (Diercks-Gransee, 2006; Diercks-Gransee et al., 2008;

Hartquist, 2006; Leverson, 2008), a few studies have examined how administration time affects the reliability and validity using 3, 5, 6, 7, 10, 15, 30, and 35 minute sample lengths. Overall, an increase in reliability has been found with longer samples. However, the validity has not been found to be significantly affected by increasing the administration time between 3 to 30 minutes using CWS and CWS-ICWS, but it has been suggested that because 7 minute writing samples meet reliability standards, this length may be best for screening purposes (Espin et al., 2008; Scierka, Weissenburger, & Espin, 2003; Weissenburger & Espin, 2005).

Very little research has been conducted to examine the validity of CBM measures in written expression with students receiving special education services. In one study, students with known learning disabilities had clearly lower CWS and CWS-ICWS scores than students without learning disabilities; however, all students made equal progress over the four week intervention (Espin et al., 2005). When Hartquist (2006) investigated the discriminate validity of CBM of TWW, CWS, and CWS-ICWS, she found that CWS and CWS-ICWS were able to differentiate between special education and general education students. Lastly, the study by Diercks-Gransee and colleagues showed that CPM and ICWS had good predictive accuracy using a 20th percentile cutoff score (Diercks-Gransee et al., 2008). This limited research showed that CPM, CWS, ICWS, and CWS-ICWS have potential promise for screening and discriminate validity purposes; however, much more research needs to be completed to discern the effectiveness and adequacy of the measures. Further, more research is needed to determine if these measures are valid for monitoring the progress of secondary-level students in special education.

Limitations of This Literature Review

A major limitation of this literature review was that it was limited to the investigation of CBMs of writing at the secondary level. Thus, it was not an exhaustive literature review of CBMs of writing across grade levels. Furthermore, this research paper only contained information derived through literature review. As such, it has not contributed new knowledge to the field.

Implications for Future Research

It is clear that there is a need for more research to be completed to determine a technically adequate method to measure and progress monitor students' developing writing skills at the secondary level. Current research supported the need for CBM measures of writing to be more complex at the secondary level due to the increase in students' age and increase in writing skills (Weissenburger & Espin, 2005). More research should be completed specifically using CWS-ICWS because it has been shown to have the strongest technical adequacy and most promise. Furthermore, more research is needed using a larger sample of students in special education. Research with a larger number of students with learning disabilities could determine which measures are technically adequate for screening and progress monitoring purposes when working with this population.

Additionally, research to date indicated that the length of the writing sample needs to be longer than 3 minutes as children become older to establish the reliability of the measures. Current research suggests that 7 minute writing samples may be reliable; however, more research is needed to pinpoint the ideal length of time for older students to write in order to obtain valid and reliable indicators of the student's level of academic

achievement. Also, once the length of time has been established for screening purposes, additional research needs be conducted to determine the ideal length of time needed for progress monitoring purposes.

Implications for Practice

Currently, technically adequate CBM measures of writing have been identified for use in elementary and middle schools. These methods of measuring students' academic competence are currently used in the education field today. However, technically adequate methods for measuring written expression at the secondary level, specifically high school, have not been identified and confirmed through replication studies. Although CBM should be used extensively in the elementary schools to assess and monitor students' academic achievement to address the needs of struggling students early, more research needs to be conducted to determine technically adequate curriculum-based measures of written expression for students at the secondary level before it can be used with assurance.

Summary

A literature review of all research conducted on CBM of written expression at the secondary level was completed. Findings indicate that CWS and CWS-ICWS have the best criterion-related validity for this population, and the data indicate these measures can be used with accuracy for screening purposes. Results also indicate that seven minute writing samples meet reliability and validity standards, and seven minutes may be the best administration time for CBM purposes. However, more research needs to be completed. Further, findings are very limited regarding the use of CBM measures of written expression with students receiving special education. Further research is needed

to examine CBM measures of written expression at the secondary level to determine their technical adequacy for students receiving special education services.

References

- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Diercks-Gransee, B. (2006). *Curriculum-based measurement in written expression at the high school level*. Unpublished field study, University of Wisconsin-Stout, Menomonie. Retrieved January 16, 2009, from:
www.uwstout.edu/lib/thesis/2006/2006diercks-granseeb.pdf
- Diercks-Gransee, B., Weissenburger, J. W., Johnson, C. L., & Christensen, P. (2008). Curriculum-based measures of writing for high school students. *Remedial and Special Education OnlineFirst*, 1-12. Advanced online publication. Retrieved January 24, 2009. doi: 10.1177/0741932508324398
- Espin, C. A., De La Paz, S., Scierka, B. J., & Roelofs, L. (2005). The relationship between curriculum-based measures in written expression and quality and completeness of expository writing for middle school students. *Journal of Special Education, 38*(4), 208- 217.
- Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary students. *Reading and Writing Quarterly, 15*(1), 5-27.
- Espin, C. A., Shinn, 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*(3), 140-154.

- Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214-253). New York: Guilford Press.
- Espin, C. A., 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. *Council for Exceptional Children, 74*, 174-193.
- Espin, C. A., Weissenburger, J. W., & Benson, B. J. (2004). Assessing the writing performance of students in special education. *Exceptionality, 12*, 55-66.
- 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.
- Hartquist, S. M. (2006). *The validity of curriculum-based measurement of written expression for students in special education*. Unpublished field study, University of Wisconsin-Stout, Menomonie. Retrieved January 16, 2009, from: www.uwstout.edu/lib/thesis/2006/2006hartquists.pdf
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34*, 27-44.
- Kellogg, R. T., & Raulerson, B. A. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review, 14*, 237-242.

- Levenson, M. J. (2008). *Investigating the reliability and validity of high school curriculum-based measures of writing*. Unpublished field study, University of Wisconsin-Stout, Menomonie. Retrieved January 24, 2009, from: www.uwstout.edu/lib/thesis/2006/2006levesronm.pdf
- Parker, R. I., Tindal, G., & Hasbrouck, J. (1991a). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality, 2*, 1-17.
- Parker, R. I., Tindal, G., & Hasbrouck, J. (1991b). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children, 58*(1), 61-73.
- Scierka, B. J., Weissenburger, J. W., & Espin, C. A. (2003, April). *Criterion-related validity of curriculum-based measures of writing at the secondary level*. Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.
- Tindal, G., & Parker, R. I. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education, 23*(2), 169-183.
- Watkinson, J. T., & Lee, S. W. (1992). Curriculum-based measures of written expression for learning-disabled and non-disabled 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.