The Relationship Between Curriculum-Based Measures in Written Expression and Quality and Completeness of Expository Writing for Middle School Students

Christine A. Espin, *University of Minnesota*
Susan De La Paz, *Vanderbilt University*
Barbara J. Scierka, *University of Minnesota*
Lisa Roelofs, *Rochester School District*

This study extended the work on curriculum-based measurement to examine the criterion-related validity of curriculum-based measures in written expression for middle school students, the differences in validity coefficients for various lengths of text, and the sensitivity of curriculum-based measures to change in student performance. Curriculum-based measures were the number of correct word sequences (CWS) and correct minus incorrect word sequences (CIWS) written in expository essays. Criterion measures were the number of functional elements in and quality ratings of student essays. Results revealed a strong relationship between curriculum-based and criterion measures.

Curriculum-based measurement (CBM) is a system of measurement that can be used by teachers to monitor student progress over time and to evaluate the effects of instructional programs (Deno, 1985). Research on CBM at the elementary school level has demonstrated that simple and efficient measures can be used as general indicators of student performance in an academic area (Deno, 1985). For example, in written expression, the number of words written, the number of words spelled correctly, and the number of correct word sequences (i.e., two adjacent correctly spelled words acceptable within the context to a native English speaker) written in 3 minutes all correlate at a moderate to moderately strong level with other measures of students' writing performance. These measures include scores on standardized achievement tests, holistic evaluations of writing, and teacher evaluations of writing ability (see Marston, 1989). Further, when CBM procedures are used by teachers to monitor student progress and evaluate the effects of instructional programs, students achieve more (Fuchs, 1998).

Research on CBM has revealed that the measures used at the elementary level are not necessarily reliable and valid and at the secondary level (see Espin & Tindal, 1998). For example, in the area of written expression, simple scoring metrics such as the number of words written and the number of words spelled correctly in a limited time frame (e.g., 3–6 minutes) have not been shown to be valid and reliable indicators of general writing proficiency for secondary students. Instead, somewhat more complex scoring systems involving the use of correct word sequences (CWS) seem to be required (Espin, Scierka, Skare, & Halverson, 1999; Espin et al., 2000; Fewster & MacMillan, 2002; Parker, Tindal, & Hasbrouck, 1991a, 1991b; Tindal & Parker, 1989; Watkinson & Lee, 1992).

CBM Writing Research at the Secondary Level

Tindal, Parker, and colleagues conducted the initial research on the development of CBM measures in written expression for students at the secondary level (Parker et al., 1991a, 1991b; Tindal & Parker, 1989). Their research pointed to the use of either the number (Parker et al., 1991b; Tindal & Parker, 1989) or the percentage (Parker et al., 1991a) of correct word sequences as valid indicators of student performance in written expression. The CWS scores were valid at both middle and high school levels, although correlation coefficients were somewhat stronger at the middle school level than at the high school level (Parker et al., 1991a). Neither the number nor the percentage of CWS resulted in regular increases across the school year (Parker et al., 1991b). Percentage measures were seen to present unique problems with respect to growth monitoring because, by their nature, percentage measures are not sensitive to change in performance (Parker et al., 1991b; Tindal & Parker, 1989). If a student writes 10 word sequences at the beginning of the year with 5 correct, the percentage score is 50%. If that same student writes 50 word sequences at the end of the year with 25 correct, the percentage score remains 50%. No change in performance is reflected in the score.

Address: Christine A. Espin, Department of Educational Psychology, 250B Burton Hall, 178 Pillsbury Dr. SE, University of Minnesota, Minneapolis, MN 55455; e-mail: espin001@tc.umn.edu

Because of the inherent problems associated with use of percentage scores for growth monitoring, subsequent research focused on the number rather than the percentage of correct word sequences. Espin and colleagues (Espin et al., 1999; Espin et al., 2000) confirmed the validity and reliability of the number of CWS as an indicator of general writing performance and introduced a new scoring procedure, the number of correct minus incorrect word sequences (CIWS; Espin et al., 2000). In the Espin studies, both CWS and CIWS were found to correlate at moderate to moderately strong levels with holistic ratings of writing performance. Similar to the findings of Parker et al. (1991a), correlation coefficients were somewhat stronger at the middle school level (Espin et al., 2000) than at the high school level (Espin et al., 1999). At the middle school level, both CWS and CIWS were found to have acceptable alternate-form reliabilities, and validity and reliability were found not to differ as a function of type of text (story writing vs. descriptive writing) or writing time (3 min vs. 5 min; Espin et al., 2000). Effects of type of text and writing time were not examined at the high school level.

Purpose of the Study

Although the research that has been conducted thus far on CBM written expression at the secondary level has been consistent in its findings regarding the potential use of CWS scoring metrics, several issues have yet to be addressed. For example, in previous research, only two forms of writing have been used (story writing and descriptive writing) and the length of the writing sample has been limited to 6 min or less. The validity of CWS and CIWS for indexing student writing performance using other forms of writing and the validity with respect to varying text lengths is not known.

In addition, in previous research, the criterion variable has been primarily holistic ratings of students' writing. These holistic ratings have been conducted on samples that have not been corrected for basic writing elements, such as spelling, punctuation, and capitalization. It is possible that the correlations found between the CBM scoring metrics and holistic ratings have been a function of the influence of basic writing elements on both of these measures. The validity of CWS and CIWS with respect to higher level elements of writing, such as content, coherence, and completeness, has not been examined. Finally, research has not yet examined the validity of CWS and CIWS for indexing change in performance over time.

In the current study, we addressed several previously unexplored issues. First, we examined the reliability and validity of CBM scoring metrics using a different genre of writing: expository essays. Second, we used as criterion variables the number of functional elements (units in the essay supporting the development of the writer's paper) and quality ratings of the essays after the essays had been corrected for spelling, punctuation, and capitalization. These criterion measures reflect the content, coherence, and completeness of the essay. Finally, we examined the sensitivity of CWS and CIWS for detecting improvements in writing over time. Four research questions were addressed in this study:

1. What is the relationship between the number of CWS and CIWS written in expository essays and the number of functional elements included in those essays?
2. What is the relationship between the number of CWS and CIWS written in expository essays and quality ratings of those essays?
3. Do the relationships between the number of CWS and CIWS written in expository essays and the criterion variables differ with respect to the length of the text?
4. Are CWS and CIWS sensitive to changes in students' writing performance over time?

Based on previous research at the middle school level, we hypothesized that the relationship between CBM and criterion measures would be moderate to moderately strong. We made no hypotheses regarding the influence of text length nor the sensitivity of the measures to change over time because, to date, little research has been conducted on these issues at the secondary level.

Data Set

One difficulty associated with examining the validity of CBM writing measures for indexing growth is that, because instruction in composition is not always a part of the regular curriculum at the secondary school level (e.g., see Greenwald, Persky, Campbell, & Mazzeo, 1999), it is never clear whether a lack of improvement on the CBM measures is due to technically inadequate measures or to a lack of improvement in students' writing proficiency. Our study consists of a reanalysis of data collected as a part of an earlier study designed to investigate the effects of strategy instruction on the writing performance of middle school students with and without learning disabilities (LD; see De La Paz, 1999). Instruction was designed to teach students to plan a composition in advance of composing and to continue planning throughout the composing process. Previous research has found that students with and without LD do little advanced planning (Scardamalia & Bereiter, 1986); yet, when taught to do so, these students produce substantially better papers (Danoff, Harris, & Graham, 1993; De La Paz & Graham, 1997a, 1997b, 2002; Graham, MacArthur, Schwartz, & Page-Voth, 1992; Harris & Graham, 1985; MacArthur, Schwartz, Graham, Molloy, & Harris, 1996; Page-Voth & Graham, 1999).

The advantage of using an existing data set was that we could be assured that an intensive intervention had been delivered to the students and that the students' writing performance had improved. Results of the original single-subject design study revealed that students increased the length, quality, and completeness of their essays following implementation of strategy instruction.

Method

Participants and Setting

Participants in the study were 22 students (11 boys and 11 girls) in the seventh and eighth grades. Students were selected from five different classrooms of three language arts teachers in two suburban middle schools in the southeastern part of the United States. The schools had populations of 504 and 540 students. Students in both schools were primarily Caucasian (approximately 94%), with a small number of African American, Asian American, and Hispanic students. Eighteen percent of students in the first school and 12% of students in the second school received free and reduced-cost lunches. Less than 1% of the students received services in English as a second language. In the current study, 91% of the participants were Caucasian and 9% were African American.

Participants included students diagnosed with LD (n = 6), and low- (n = 6), average- (n = 6), and high-achieving writers (n = 4). Students without LD were classified into low-, average- and high-achieving groups, based on their scores on the written expression subtest of the *Wechsler Individual Achievement Test* (WIAT; 1992). Low-achieving writers (LA) were those with standard scores of 79 to 91, average-achieving writers (AA) were those with scores of 96 to 105, and high-achieving writers (HA) were those with scores of 116 to 123.

Students in the LD group had been identified as LD according to district criteria. Students had verbal IQ scores between 85 and 125 on individually administered norm-referenced intelligence tests, scored at least 1 *SD* below average in reading, writing, and/or mathematics on a norm-referenced achievement test, had no other handicapping conditions, and used English as their primary language. Students with LD who participated in the study had been nominated by their teachers as having difficulty with writing composition. The average WIAT standard score for the LD students was 81.

Students' percentile rank scores on the Language Arts subtest of the *Comprehensive Tests of Basic Skills* (1989), a group-administered achievement battery, were as follows: LD, 28; LA, 41; AA, 70; and HA, 73.

Procedures

Students wrote expository essays. Expository essays were chosen because seventh- and eighth-grade students were required to write expository essays to pass the state's competency test. A bank of topics was developed based on previous state exams. This bank was then shown to one special education and two general education middle school teachers, who eliminated or modified topics, based on interest and difficulty levels for middle school students. The following are some examples of topics: "Choose a country you would like to visit. Write an essay explaining why you want to go to this country," "Think about how students can improve their grades. Write an essay telling why it is important to get good grades, and explain how students can improve their grades," and "Think about rules you think are not fair. In an essay, state what rules you think should be changed, and give reasons explaining why you think so."

Essays were administered and monitored by the classroom teacher. Teachers provided students with a copy of the topic, read the topic aloud, and then read the following directions:

> Look carefully at the prompt and make up a good essay to go with it. Remember to plan your essay before you begin writing. Try to remember everything you know about writing essays. Also, it is okay to change your plan or go back to add ideas to your plan when you are composing your essay. Do you understand these instructions? After we begin writing, I cannot help you with anything.

Students were given 35 minutes to write their essays by hand. No assistance was given to students for spelling or grammar.

Students wrote a minimum of six expository essays at the beginning of the study. Following collection of the pretest data, students were instructed in writing using composition strategies designed to help them plan, organize, and write expository essays. Instruction was 4 weeks long, averaging 4 days per week. Within 1 week following instruction, students were again asked to write expository essays. Results of the multiple-baseline study revealed that the students improved in their writing performance. Students in all four groups (LD, LA, AA, and HA) wrote longer, more complete, and higher quality essays (see De La Paz, 1999, for details).

To address the research questions for the current study, a random sample of three pretests and posttests were selected from each student to be scored for the CWS and CIWS. CWS was defined as any two adjacent, correctly spelled words that were acceptable within the context of the sample to a native English speaker. Acceptable meant that a native speaker would judge the word sequence as syntactically and semantically correct (Videen, Deno, & Marston, 1982). End punctuation and beginning capitalization were also taken into account in scoring CWS and CIWS (Tindal & Parker, 1989). Pretests and posttests were scored by the first author and two graduate students. Rules for scoring CWS and CIWS were reviewed and then coders scored several essays together, discussing issues as they arose. Coders then independently scored approximately 10% each of the pretests and posttests. Percentage of scoring agreement between pairs of coders was calculated by dividing agreements by the total number of agreements plus disagreements and multiplying by 100. Rates of agreement between pairs of coders were 96.62%, 97.49%, and 97.06% for CWS and 90.02%, 90.32%, and 91.23% for CIWS.

Criterion variables in the study were the number of functional essay elements and quality ratings of the essays. Functional elements were defined as units in the essay that directly supported the development of the writer's paper (Graham, 1990). Functional elements included premises (statements spec-

ifying the writer's position on the topic), reasons (explanations to support or refute a position), elaborations (extensions or examples of a premise, reason, or conclusion), and conclusions (closing statements). Nonfunctional elements were units that were repeated without any discernible rhetorical purpose, were unrelated to the topic, or were not appropriate for an expository genre. Essays were divided into minimally parsable units (i.e., the smallest units of an argument that convey meaning; Scardamalia, Bereiter, & Goelman, 1982) and were scored as functional or nonfunctional. The number of elements in each essay ranged from 5 to 40. The second author scored all essays for functional elements. Twenty-five percent of the essays were scored by an independent rater. Interrater reliability for the total number of functional essay elements, determined in the same manner as percentage of scoring agreement, was 84%.

The quality of the essays was assessed by trained raters using a holistic rating system. Raters were unfamiliar with the purpose or design of the study. Prior to scoring, essays were typed and corrected for spelling, punctuation, and capitalization, to remove the effects of these factors on the ratings assigned to each essay (see Graham, 1992). These factors would especially penalize students with LD, who make considerably more mechanical errors than their normally achieving peers (Deno, Marston, & Mirkin, 1982).

Essays were scored by two general education teachers (1 seventh- and 1 eighth-grade teacher from another suburban middle school) who were unfamiliar with the design of the study. Raters scored the essays on the basis of their general impression of overall quality. Essays were rated on a scale of 0 to 7, with 0 = *low* and 7 = *outstanding*. Raters were instructed to consider the ideas and development of the essay; the organization, unity, and coherence; and the breadth of the vocabulary in assigning a score to the essay. Anchor points were established by selecting a high essay (score of 7), a middle essay (score of 4), and a low essay (score of 1). These essays were obtained from seventh- and eighth-grade students who were in the target schools but were not participating in the study. Interrater agreement between the two raters, as calculated by Pearson product–moment correlations, was .90. Differences between raters were resolved through discussion.

In addition to the functional elements and quality ratings, the number of words written in the essays was scored via computer. Any word that represented a spoken word, regardless of its spelling, was counted as a written word.

Results

Relationship Between CBM Scores and Criterion Measures

To address the first two research questions, correlations between the CBM scoring metrics (CWS and CIWS) and the criterion measures (functional elements and quality ratings) were examined. Pretest and posttest scores for students on the CBM and criterion measures are presented in Table 1. Correlations between the measures are presented in Table 2. Correlations for pretests and posttests were calculated separately. The magnitude of the correlations between the predictor and criterion variables was strong, ranging from .66 to .83. In general, the obtained correlations for CWS and CIWS were similar in magnitude. Correlations between the CBM scoring metrics and the quality posttest rating were lower than for the quality pretest ratings, most likely due to a bunching of posttest quality scores: While the overall range of scores for the posttest was greater than for the pretest (thus the larger standard deviation for the posttest), there was a greater bunching of scores on the posttest, with 17 of 22 students receiving average quality ratings between 5 and 7.

Moderately strong to strong correlations between the number of words written in the essay and the dependent variables were also found (r = .58–.90; see also Table 2). Although previous research on writing (see Hillocks, 1986, for a review) has revealed a relationship between length and essay quality, our finding is unusual in light of previous CBM research at the secondary level. In that research, the relationship between the number of words written and other measures of written expression proficiency, including quality of writing, has been

TABLE 1. Means and Standard Deviations on Pretest and Posttest for CBM and Criterion Measures

Measure	Pretest		Posttest	
	M	*SD*	*M*	*SD*
CWS	97.74	47.96	183.93	45.33
CIWS	80.33	47.93	151.39	56.65
Number of words written	106.83	46.11	203.77	36.96
Functional elements	12.18	3.06	27.08	4.77
Quality ratings	2.58	.63	5.26	.86

Note. CBM = curriculum-based measurement; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.

TABLE 2. Correlations Between CBM and Criterion Measures

Measure	Functional elements		Quality ratings	
	Pretest	Posttest	Pretest	Posttest
CWS	.70	.79	.83	.68
CIWS	.70	.66	.82	.67
Number of words written	.68	.90	.82	.58

Note. CBM = curriculum-based measurement; CWS = correct word sequences; CIWS = correct minus incorrect word sequences. All correlations significant at $p < .01$.

in the low to moderate range (r = .0–.47; Espin et al., 1999; Espin et al., 2000; Parker et al., 1991b; Tindal & Parker, 1989).

With our third research question, we further explored the issue of length of text. To address this question, we calculated the correlations between CWS and CIWS in the first 50 words of the writing sample and the criterion variables. This analysis addressed the issue of whether length alone was responsible for the correlations between the CBM and the criterion measures, or whether the number of correct and incorrect word sequences also were important factors.

Examination of scattergrams between the CBM and criterion variables revealed an outlier in the correlations for the pretest scores (see Note 1). This outlier was removed for subsequent analyses. Means and standard deviations for the number of CWS and CIWS on the pretest and posttest for the 50-word sample are reported in Table 3. Correlations between predictor and criterion variables are reported in Table 4.

As might be expected, the magnitude of the correlations is lessened when the number of words written is limited to 50 words because limiting the length also limits the range of CWS and CIWS scores. Nevertheless, the obtained correlations are still quite respectable, ranging from .33 to .59, and are in line with the results of McCulley (1985), who found correlations of .41 between measures of text cohesion and quality when length was held constant. These results indicate that the relationship between CWS and CIWS is not solely a result of the influence of length of text: Other factors contribute to the correlations.

TABLE 3. Means and Standard Deviations for CBM Measures on First 50 Words of Pretest and Posttest

	Pretest[a]		Posttest	
Measure	*M*	*SD*	*M*	*SD*
CWS	44.67	5.39	46.86	4.24
CIWS	36.10	10.17	39.55	8.62

Note. CBM = curriculum-based measurement; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.
[a]Means and standard deviations of pretest before outlier was removed were as follows: M = 43.21, SD = 8.64, for CWS; M = 33.61, SD = 15.30, for CIWS.

TABLE 4. Correlations Between CBM Scores for First 50 Words and Criterion Measures

	Functional elements		Quality ratings	
Measure	Pretest[a]	Posttest	Pretest[a]	Posttest
CWS	.43*	.35	.59**	.56**
CIWS	.44*	.33	.58**	.54*

Note. CBM = curriculum-based measurement; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.
[a]Correlations between the number of functional elements and CWS and CIWS before the outlier was removed were r =.53 and r = .54, respectively. Correlations between quality ratings and CWS and CIWS were r = .43 and r = .44, respectively.
*p <. 05. **p < .01.

Sensitivity to Change Over Time

Our final research question addressed the sensitivity of the CBM scoring metrics to change in performance over time. To address this question, we examined the differences from pretest to posttest for both the CBM scores and the criterion measures. We expected significant pre–post differences in the criterion measures, given the improvements demonstrated in the single-subject design study conducted by De La Paz (1999). Our primary interest, however, was whether the CBM scoring metrics would also be sensitive to change in performance over time. Due to the unexpected results related to the number of words written in text, we present words written in this set of analyses as a potential CBM scoring metric.

Given the multiple number of dependent variables we wished to analyze, we first ran a MANOVA, with time (pretest to posttest) as a within-subjects factor. Dependent variables entered into the analysis included functional essay elements, quality ratings, CWS, CIWS, and number of words written. Results of the MANOVA revealed significant effects, $\Lambda = .13$, $F(5, 38) = 52.57$, $p <. 001$. Follow-up univariate F tests revealed significant changes from pre- to posttests on the number of functional essay elements, $F(1, 42) = 151.83$, $p < .001$, $\eta^2 = .78$, and quality ratings, $F(1, 42) = 139.59$, $p < .001$, $\eta^2 = .77$, confirming the results found in the De La Paz (1999) single-subject design study. Of interest to us was the fact that significant differences between pre- and posttest also were found for CWS, $F(1, 42) = 37.52$, $p < .001$, $\eta^2 =.47$; CIWS, $F(1, 42) =20.18$, $p <. 001$, $\eta^2 = 32$; and number of words written, $F(1, 42) = 59.21$, $p < .001$, $\eta^2 = .59$. These differences indicate that the CBM scoring metrics also were sensitive to change over time. Of note, the eta-squared value for number of words written was larger than that for CWS and CIWS.

The number of students within each group was too small to allow for statistical testing of group differences in growth; however, inspection of the obtained group differences in growth reveal patterns that are worthy of mention. Figures 1 and 2 display changes over time by group for the number of CWS and CIWS in the entire essay. Examination of these figures reveals that although, as might be expected, the levels of performance for the LD and LA groups are below that of the AA and HA groups, rates of growth are fairly similar for students in all four groups (see Note 2).

Figures 3 and 4 display changes over time by group on the number of CWS and CIWS written in the first 50 words of the essay. These figures reveal that the HA, AA, and LA students in our study showed little change from pretest to posttest (approximately 1.5 CWS and 7 CIWS on average across the three groups) when the length of the text was limited to 50 words. In comparison, students with LD showed more substantial changes (approximately 9 CWS and 14 CIWS)

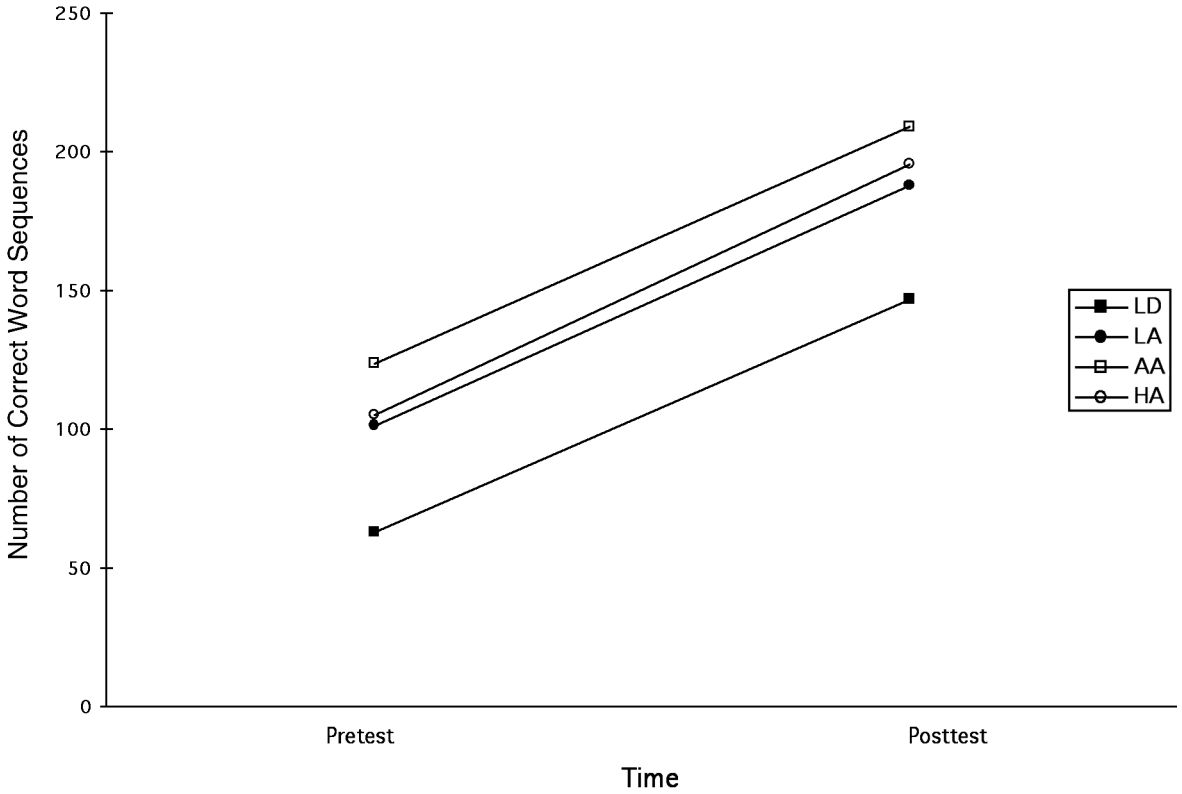

FIGURE 1. Number of correct word sequences on pretests and posttests for learning disabled (LD), low-achieving (LA), average-achieving (AA), and high-achieving (HA) students.

when length was limited to 50 words. These results indicate that for the students who were at the lowest end of the writing performance continuum (i.e., students with LD), a fairly short sample of writing revealed growth over time; however, for students at the higher end of the continuum, a longer sample was necessary.

Discussion

Results of this study provide support for the use of CBM scoring procedures in written expression. Both CWS and CIWS were strongly correlated with the criterion measures of the number of functional elements and quality ratings of the essays. In addition, both measures were sensitive to change in student performance over time.

This study contributes to our current body of knowledge about CBM in several ways. First, it provides confirmation that simple scoring procedures such as CWS and CIWS can be used as valid indicators of student performance in written expression. Second, it extends previous work to reveal that simple measures such as CWS and CIWS are valid not only for narrative and descriptive writing but also for a type of writing often required in secondary schools, expository essay writing. Finally, it adds to our confidence in the use of CWS and CIWS scores as general indicators of writing proficiency because the criterion variables in this study—functional elements and quality—focused on the content, coherence, and completeness of the writing, while controlling for basic elements of writing such as spelling, capitalization, and punctuation.

The results of this study are important both practically and scientifically. Practically speaking, training teachers to score CWS and CIWS is much simpler and less time-consuming than training them to score functional essay elements or quality ratings. In addition, it is much more likely that teachers will use simple scoring procedures such as CWS and CIWS on a repeated basis for monitoring growth over time than more time-consuming measures, such as functional elements or quality ratings. Scientifically, the results of this study indicate that researchers may also choose to use a simpler measure of student performance when evaluating the effect of their interventions in written expression.

Although the results of this study are encouraging and contribute to our knowledge base, they also raise some important questions. First, based on previous research, we hypothesized that correlations between CWS and CIWS and the criterion variables would be in the moderate to moderately strong range. The magnitude of the correlations found in this study are large compared to those found in previous studies of CBM for middle school students (e.g., Espin et al., 2000). We

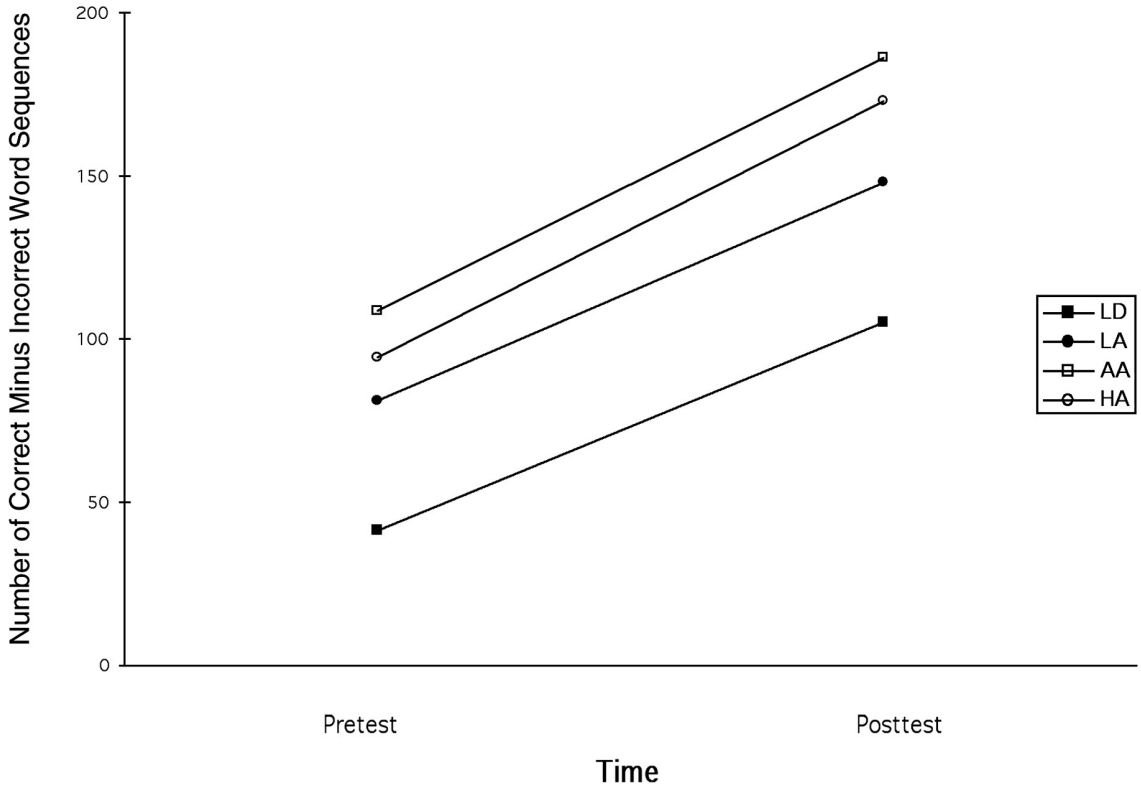

FIGURE 2. Number of correct minus incorrect word sequences on pretests and posttests for learning disabled (LD), low-achieving (LA), average-achieving (AA), and high-achieving (HA) students.

speculate that these differences are due to variations in the time allowed for students to write. It may be necessary to give students more time to write, to obtain a more valid and reliable sample of their performance. Indeed, the analysis conducted with 50 words supports this conclusion. Correlations between the predictor and criterion measures dropped when CWS and CIWS were counted for a 50-word sample. These results indicated that although length of text alone did not account for the relationship between the predictor and criterion variables, it was an important factor. The need for a longer sample of text is also supported by the results of Parker et al. (1991b), who found that 6-min samples of writing did not produce indicators that were stable measures of growth over time.

The question that arises with respect to length of text is, How long is long enough? Though preliminary, our results indicate that the amount of writing time needed may vary with students' level of writing proficiency. Examination of growth patterns by group indicated that for students diagnosed with LD, fairly short writing samples were sufficient for reflecting growth over time, but for students who were better writers, longer writing samples were necessary. One possible reason for these differences is that CWS and CIWS reflect both length and errors. For writers who make fewer errors, the measures reflect mostly length; for writers who make more errors, the measures reflect both length and errors. Thus, even when length is controlled, CWS and CIWS predict writing quality for poor writers but not for good writers.

Practically speaking, the length of the sample that is needed to reveal changes over time must be balanced with the efficiency of the measurement system. The amount of writing time given in this study is probably too long in CBM terms. If teachers are to collect samples of student work on a weekly basis, and score and graph the data, a 35-min time frame is too long. In future research studies, time frames such as 3, 5, 10, and 15 minutes should be compared for students at various levels of writing proficiency. Our speculations about the effects of length of text are preliminary and conjectural and must be substantiated in future research in which larger sample sizes are employed.

A second question raised by this study relates to the use of the number of words written as a CBM indicator. In the current study, unlike previous CBM studies, the number of words written correlated at a moderately strong to strong level. Further, eta-squared values indicated that number of words written was more sensitive to change in performance over time than was CWS or CIWS. The use of the number of words written score needs to be examined more closely. In none of the previous CBM research studies conducted at the secondary level has the number of words written been found to be a valid indicator of students' general writing performance. We spec-

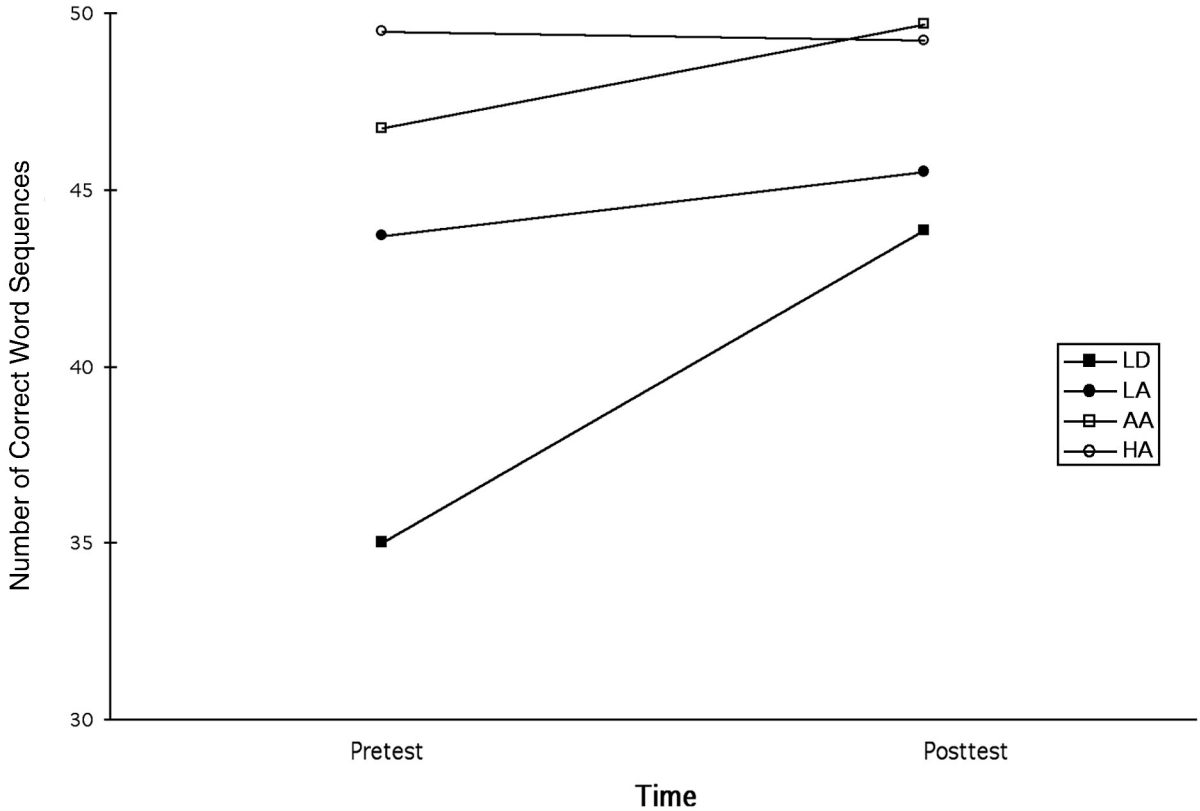

FIGURE 3. Number of correct word sequences on first 50 words of pretests and posttests for learning disabled (LD), low-achieving (LA), average-achieving (AA), and high-achieving (HA) students.

ulate again that this finding may be related to the length of the sample collected in this study. This finding is worth pursuing further because, although CWS and CIWS are easier to score than quality ratings and functional elements, counting the number of words written is even easier to score than CWS or CIWS.

Our research was limited by the small sample size, which may have affected the magnitude of our correlations. A future study with a larger sample of students with diverse writing abilities would allow for a more systematic examination of some of the questions raised in this study. In addition, our study was limited by the use of a pretest–posttest design for examining change in performance. Because the order in which the measures were administered was not counterbalanced, the observed changes may have been influenced by topic. This concern is somewhat diminished by the fact that for each student, three pretests and three posttests were randomly selected from a pool of six or more potential essays; thus, change scores were not calculated on the same set of pretests and posttests for each student. Nonetheless, topic may still have exerted an influence on the observed change scores.

An additional limitation to our study is the fact that we examined change based on a small number of data points collected at the beginning and end of an intervention. CBM measures, however, are designed to be given on a repeated and frequent basis (i.e., once a week) so that teachers can make decisions regarding the effects of their interventions throughout the school year. A stronger test of the validity and reliability of the CBM writing measures for growth monitoring would be to examine the technical adequacy of the growth trajectories created by the measures. (See Shin, Deno, & Espin, 2000, for an example of this type of study using reading data collected at the elementary school level.)

In conclusion, the results of this study support the use of CWS and CIWS as indicators of students' general writing performance and introduce the possibility of using the number of words written as a CBM indicator. The results support previous research on the use of CWS and CIWS as general indicators of performance and contribute to our knowledge base about the use of CBM procedures for students at the middle school level.

AUTHORS' NOTES

1. This study consists of a reanalysis of data collected for a previous study (see De La Paz, 1999).
2. This research was funded in part by Vanderbilt University's Research Council. We wish to thank the teachers and students of the

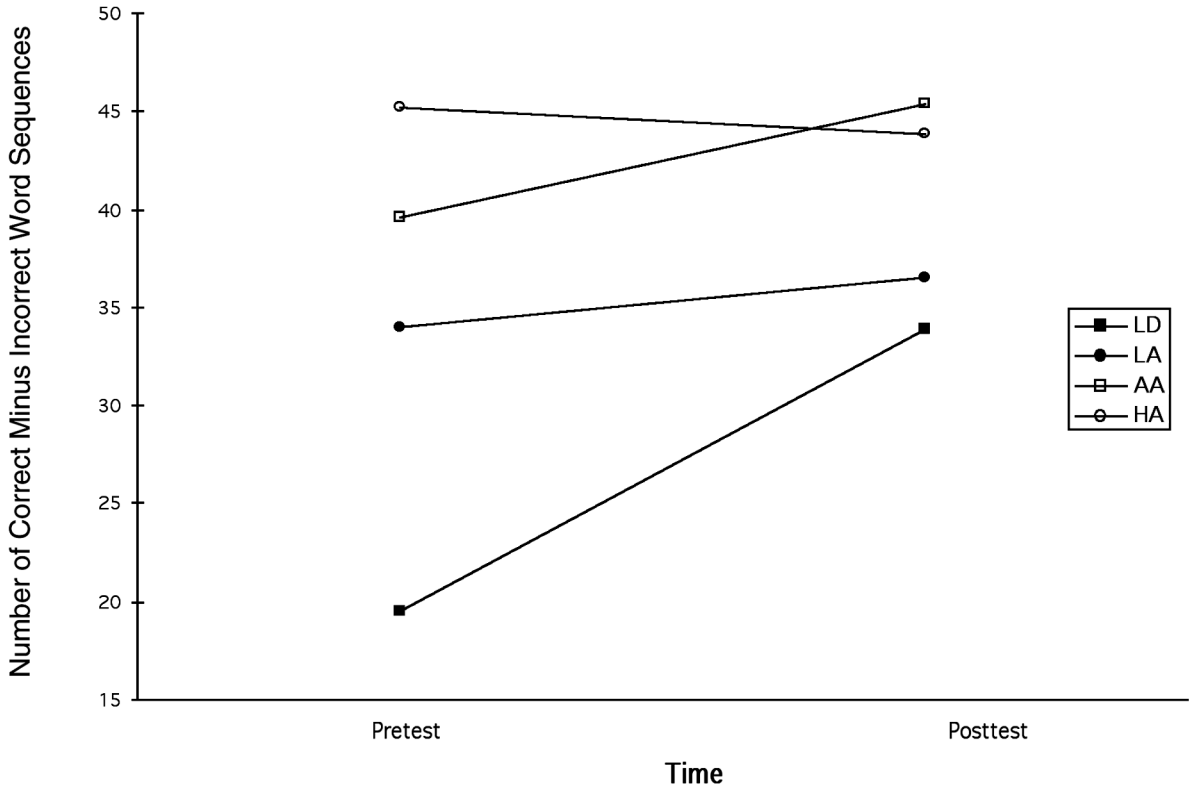

FIGURE 4. Number of correct minus incorrect word sequences on first 50 words of pretests and posttests for learning disabled (LD), low-achieving (LA), average-achieving (AA), and high-achieving (HA) students.

Sumner County Public School District for their participation in this study and to acknowledge the Netherlands Institute for Advanced Study in the Humanities and Social Sciences for its support in the preparation of this manuscript. We also wish to thank Stanley Deno, Jongho Shin, Paul van den Broek, and Charles MacArthur for their helpful comments on an earlier version of this article.

NOTES

1. This data point fell within the normal range of scores for analyses involving all words, and thus was not removed for those analyses.
2. The figures demonstrate that the AA group did somewhat better on the writing performance measures than the HA group. These differences were evident in the levels of performance on the functional elements and quality ratings as well.

REFERENCES

Comprehensive tests of basic skills (4th ed.). (1989). Monterey, CA: School Publishing.

Danoff, B., Harris, K. R., & Grahams, S. (1993). Incorporating strategy instruction within the writing process in the regular classroom: Effects on the writing of students with and without learning disabilities. *Journal of Reading Behavior, 25,* 295–322.

De La Paz, S. (1999). Self-regulated strategy instruction in regular education settings: Improving outcomes for students with and without learning disabilities. *Learning Disabilities Research & Practice, 14,* 92–106.

De La Paz, S., & Graham, S. (1997a). Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology, 89,* 203–222.

De La Paz, S., & Graham, S. (1997b). Strategy instruction in planning: Effects on the writing performance and behavior of students with learning difficulties. *Exceptional Children, 63,* 167–181.

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94,* 687–698.

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

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

Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Curriculum-based measures in writing for secondary students. *Reading and Writing Quarterly, 15,* 5–28.

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

Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214–253). New York: Guilford Press.

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

Fuchs, L.S. (1998). Computer applications to address implementation difficulties associated with curriculum-based measurement. In M. Shinn (Ed.), A*dvanced applications of curriculum-based measurement* (pp. 89–112). New York: Guilford Press.

Graham, S. (1992). Composition research and practice: A unified approach. *Focus on Exceptional Children, 14*(8), 1–16.

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82,* 781–791.

Graham, S., MacArthur, C., Schwartz, S., & Page-Voth, T. (1992). Improving the compositions of students with learning disabilities using a strategy involving product and process goal setting. *Exceptional Children, 58,* 322–344.

Greenwald, E.A., Persky, H.R., Campbell, J.R., & Mazzeo, J. (1999). *The NAEP 1998 writing report card for the nation and the states* (NCES 1999-462). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Harris, K. R., & Graham, S. (1985). Improving learning disabled students composition skills: Self control strategy training. *Learning Disability Quarterly, 19,* 179–200.

Hillocks, G. (1986). *Research on written composition.* ERIC Clearinghouse on Reading and Communication Skills and the National Conference on Research in English.

MacArthur, C., Schwartz, S., Graham, S., Molloy, D., & Harris, K. (1996). Integration of strategy instruction into a whole language classroom: A case study. *Learning Disabilities Research & Practice, 11,* 168–176.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shin (Ed.), *Curriculum-based measurement* (pp. 18–78). New York: Guilford Press.

McCulley, G. (1985). Writing quality, coherence, and cohesion. *Research in the Teaching of English, 19,* 269–282.

Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology, 91,* 230–240.

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

Parker, R., Tindal, G., & Hasbrouck, J. (1991b). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children, 58,* 61–73.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 778–803). New York: Macmillan.

Scardamalia, M., Bereiter, C., & Goelman, H. (1982). The role of production factors in writing ability. In M. Nystrand (Ed.), *What writers know: The language, process, and structure* (pp. 173–210). New York: Academic Press.

Shin, J., Deno, S. L., & Espin, C. A. (2000). Technical adequacy of the maze tasks for Curriculum-based measurement of reading growth. *The Journal of Special Education, 34,* 164–172.

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

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Research Report 84). Minneapolis, MN: Institute for Research on Learning Disabilities.

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.

Wechsler Individual Achievement Test. (1992). San Antonio, TX: Psychological Corp.