Hammill Institute on Disabilities

The Journal of Special Education
44(4) 195–204

DOI: 10.1177/0022466909333516
http://journalofspecialeducation.sagepub.com
SAGE

The Predictive Validity of CBM Writing Indices for Eighth-Grade Students

Janelle M. Amato[1] and Marley W. Watkins[1]

Abstract

Curriculum-based measurement (CBM) is an alternative to traditional assessment techniques. Technical work has begun to identify CBM writing indices that are psychometrically sound for monitoring older students' writing proficiency. This study examined the predictive validity of CBM writing indices in a sample of 447 eighth-grade students. Regression analyses revealed that simple fluency measures were not adequate for assessing secondary students' writing. A more complex fluency measure, the number of correct punctuation marks, and an accuracy-based measure, the percentage of correct word sequences, were the best predictors of a written expression test for eighth-grade students. However, overall results of the current study provided only limited support for the use of CBM to assess writing skill at the secondary level.

Keywords

curriculum-based measurement, written expression, writing assessments, secondary students

Writing has always been an important part of the school curriculum (U.S. Department of Education, Institute of Education Sciences, & National Center for Education Statistics, 2003) but has received less attention from researchers than reading and mathematics (Bradley-Johnson & Lesiak, 1989). In particular, the development of psychometrically sound methods to assess writing has lagged other academic areas (Watkinson & Lee, 1992). However, in the past two decades, an increased number of research studies in the area of written expression have been conducted.

Efforts to improve written language instruction in the schools have been accompanied by an increased number of research studies examining different aspects of written language. There are many studies that describe the characteristics of essays that were written by skilled and unskilled writers and the processes used by skilled and unskilled writers to create their final product (Cole, Haley, & Muenz, 1997; Graham, 1990; Graham & Harris, 1997; Hooper et al., 1994; MacArthur & Graham, 1987). Although not a theoretical model of the writing process, these distinctions help define good writing. The characteristics of the final product and of the processes used are important not only in distinguishing between skilled and unskilled writers but also for developing treatment and instructional programs (Hooper et al., 1994).

In general, skilled writers are described as goal-directed learners who apply various writing and self-regulation strategies such as planning, revising, organizing, monitoring, and evaluating (Cole, Haley, et al., 1997; Graham & Harris, 1997). They understand the goals of the writing assignment and have a greater knowledge about their writing topic and their audience (Hooper et al., 1994). In addition, skilled writers generate more ideas and eliminate their less productive ideas as they revise and edit. These characteristics add to the smoothness and cohesiveness of the writer's final product (Graham & Harris, 1997; Hooper et al., 1994).

In contrast, children who find writing challenging use less sophisticated approaches to writing. They not only display deficits in their use of self-regulation strategies but also show problems in generating text ideas. Unskilled writers are less likely to revise their spelling, punctuation, grammar, or text ideas, resulting in poorly written text (Graham & Harris, 1997; Hooper et al., 1994). In addition, students who experience difficulties in writing tend to have shorter compositions and provide the audience with little detail or elaboration when compared to skilled writers.

Graham and Harris (1997) provided three possible reasons why unskilled writers produce shorter essays. The first reason may be because students who struggle with writing terminate their writing process too soon. A study by Graham

[1]Arizona State University, Tempe, AZ, USA

Corresponding Author:
Janelle Amato, PhD, 1 Lincoln Drive, Cresskill, NJ 07626
E-mail: jamato@cresskillboe.k12.nj.us

(1990) provided some evidence for this proposition. Fourth- and sixth-grade students with learning disabilities (LD) generally wrote for 6 or 7 minutes when writing an essay. However, when verbally prompted to write more, these students generated substantial increases in the amount of text written.

Graham and Harris (1997) also suggested that unskilled writers may produce shorter essays when compared to skilled writers due to poorly developed mechanical skills. In a study by MacArthur and Graham (1987), fifth- and sixth-grade students with LD produced longer stories and improved the quality of their stories when they dictated their essays versus handwriting them or typing them on a word processor. These results are consistent with Graham (1990), who examined the effects of mechanical skills on writing for fourth- and sixth-grade students with LD. Graham also found that under normal conditions, students generated better quality essays when the stories were dictated rather than written. However, unlike McArthur and Graham, Graham did not observe an increase in the length of output as a result of mode.

Finally, the third possible reason provided by Graham and Harris (1997) is not related to writing skills but related to topic knowledge and interest. Graham and Harris proposed that students who lack knowledge or interest in a topic will generate less text than those students who are knowledgeable and interested in the topic. In the past, researchers had largely theorized that individual interest influences students' writing; however, these claims are not supported by research (Hidi & McLaren, 1990, 1991). For example, Hidi and McLaren (1991) found that students in the fourth and sixth grades did not write longer or qualitatively better essays on topics that they identified as interesting when compared with topics that they identified as uninteresting. In contrast, other research has supported the hypothesis that topic knowledge does influence length and quality of a written product (DeGroff, 1987; Kellogg, 1987; McCutchen, 1986). For example, McCutchen found that high school students who were knowledgeable about football wrote lengthier and more coherent texts than did students who had low levels of knowledge about football.

Given that students with writing difficulties tend to have more difficulties with writing lengthier texts and produce poorly written compositions than those students who do not have difficulties in writing, it is important to examine the lower level skills that are needed in writing such as handwriting, spelling, grammar, punctuation, and vocabulary. It is theorized that students who are not fluent in these lower level skills will have more difficulty successfully utilizing higher level writing strategies, such as planning, generating text, and revising (Graham & Perin, 2007). Thus, utilizing an assessment technique that reliably and validly targets students who are having difficulty with these lower level skills may facilitate classroom interventions and instructions in regard to these lower level skills.

One type of assessment technique that measures the lower level skills in writing is curriculum-based measurement (CBM). CBM scores are indicators of student performance primarily in the area of basic skills. Therefore, CBM scores are not direct measures of the underlying construct (Shinn & Bamonto, 1998). In contrast, CBM is intended to index only basic skills, such as the lower level skills needed for written expression (i.e., spelling, punctuation, and grammar). Unfortunately, the evidence supporting the reliability and validity of CBM writing scores is less convincing than the evidence supporting the psychometric properties of reading and math CBM scores (Fewster & Macmillan, 2002; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Good & Jefferson, 1998; Marston, 1989; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). Nevertheless, past research has provided moderate support for the reliability and validity of CBM scores in written expression, thus warranting detailed research on the criterion-related validity of CBM writing assessment (Fewster & Macmillan, 2002; Good & Jefferson, 1998; Marston, 1989). The present authors will follow guidelines that were developed in previous research in CBM. Thus, reliability and validity coefficients that are .70 and above are considered strong, coefficients between .50 and .70 are considered moderate, and coefficients below .50 are considered weak (McMaster & Espin, 2007; Wayman et al., 2007).

Generally, three CBM writing indices have been examined at the elementary level: total words written, total words spelled correctly, and the number of correct word sequences (CWS). These production-dependent scoring indices are fluency based (Tindal & Parker, 1989, 1991; Watkinson & Lee, 1992). In regard to total words written, research has consistently demonstrated that students who struggled with writing tended to produce shorter essays or compositions than skilled writers (Deno, Marston, & Mirkin, 1982; Graham & Harris, 1997; Houck & Billingsley, 1989; Poteet, 1979). For example, in a study comparing the writing of 48 students with LD and 48 normally achieving students in Grades 4, 8, and 11, on average, students with LD wrote fewer words (M = 140.9 words) than students who were achieving at an average level (M = 112.3 words; Houck & Billingsley, 1989). In the majority of research examining CBM scores with elementary students, total words written was the strongest index correlated with various writing criteria (Deno, Marston, & Mirkin, 1982; Deno, Marston, Mirkin, Lowry, et al., 1982; Marston, Lowry, Deno, & Mirkin, 1981). Consequently, total words written has been recommended for primary use in Grades 1 through 3 and optional use in Grades 4 through 12 (Malecki, 2008).

Correct spelling is an important component of writing because readers of the written product must be able to read

the words written by the student. Several studies have compared the spelling skills of students with LD and students who are achieving normally. Consistently, research has shown that students without LD spell more words correctly than students with LD (Barenbaum, Newcomer, & Nodine, 1987; Houck & Billingsley, 1989; Moran, 1981). Total words spelled correctly was also found to be a moderate to strong predictor of general writing performance for elementary students (Deno, Marston, & Mirkin; Tindal & Parker, 1991), but more recent studies with elementary students reported that total words spelled correctly did not predict criterion measures of writing (Gansle et al., 2002). Thus, results for this index have been incongruent.

Finally, researchers have theorized that CWS would be a useful indicator of written expression given that it simultaneously takes into account the number of words written and the grammar, spelling, punctuation, and capitalization of the written product (Good & Jefferson, 1998). The number of CWS was reported to correlate at a moderate to strong level with criterion measures of writing in several studies (Espin et al., 2000; Tindal & Parker, 1991) and was shown to discriminate between students in a classroom for LD and a general education classroom (Tindal & Parker, 1991). In addition, the number of CWS was reported to increase from Grades 3 through 6 (Tindal & Parker, 1991) and from Grades 1 through 5 (Malecki & Jewell, 2003). Thus, research at the elementary level has favored total words written and the number of CWS as reliable and valid indicators of students' writing performance (Malecki, 2008).

However, at the secondary level, these two production-dependent indices have consistently failed to reliably or validly measure students' performance in written expression (Espin et al., 2000; Espin, Scierka, Skare, & Halverson, 1999; Gansle et al., 2002; Malecki & Jewell, 2003; Parker, Tindal, & Hasbrouck, 1991; Tindal & Parker, 1989; Watkinson & Lee, 1992; Weissenburger & Espin, 2005). In contrast, research has supported the reliability and validity of production-independent indices, or measures of accuracy, as the strongest indicators of older students' writing performance (Espin, De La Paz, Scierka, & Roelofs, 2005; Parker et al., 1991; Tindal & Parker, 1989; Watkinson & Lee, 1992). In particular, the percentage of correctly spelled words and the percentage of CWS have moderately to strongly predicted writing criteria. These two indices significantly discriminated between students with and without LD and significantly discriminated between students in remedial and general education (Tindal & Parker, 1989; Watkinson & Lee, 1992).

In addition to production-independent indices, a relatively new index, CWS-IWS (the number of correct word sequences minus the number of incorrect word sequences) has been a moderate to strong indicator of writing performance for secondary students (Espin et al., 2000; Espin et al., 2005; Malecki & Jewell, 2003). Preliminary evidence has also shown the number of CWS-IWS to significantly increase from fall to spring in Grades 1 through 8 (Malecki & Jewell, 2003), demonstrating sensitivity to change over time. Thus, at the secondary level, the percentage of correctly spelled words, the percentage of CWS, and the number of CWS-IWS have been the most reliable and valid indicators of students' writing performance.

In summary, most of the research at the elementary level has supported words written and total CWS as valid estimates of general writing ability. However, these production-dependent indices have not proven to be valid predictors of older students' written expression skills. Thus, additional research is needed to examine the validity of CBM writing indices among older students with higher levels of complex writing.

Consequently, the purpose of the present study was to investigate the predictive validity of 10 CBM measures of written expression for a sample of students in Grade 8. In addition to examining the overall contribution of the predictor variables, the unique contribution of each predictor variable was also scrutinized. Past research most often included criterion measures that represented either direct or indirect assessments of writing but not both forms of writing. Thus, this study utilized a criterion measure that included both forms of writing. If results support the predictive validity of curriculum-based scoring indices, school personnel can utilize these scores as valid and efficient tools for assessing older students' writing skills.

Method

Participants

Participants included all eighth-grade students (n = 447) in a New Jersey school district. Participants ranged from ages 12 to 16, with 66% of the sample aged 13 and 30% of the sample aged 14. The sample was 60% Caucasian, 14% Black, 14% Hispanic, and 12% Asian/Pacific Islander and was 52% male and 48% female. Sixty-six students, 15% of the sample, were enrolled in special education programs, and 381 students, 85% of the sample, were in regular education. Around 9% of the students received free lunch, and 15% of the students received reduced lunch.

Measures

The predictor variables in this study were the 10 writing indices often used for scoring CBM writing probes. To remove the variability associated with different story starters, only one story starter was used: "One day our teacher was sick. We had a substitute teacher and. . . ."

Total words written. Total words written was defined as the number of words that the student wrote in 3 minutes.

Spelling, grammar, and content were not taken into consideration when counting the number of words. Numerical representations and symbols were not included in this total.

Words spelled correctly. Total words spelled correctly was defined as the number of correctly spelled words written by the student. Each word counted as correct had to be able to stand alone in the English language. However, context and grammar were not taken into account. Therefore, the word did not need to be used correctly within the context of the sentence; it needed only to be spelled correctly.

Percentage of words spelled correctly. The percentage of words spelled correctly was the ratio of the number of words spelled correctly to the total number of words written in the composition.

Number of CWS. CWS was defined as two adjacent, correctly spelled words that were syntactically and semantically appropriate given the context of the sentence. Thus, words were examined for correct meanings, tenses, number agreement (singular or plural), and noun-verb correspondences, when identifying CWS. In addition, punctuation, capitalization, and spelling were taken into account when scoring correct word sequences.

Percentage of CWS. The percentage of CWS was the ratio of the number of correct word sequences divided by the total number of possible word sequences.

Number of CWS-IWS. The number of CWS-IWS was calculated by subtracting the number of incorrect word sequences from the total number of correct word sequences.

Number of sentences. A sentence was defined as any series of words separated from another series of words by a space or punctuation mark, such as a period, question mark, or exclamation point. The series of words had to include a recognizable subject and verb but did not need to contain the appropriate beginning capitalization or correct ending punctuation.

Number of correct capitalization. Correct uses of capital letters were counted to determine the number of correct capitalization. This included the first letter of a word used to begin a sentence, days of the week, months, holidays, countries, languages, nationalities, religions, people's names and titles, trademarks and names of companies, places and monuments, names of vehicles, and titles of books, poems, songs, plays, and films. In addition, capital letters were required for the personal pronoun I and for acronyms.

Number of punctuation marks. The number of punctuation marks was defined as the total number of punctuation marks used, regardless of whether they were appropriate for the sentence.

Number of correct punctuation marks. The number of correct punctuation marks was defined as only those punctuation marks that were determined to be used appropriately for that sentence. In addition, students had to correctly place the punctuation mark in the sentence.

Although it is difficult to identify a writing criterion that is unanimously accepted as a good measure of writing (Cole, Haley, et al., 1997; Hooper et al., 1994), there are several factors that should be considered when selecting a measure of written expression. First, to appropriately evaluate student performance, the data obtained from an assessment method should be reliable and valid (Cole, Muenz, Ouchi, Kaufman, & Kaufman, 1997; Salvia & Ysseldyke, 2001). Second, Hooper et al. (1994) recommended that writing assessments should include production components, such as spelling, proofreading, and mechanics, in order to determine students' knowledge of writing conventions. Third, writing assessments should contain some aspect of direct measurement, in which students can apply writing conventions to an actual writing task (Cole, Muenz, et al., 1997; Hooper, 2002; Hooper et al., 1994), thereby increasing the ecological validity of the assessment. Fourth, Hooper et al. suggested that a picture stimulus be used to elicit a writing sample. This stimulus should be a color photograph, contain at least two characters, display a novel and interesting depiction, and portray a state of conflict. Fifth, the writing measure should contain scoring criteria that differentiate among poor and skilled writing qualities (Cole, Muenz, et al., 1997). Given these assessment recommendations, the *Test of Written Language–Third Edition* (TOWL-3; Hammill & Larsen, 1996) was chosen as the criterion variable. The TOWL-3 satisfied all but one (using a colored photograph as a picture stimulus) of the recommendations for a well-designed written expression measure. In addition, the reliability and validity of the TOWL-3 scores are reported to be sufficient for making high-stakes individual decisions (Hammill & Larsen 1996; Salvia & Ysseldyke, 2001).

Procedure

After obtaining consent from school district officials and the university institutional review board, all eighth-grade students were administered predictor and criterion writing measures by the primary researcher. Of the 464 students enrolled in the eighth grade, 17 did not participate in the study due to absence, suspension, lack of English language skills, or physical disability. The remainder, 447 students, completed the writing measures over two or three 40-minute English periods. All tests were group administered to intact classes. The CBM writing probe was administered prior to the TOWL-3. There were 23 general education classes that included students with special education needs. These classes ranged in size from 14 students to 27 students. In addition, there were four self-contained classrooms that included only students with special education

needs. These classes ranged from 2 students to 6 students in each class.

Ten doctoral students in school psychology were trained in a 4-hour session to score the writing measures. Scorers received a combination of two training manuals that included detailed descriptions and scoring instructions for the CBM indices (Powell-Smith & Shinn, 2004; Wright, 1992) as well as a handout that reviewed the rules of grammar (American Psychological Association, 2001; *Grammar Slammer*, 1997). Finally, the TOWL-3 *Examiner's Manual* (Hammill & Larsen, 1996) guided scoring of TOWL-3 protocols.

Scorers were initially trained on three sample CBM probes and two sample TOWL-3 protocols. Once trained, they were tested for accuracy with a final protocol of each test and proceeded to score student protocols upon attainment of 95% or greater accuracy. To ensure scoring consistency throughout the scoring period, every eighth protocol from each rater was checked by the primary investigator to identify common errors. If errors were found, the packet of eight protocols was returned to that rater for rescoring. Following these procedures, 33 cases were rescored to ensure accurate results. In addition, scorers posed specific scoring questions to the investigator via e-mail throughout the scoring process. The e-mailed questions and the responses were forwarded to the other scorers to promote scoring consistency.

Data Analysis

Multiple regression analyses were completed to determine the relationship between CBM indices and written expression, as measured by the TOWL-3. Simultaneous multiple regression was applied because it assesses the relationships among predictor and criterion variables and answers two fundamental questions: (a) What is the size of the overall relationship between the criterion variable and the set of predictor variables? and (b) How much is each predictor variable contributing uniquely to the prediction of the criterion variable (Tabachnick & Fidell, 2001)?

Results

Student protocols were divided among the 10 scorers, with each rater scoring approximately 40 CBM protocols and 40 TOWL-3 protocols. Each protocol was initially scored by the primary researcher, independently scored by a trained doctoral student in school psychology, and rescored for clerical and computational errors by the primary researcher. Average interscorer reliability between the primary researcher and independent raters was extremely high. All correlations were .945 or above for the CBM indices and .99 for the Overall Writing Quotient of the TOWL-3.

Table 1. Descriptive Statistics of CBM Indices and the Overall Writing Quotient

Measure	*M*	*SD*	Skew	Kurtosis
TWW	59.62	16.94	0.05	0.03
WSC	58.36	17.01	0.01	0.02
%WSC	97.63	3.92	–4.56	34.25
CWS	60.29	19.36	0.06	0.01
%CWS	90.58	9.62	–2.34	8.53
CWS-IWS	54.63	20.92	–0.17	0.18
SEN	5.40	2.02	0.47	0.50
CC	6.56	3.02	0.72	1.06
TPM	8.23	3.54	0.59	0.99
CPM	7.80	3.33	0.59	1.29
Overall Writing Quotient	96.35	13.51	–0.64	0.67

Note: %CWS = percentage of correct word sequences; %WSC = percentage of words spelled correctly; CC = number of correct capitalizations; CPM = number of correct punctuation marks; CWS = number of correct word sequences; CWS-IWS = number of correct word sequences minus incorrect word sequences; SEN = number of sentences; TPM = number of punctuation marks; TWW = total words written; WSC = number of words spelled correctly.

Cronbach's coefficient alpha was used to determine the degree of homogeneity among the test items of the TOWL-3. Although the coefficient alphas for the scores obtained from the Overall Writing Quotient ($\alpha = .84$) fell below that of the test authors' research (Hammill & Larsen, 1996), it was above the acceptable range of .80 (Salvia & Ysseldyke, 2001). However, the alpha coefficients for the Contrived and Spontaneous Writing Quotients (.79 and .73, respectively) were below those found in the normative sample and what is recommended for screening and decision making (Salvia & Ysseldyke, 2001). Thus, analyses were conducted with only the more reliable Overall Writing Quotient.

Descriptive statistics for all variables are provided in Table 1 and an intercorrelation matrix is presented in Table 2. Prior to analysis, all 10 CBM indices and the Overall Writing Quotient were examined through various scatterplots and statistical equations for accuracy of data entry, absence of outliers, absence of multicollinearity, and fit between their distributions and the assumptions of multivariate analysis. Results of the regression diagnostics, which included leverage, discrepancy, and influence, revealed one statistical outlier (Cohen, Cohen, West, & Aiken, 2003). That student was enrolled in special education, so was retained in the data sample as a legitimate member of the sample (Meyers, Gamst, & Guarino, 2006). The distribution of residuals was normal and the small skewness that appeared in the histogram should not affect conclusions. The Durbin-Watson coefficient indicated that the residuals were independent. There was no substantial departure from the assumption of linearity, but homoscedasticity was violated by percentage of words spelled correctly (%WSC) and percentage of correct word sequences (%CWS).

Table 2. Intercorrelations Among CBM Indices and the *Test of Written Language* Writing Quotient

Measure	TWW	WSC	%WSC	CWS	%CWS	CWS-IWS	SEN	CC	TPM	CPM	Writing Quotient
TWW	1.000	0.994	0.222	0.949	0.272	0.857	0.655	0.525	0.578	0.587	0.338
WSC		1.000	0.315	0.964	0.337	0.888	0.654	0.523	0.588	0.598	0.371
%WSC			1.000	0.386	0.740	0.491	0.191	0.144	0.255	0.268	0.414
CWS				1.000	0.510	0.971	0.683	0.574	0.686	0.701	0.485
%CWS					1.000	0.677	0.232	0.222	0.326	0.362	0.606
CWS-IWS						1.000	0.624	0.530	0.653	0.677	0.561
SEN							1.000	0.677	0.711	0.726	0.277
CC								1.000	0.663	0.670	0.225
TPM									1.000	0.979	0.429
CPM										1.000	0.443
Writing Quotient											1.000

Note: All correlations are significant at ≥.01. %CWS = percentage of correct word sequences; %WSC = percentage of words spelled correctly; CC = number of correct capitalizations; CPM = number of correct punctuation marks; CWS = number of correct word sequences; CWS-IWS = number of correct word sequences minus incorrect word sequences; SEN = number of sentences; TPM = number of punctuation marks; TWW = total words written; WSC = number of words spelled correctly.

Table 1 reveals the extreme skewness and kurtosis of these two variables. %WSC and %CWS were transformed, but results did not differ. Consequently, the untransformed variables were used for ease of interpretation.

When examining all 10 CBM indices as independent variables, extreme multicollinearity was present. Linear transformations of the variables were attempted; however, multicollinearity continued to exist. To reduce estimation problems, three of the predictor variables (i.e., words spelled correctly, correct word sequences, and total punctuation marks) were omitted from the study (Cohen et al., 2003; Morrow-Howell, 1994). These three predictor variables were chosen based on their high correlations with other predictor variables and past research showing more empirical support for the use of accuracy-based measures than fluency-based measures with secondary students (Jewell & Malecki, 2005; Tindal & Parker, 1989, 1991; Watkinson & Lee, 1992).

A simultaneous regression analysis was conducted to explore the relationship between the CBM indices and the TOWL-3 Overall Writing Quotient. The seven predictor variables collectively accounted for 44% of the variance in students' Overall Writing Quotient scores, $F(7, 439) = 49.40$, $p < .001$. However, only three of the predictor variables significantly contributed to prediction of the Overall Writing Quotient scores. As seen in Table 3, percentage of CWS contributed the most unique variance ($\beta = .490$, $p < .01$), followed by the number of correct punctuation marks ($\beta = .296$, $p < .01$), and the number of correct capitalizations ($\beta = -.117$, $p < .05$). However, correct capitalization was probably being suppressed by another variable as evident by the negative sign of the regression weight and positive correlation ($r = .23$; Cohen et al., 2003; Tabachnick & Fidell, 2001).

Table 3. Summary of CBM Index Scores as Predictors of Overall Writing Quotient Scores

Variable	*B*	*SE B*	β	*t*	*p*
Intercept	49.301	15.766			
TWW	0.018	0.101	.022	0.176	.860
%WSC	–0.257	0.186	–.074	–1.379	.169
%CWS	0.689	0.143	.490	4.813	.001
CWS-IWS	0.094	0.113	.146	0.836	.403
SEN	–0.446	0.404	–.067	–1.103	.270
CC	–0.521	0.233	–.117	–2.237	.026
CPM	1.199	0.254	.296	4.725	.001

Note: %CWS = percentage of correct word sequences; %WSC = percentage of words spelled correctly; CC = number of correct capitalizations; CPM = number of correct punctuation marks; CWS-IWS = number of correct word sequences minus incorrect word sequences; SEN = number of sentences; TWW = total words written.

Discussion

Assessment measures are commonly used to make important decisions regarding students' lives. Although there are several measures available for assessing writing skills, many are subjective, difficult to score, and time consuming (Watkinson & Lee, 1992). For these reasons, researchers continue to explore more straightforward, time-efficient, and informative methods of writing assessment, such as CBM writing indices. If valid, school personnel can utilize these CBM scores for assessing students' writing skills. Thus, criterion-related evidence for the validity of CBM writing indices was gathered by examining the ability of these indices to predict scores from a well-designed measure of written expression, the TOWL-3.

Past research has suggested that CBM indices must measure more complex skills in older students' writing than the traditional total words written index found sufficient at the elementary school level. However, as indices become more complex, interscorer reliability may suffer. In the present study, all CBM indices, regardless of the level of complexity, were scored reliably following training. However, it is important to note that although interscorer reliability was high, scorers received extensive training and monitoring throughout the scoring period. As the indices become more complex, more time is needed in training and monitoring of scorers and in scoring the index itself.

The results of the present study provide limited support for the use of CBM indices in written expression with secondary students. The seven CBM indices together accounted for only 44% of the variance in TOWL-3 scores. Moreover, only three of the seven indices uniquely contributed to the prediction of TOWL-3 scores: percentage of correct word sequences, correct punctuation marks, and correct capitalizations.

Past research has shown the percentage of CWS to serve as a moderate to strong indicator of writing performance for secondary students (Jewell & Malecki, 2005; Parker et al., 1991; Tindal & Parker, 1989; Watkinson & Lee, 1992). Tindal and Parker's initial factor analytic study suggested that the percentage of CWS represents a production-independent factor that measures accuracy and not fluency. As students' age increases, measures of writing accuracy may be more strongly related to students' performance on writing criteria than measures of writing fluency (Jewell & Malecki, 2005; Parker et al., 1991; Tindal & Parker, 1989; Watkinson & Lee, 1992). This is consistent with the present study in which percentage of CWS had the strongest bivariate correlation with TOWL-3 scores (r = .61) and contributed the most unique variance in the prediction of TOWL-3 scores (β = .49).

Although this study, along with past research, found the percentage of CWS to be a modest indicator of writing performance, this index should be used with caution for progress monitoring (Espin et al., 2000; Parker et al., 1991; Tindal & Parker, 1989). Percentage measures, including the percentage of CWS, are linear transformations of raw scores. Movement across percentage values depends on the number of opportunities the student has to respond. Thus, students' progress may be masked depending on the number of CWS the student has written. For example, a student can write 40 word sequences with 38 correct in the fall and 80 word sequences with 70 correct in the spring. Although this student's number of CWS has increased from 40 to 80, the percentage of CWS has decreased from 95% to 88%.

The second CBM scoring index that significantly contributed unique variance to the prediction of TOWL-3 scores was correct punctuation marks (β = .30). However, the correlation between correct punctuation marks and TOWL-3 scores was weak (r = .44) and similar to past studies examining the relationship between correct punctuation marks and standardized writing assessments (Gansle et al., 2002; Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006). It is important to note that a very limited number of studies have examined the use of correct punctuation marks (Gansle et al., 2002; Gansle et al., 2004; Gansle et al., 2006). In addition, these studies were conducted by the same researchers and the results of these studies have been inconsistent. Although the present study and Gansle et al.'s (2002) research found that the number of correct punctuation marks significantly contributed to the prediction of a writing criterion, Gansle et al. (2004) found that the number of correct punctuation marks did not significantly contribute to the prediction of *Woodcock Johnson–Revised* writing samples. However, given the good interscorer agreement and its relative ease in scoring, correct punctuation marks may be a promising CBM index if more research supports its reliability and validity.

The last CBM index to significantly contribute to the prediction of TOWL-3 scores was correct capitalizations. However, this should be viewed with caution. It is probable that another predictor variable was a negative suppressor for correct capitalizations. Thus, the relationship between one of the predictor variables and correct capitalization is hiding the real relationship between correct capitalization and TOWL-3 scores. This is largely evident by the negative sign of the regression weight of correct capitalization (β = –.12), which is opposite of what would be expected because correct capitalization and TOWL-3 scores are positively correlated (r = .23; Cohen et al., 2003; Tabachnick & Fidell, 2001). It is important to note that the weak bivariate correlation between correct capitalization and TOWL-3 scores was consistent with past research, indicating that correct capitalization may not be a robust indicator of written expression as measured by different writing criteria. Also, the number of correct capitalization can be highly dependent on what students write. For example, one student may choose to write a story referring to several individuals or places using proper names, whereas another student may write about an event that does not provide opportunities for correct capitalizations (Gansle et al., 2006). Therefore, the number of correct capitalizations should not be used to measure students' writing ability, unless future research finds additional supportive evidence.

As with all research, there are limitations of this study that require discussion. First, multicollinearity was present among the 10 CBM indices that were collected from the writing samples. To reduce estimation problems, three of the predictor variables (i.e., words spelled correctly, correct word sequences, and total punctuation marks) were omitted from the study. These three predictor variables were chosen based on their high correlations with other predictor

variables (i.e., .964 to .871) and past research. Secondary school studies have shown more empirical support for the use of accuracy-based measures (i.e., percentage of WSC and percentage of CWS) than fluency-based measures (i.e., WSC and CWS; Jewell & Malecki, 2005; Tindal & Parker, 1989, 1991; Watkinson & Lee, 1992). Thus, the two fluency-based measures, WSC and CWS, were removed from the analyses, and their related indices, percentage of WSC and percentage of CWS, were retained in the analyses. As for total punctuation marks, it was highly related to correct punctuation marks (r = .979), which also takes into consideration the quality of student writing.

A second limitation to the study is that all students who participated were from one grade level in a single school district in New Jersey. Thus, generalizations beyond this sample might be limited. Replication of this study with students from different grade levels and from different regions of the country would allow for a more constructive examination of the research question raised in this study. Specifically, future research needs to explore grade-level trends to further delineate which CBM indices are most appropriate to use with different grade levels.

Finally, this study did not investigate the suitability of any of the 10 CBM indices for progress monitoring, program evaluation, eligibility decisions, or skill diagnosis. In the current study, both the CBM and the criterion measure were given at one point in time, only providing evidence of criterion-related validity. Future research needs to examine the utility of CBM indices for a variety of academic purposes. In addition, when research examines the application of CBM indices, different grade levels should be used to further determine the reliability and validity of CBM indices in written expression.

Conclusion

The present study examined the relationship between CBM scores and a well-designed standardized test of writing, the TOWL-3. Results of this study were consistent with past research that showed simple fluency measures, such as total words written, are not adequate for assessing secondary students' writing ability. A more complex fluency measure, the number of correct punctuation marks, and an accuracy-based measure, the percentage of CWS, were the best predictors of TOWL-3 scores for these eighth-grade students. Of these two CBM indices, the percentage of CWS contributed the most unique variance in predicting TOWL-3 scores and had a moderate bivariate correlation with TOWL-3 scores. However, it is important to note that the percentage of CWS accounted for only 37% of the variance in TOWL-3 scores. In addition, past research has not shown the percentage of CWS to be adequate for progress monitoring (Espin et al., 2000; Parker et al., 1991; Tindal & Parker, 1989).

Results of the current study, along with past research, provide limited support for the use of existing CBM indices to assess and monitor writing skill at the secondary level. Although future research may identify a CBM index that yields reliable and valid scores and that can be used to monitor progress, current research suggests that existing CBM measures are unlikely to fulfill this need. At this point in time, educators may wish to rely on other qualitative and quantitative aspects of student writing, including published norm-referenced tests, for instructional or high-stakes decisions.

Authors' Note

This article is based on a dissertation completed by the first author and supervised by the second author. The authors are grateful for the guidance of Professors Liza Conyers, Richard Kubina, and Barbara Schaefer. Janelle Amato is now affiliated with the Cresskill School District.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Barenbaum, E., Newcomer, P., & Nodine, B. (1987). Children's ability to write stories as a function of variation in task, age and developmental level. *Learning Disability Quarterly*, *7*, 19–29.

Bradley-Johnson, S., & Lesiak, J. L. (1989). *Problems in written expression: Assessment and remediation*. New York: Guilford.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cole, J. C., Haley, K. A., & Muenz, T. A. (1997). Written expression reviewed. *Research in the Schools*, *4*, 17–34.

Cole, J. C., Muenz, T. A., Ouchi, B. Y., Kaufman, N. L., & Kaufman, A. S. (1997). The impact of the pictorial stimulus on the written expression output. *Psychology in the Schools*, *34*, 1–9.

DeGroff, L. C. (1987). The influence of prior knowledge on writing, conferencing, and revising. *The Elementary School Journal*, *88*, 105–118.

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

Deno, S. L., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure*

achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. Abstract retrieved February 22, 2004, from ERIC database.

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*, 208–217.

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

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

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

Gansle, K. A., Noell, G. H., VanDerHeyden, A., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review*, *31*, 477–497.

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

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review*, *35*, 435–450.

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advances in curriculum-based measurement* (pp. 61–88). New York: Guilford.

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

Graham, S., & Harris, K. R. (1997). It can be taught but it does not develop naturally: Myths and realities in writing instruction. *School Psychology Review*, *26*, 414–415.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A report to Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

Grammar slammer. (1997). Retrieved August 8, 2004, from http://englishplus.com/grammar/

Hammill, D. D., & Larsen, S. C. (1996). *The Test of Written Language* (3rd ed.). Austin, TX: Pro-Ed.

Hidi, S., & McLaren, J. (1990). The effect of topic and theme interestingness on the production of school expositions. In H. Mandl, E. DeCorte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 1, pp. 295–308). Oxford, UK: Pergamon.

Hidi, S., & McLaren, J. (1991). Motivational factors and writing: The role of topic interestingness. *European Journal of Psychology of Education*, *6*, 187–197.

Hooper, S. R. (2002). The language of written language: An introduction to the special issue. *Journal of Learning Disabilities*, *35*, 2–6.

Hooper, S. R., Montgomery, J., Swartz, C., Reed, M. S., Sandler, A. D., Levine, M. D., et al. (1994). Measurements of written language expression. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 375–417). Baltimore: Brookes.

Houck, C. K., & Billingsley, B. S. (1989). Written expression of students with and without learning disabilities: Differences across the grades. *Journal of Learning Disabilities*, *22*, 561–572.

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. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, *15*, 255–266.

MacArthur, C., & Graham, S. (1987). Learning disabled students' composing with three methods: Handwriting, dictation and word processing. *Journal of Special Education*, *21*, 22–42.

Malecki, C. (2008). Best practices in written language assessment and intervention. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 477–488). Bethesda, MD: National Association of School Psychologists.

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

Marston, D., Lowry, L., Deno, S., & Mirkin, P. (1981). *An analysis of learning trends in simple measure of reading, spelling, and written expression: A longitudinal study* (Research Report No. 49). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. Abstract retrieved February 22, 2004, from ERIC database.

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

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language*, *25*, 431–444.

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

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.

Moran, M. R. (1981). Performance of learning disabilities and low achieving secondary students on formal features of a paragraph-writing task. *Learning Disabilities Quarterly*, *4*, 271–280.

Morrow-Howell, N. (1994). The M word: Multicollinearity in multiple regression. *Social Work Research*, *18*, 247–251.

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

Poteet, J. A. (1979). Characteristics of written expression of learning disabled and non-learning disabled elementary-school students. *Diagnostique*, *4*, 60–74.

Powell-Smith, K. A., & Shinn, M. R. (2004). *Administration and scoring of written expression curriculum-based measurement (WE-CBM) for use in general outcome measurement.* Retrieved July 18, 2004, from http://www.aimsweb.com/uploaded/files/scoring_wecbm.pdf

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston: Houghton Mifflin.

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

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins.

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

Tindal, G., & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research and Practice*, *6*, 211–218.

U.S. Department of Education, Institute of Education Sciences, & National Center for Education Statistics. (2003). *The nation's report card: Writing 2000* (NCES 2003-529). Washington, DC: Author.

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

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*, 85–120.

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

Wright, J. (1992). *Curriculum-based measurement: Directions for administering and scoring CBM probes in writing.* Retrieved July 18, 2004, from http://www.jimwrightonline.com/pdfdocs/cbmresources/cbmdirections/cbmwrit.pdf

About the Authors

Janelle M. Amato, PhD, is now affiliated with the Cresskill School District. Her research interests include curriculum-based measurement and academic interventions.

Marley W. Watkins, PhD, is a professor and director of training at Arizona State University. His current interests include professional roles and functions, academic and behavioral interventions, the psychometrics of assessment and diagnosis, computer applications, and consultation.