Psychology in the Schools, Vol. 40(4), 2003
Published online in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/pits.10096

DEVELOPMENTAL, GENDER, AND PRACTICAL CONSIDERATIONS IN SCORING CURRICULUM-BASED MEASUREMENT WRITING PROBES

CHRISTINE KERRES MALECKI AND JENNIFER JEWELL

Northern Illinois University

The present study focused on CBM written language procedures by conducting an investigation of the developmental, gender, and practical considerations surrounding three categories of CBM written language scoring indices: production-dependent, production-independent, and accurate-production. Students in first- through eighth-grade generated a three-minute writing sample in the fall and spring of the school year using standard CBM procedures. The writing samples were scored using all three types of scoring indices to assess the trends in scoring indices for students of varying ages and gender and of the time required to score writing samples using various scoring indices. With only one exception, older students outperformed younger students on all of the scoring indices. Although at the middle school level students' levels of writing fluency and writing accuracy were not closely associated, at the younger grade levels the CBM indices were significantly related. With regard to gender differences, girls outperformed boys on measures of writing fluency at all grade levels. The average scoring time per writing sample ranged from $1\frac{1}{2}$ to $2\frac{1}{2}$ minutes (depending on grade level). © 2003 Wiley Periodicals, Inc.

Curriculum-based measurement (CBM), or standardized, short-duration fluency measures of academic skills (Shinn, 1995) are gaining prominence in the schools for use within a problem-solving framework. Specifically, CBM is being used in all steps of a problem-solving process including problem identification, instructional placement, goal-setting and intervention planning, progress monitoring, and eligibility decisions (Fuchs & Fuchs, 1997). The very nature of CBM makes it a useful assessment technology in an educational system as it can be used frequently and reliably, and provides practical information on which to base instructional decisions. However, most of the above conclusions are based on research that investigated reading or math CBM procedures.

CBM in Written Language

Researchers investigating CBM in the area of written language first focused on the psychometric properties of the assessment technique. This work has established adequate reliability and validity evidence for the use of CBM writing probes and scoring measures in a problem-solving model (Deno, Marston, & Mirkin, 1982; Deno, Mirkin, & Marston, 1980; Espin et al., 2000; Marston & Deno, 1981; Tindal & Parker, 1989; Videen, Deno, & Marston, 1982). Deno et al. (1980) were the first to investigate what types of measures could be used to routinely and appropriately assess students' writing performance. They found the production-dependent indices (dependent because they depend on how much the student writes) of Total Words Written, Words Spelled Correctly, Correct Letter Sequences, and Mature Words to be valid measures of students' writing ability. Specifically, these scores were significantly correlated with criterion measures such as published standardized tests ($r = .41$ to $.88$). Marston and Deno (1981) found these same scoring indices to be reliable and stable indicators of writing performance. Furthermore, Videen et al. (1982) examined Correct Writing Sequences and defined it as "two adjacent, correctly spelled words that are acceptable within the context of the phrase to a native speaker of the English language." Correct Writing Sequences, which takes into account grammar, capitalization,

Correspondence to: Christine Malecki, Northern Illinois University, Department of Psychology, DeKalb, IL 60115; cmalecki@niu.edu.

punctuation, spelling, etc., was found to be both a reliable and valid method for assessing students' writing proficiency. The production-independent measures (which do not depend on how much the student writes) of Percentage of Words Spelled Correctly and Percentage of Correct Writing Sequences were investigated by Tindal and Parker in 1989. Tindal and Parker found these scoring indices to be reliable and valid with a sample of middle school students. Recently Espin et al. (2000) examined the use of an accurate-production index, Correct Minus Incorrect Writing Sequences, and found that it was also reliable and valid to use with middle school students.

In addition to reliability and validity research, a trend has been identified in the literature suggesting that certain types of CBM writing scores are more appropriate for use with students of certain ages (Espin et al., 2000; Jewell & Malecki, 2003; Tindal & Parker, 1989). In other words, the relationship between students' CBM scores and scores on other writing criterion, such as published standardized tests, changes with age. For example, Tindal and Parker (1989) and Watkinson and Lee (1992) found that at the middle school level, production-independent scores were strongly related to teachers' holistic ratings, whereas production-dependent scores were only moderately related. Similarly, Espin et al. (2000) reported that Correct Minus Incorrect Writing Sequences (an accurate-production measure) was the most highly related to holistic judgements and scores on a district writing test when compared to middle school students' scores on production-dependent and production-independent measures. Jewell and Malecki (2003) also found that with students across grade levels, production-independent and accurate-production measures are generally more strongly related to other types of writing criterion than production-dependent measures. This holds true even at the younger grade levels where production-dependent measures are typically significantly related to the writing criterion. For instance, Deno et al. (1980, 1982) and Videen et al. (1982) reported that production-dependent indices were valid measures of elementary students' writing skills. Thus, it appears that most of the CBM scoring indices are valid to use with younger students. However, with older students, the production-independent and accurate-production indices are typically more related to students' writing abilities.

A recent study has identified gender differences in CBM fluency scores at all ages (Jewell & Malecki, 2003). Specifically, girls were more fluent than boys, as measured by higher average scores on the production-dependent measures (Total Words Written, Words Spelled Correctly, and Correct Writing Sequences). No differences were found between boys and girls in their performance on the accurate-production and production-independent measures. This result suggests that boys and girls differ in the amount that they write in a given time limit, but their writing accuracy is not significantly different. Thus, both grade level and gender of students must be considered when deciding which CBM scoring index is the most appropriate to use.

In addition to the grade and gender of the student, the *purpose* of the assessment also needs to be taken into account when deciding which scoring index to use. For instance, some researchers promote using fluency measures over "percentage" indices when monitoring student progress (Espin et al., 2000; Tindal & Parker, 1989). However, the sensitivity of the various CBM measures to student progress over time has not been clearly delineated. Because the present study collected writing samples from students at two separate time points, the growth of the CBM scoring indices over time can be examined.

Finally, the issue of practicality may be the most compelling consideration for practitioners and educators in deciding which CBM writing indices to use. Although writing probes can be administered in a group setting and take only three minutes, it takes time to score writing samples. From the authors' experiences, many schools and school districts are collecting normative data only on Total Words Written and Words Spelled Correctly. These fluency measures are the easiest to score and can be scored rather quickly. However, due to the research suggesting that Correct Writing Sequences and production-independent scores may be more useful in certain cases (especially with older students), it is worth considering the increased time needed to score writing samples using these more complex scoring indices.

Research Goals

The present study aims to broaden the research on CBM written language procedures by conducting an investigation of the developmental, gender, and practical considerations surrounding the three categories of CBM written language indices discussed above: production-dependent indices (Total Words Written, Words Spelled Correctly, Correct Writing Sequences), production-independent indices (Percentage of Words Spelled Correctly and Percentage of Correct Writing Sequences), and an accurate-production indicator (Correct Minus Incorrect Writing Sequences). The present study investigated which scoring indices are appropriate for use with early elementary, elementary, and middle school boys and girls. In addition, the present study investigated the time it takes to score writing samples using various scoring indices in order to provide practical recommendations regarding how to choose the appropriate scoring indices.

Method

Participants

The participants in this study were 946 first- through eighth-grade students from three school districts and five separate schools in suburban and rural northern Illinois. The participants consisted of 133 (14.1%) first-graders, 200 (21.1%) second-graders, 168 (17.8%) third-graders, 192 (20.3%) fourth-graders, 127 (13.4%) fifth-graders, 57 (6.0%) sixth-graders, 44 (4.7%) seventh-graders, and 25 (2.6%) eighth-graders. The sample consisted of 48% males and 51% females with 1% ($n = 6$) having missing gender data. No other demographic data were collected from students. Participation by students was voluntary and parental consent was obtained from each participant's parent or guardian.

Procedure

The researchers provided schools a description of the study as well as the opportunity to gather CBM writing data on their students. Student recruitment was school-wide, thus, consent letters were sent home with every student in the school. Those students in participating classrooms who had obtained parental consent provided a three-minute writing sample in response to a story starter in the fall and spring of the school year. Typically in individual assessments, three writing samples are collected from a student. However, given the large number of writing samples being collected across participants and the aggregate nature of the data, only one writing sample was collected from each participant. Teachers reported that the administration process (passing out papers, answering questions, monitoring the students while they wrote, and collecting papers) took approximately 10 minutes. Specifically, classroom teachers passed out lined paper with a story starter typed at the top of the page (a different story starter was used in the fall and spring). Students used the writing utensil they typically used for writing assignments in their classroom (pen or pencil). The students were given the following directions:

> You are going to write a story. First I will read the story starter at the top of your paper and you will write a story about it. You will have one minute to think about what you will write, and three minutes to write your story. Remember to do your best work. If you don't know how to spell a word, you should guess. Are there any questions? (Pause—respond to questions). Put your pencils down and listen. For the next minute, think about . . . (insert story starter and let the students think for one minute). Now please begin writing. (Students write for three minutes). Please stop and put your pencils down.

Measures

The three-minute writing samples were scored in the following methods:

Production-dependent indices.

1. *Total Words Written (TWW):* Any word-like string of letters with space before and after it was counted as a word. In this instance, the "word" did not have to be spelled correctly nor did it have to be an actual English word. Scorers counted and recorded the number of words written.
2. *Words Spelled Correctly (WSC):* Words that were spelled correctly in the English language were counted. If the word was used incorrectly in context (e.g. bear instead of bare), it was still counted as correct if it could stand alone as a correctly spelled English word. Scorers counted and recorded the number of words spelled correctly.
3. *Correct Writing Sequences (CWS):* Writing sequences (two adjacent writing units) were scored according to their correctness in context. Several things were taken into account including proper sentence-ending punctuation, capitalization of the first word in a sentence and proper nouns, commas in a sequence, spelling, correct use of pronouns, contractions, and possessives, and syntactical and semantic correctness (Videen et al., 1982). Scorers counted and recorded the number of correct writing sequences and the number of possible correct writing sequences.

Accurate-production indicator.

4. Correct Minus Incorrect Writing Sequences (CMIWS): This index was calculated by subtracting the number of incorrect writing sequences from the total number of correct writing sequences (CWS; Espin et al., 2000).

Production-independent indices.

5. Percentage of Words Spelled Correctly (%WSC): This measure is equal to the number of words spelled correctly (WSC) divided by the total number of words in the sample (TWW).
6. Percentage of Correct Writing Sequences (%CWS): This indicator was calculated by dividing the number of correct writing sequences (CWS) by the total number of possible writing sequences in the sample.

After the administration, the writing samples from the three schools were scored. Each student was assigned an ID number and the names on the samples were removed to ensure the students' confidentiality. One graduate and seven undergraduate students were taught the proper procedure for scoring the writing samples using the measures described above. The scorers were blind to the purposes of the study and were given no indication of the gender of the student participants. The students were trained and practiced scoring several writing samples. Following their training, students scored twenty writing samples per Grades 1–8 each from the fall and the spring samples (320 samples). Using these double scored samples, inter-rater reliability coefficients were calculated for the scoring indices of Total Words Written, Words Spelled Correctly, and Correct Writing Sequences. The coefficients were extremely high for Total Words Written (.995 to .999), Words Spelled Correctly (.994 to .998), and Correct Writing Sequences (.983 to .995). Reliability coefficients were not calculated for Correct Minus Incorrect Writing Sequences, Percentage of Words Spelled Correctly, or Percentage of Correct Writing Sequences because each of these scoring indices is imbedded in the other three scoring indices.

In addition, three students timed how long it took them to score each of 20 writing samples per grade (first through eighth) using all three methods. They started a stopwatch before starting each scoring and stopped the stopwatch and recorded the time it took them after writing down the score. The three timers' scores were averaged to create the average time taken to score writing samples from each grade. For purposes of reducing the number of analyses conducted, for many analyses students in various grade levels were grouped as follows: first- and second-graders as "early elementary students", third- through fifth-graders as "elementary students", and sixth- through eighth-graders as "middle school students."

Results

Because of the number of analyses conducted and a desire to reduce Type I error, a more conservative critical value ($p < .01$) was set as the criterion for significance for all analyses.

Grade Level and Gender Differences in CBM Indices

A 2 (gender) × 3 (grade level) multivariate analyses of variance (MANOVA) was conducted with fall TWW, WSC, CWS, %WSC, %CWS, and CMIWS scores as dependent variables. The intercorrelations of these dependent variables ranged from .56 to .99. A main effect was found for grade level, Wilks' lambda = .423, $F(12, 1858) = 83.10$, $p < .001$ (partial $\eta^2 = .349$, small effect size). Results of follow-up univariate analyses indicated that there were significant differences between the grades in all six scoring indices (TWW, WSC, CWS, %WSC, %CWS, and CMIWS), Fs (2, 934) = 393.65, 444.22, 451.27, 196.83, 369.24, and 343.45, respectively, ps $< .001$ (partial η^2s = .457, .487, .491, .297, .442, .424, respectively, small to medium effect sizes). Follow-up Scheffe contrasts found significant differences (ps $< .001$) among students at all three grade levels (early elementary, elementary, and middle) within the six measures with two exceptions. The %WSC and %CWS scores were not significantly different between elementary and middle school students. In all significant comparisons, middle school students' scores were higher than elementary and early elementary students' scores, and elementary students' scores were higher than early elementary students' scores. See Tables 1 and 2 for descriptive data of all CBM scores by grade and grade level.

Similarly, a gender main effect was found, Wilks' lambda = .923, $F(6, 929) = 12.96$, $p < .001$ (partial $\eta^2 = .077$, very small effect size). Results of follow-up univariate analyses indicated that there were significant differences between boys and girls on all scoring indices, Fs (1, 934) = 49.57, 51.23, 48.68, 18.00, 7.45, and 27.52 respectively, ps $< .001$ (partial η^2s = .050, .052, .050, .019, .008, .029, respectively, very small effect sizes). In all cases, girls outperformed boys. Table 3 contains descriptive data for the CBM indices by gender.

Finally, a significant gender × grade level interaction was found, Wilks' lambda = .939, $F(12, 1858) = 4.99$, $p < .001$ (partial $\eta^2 = .031$, small effect size). Results of follow-up univariate analyses indicated that this interaction was present for CWS, %WSC, and CMIWS, Fs (2, 934) = 5.54, 9.07, and 5.79 respectively, ps $< .004$ (partial η^2s = .012, .019, .012, respectively, very small effect sizes). Figure 1 illustrates the interactions showing that on CWS, girls' scores are higher than boys' and this gap grows over time. This pattern is similar for the CMIWS score. The %CWS score shows girls outperforming boys in early elementary, but the gap closes in elementary and middle school.

Grade Level and Gender Differences in Scoring Time

A 2 (gender) × 3 (grade level) MANOVA was conducted with the three scoring time scores as dependent variables. The scoring times had intercorrelations ranging from .71 to .82. No gender main effect was found, Wilks' lambda = .958, $F(3, 152) = 2.21$, $p = .09$. Additionally, there was no significant gender × grade level interaction, Wilks' lambda = .928, $F(6, 304) = 1.94$, $p = .07$. Only a grade level main effect was found, Wilks' lambda = .794, $F(6, 304) = 6.20$, $p < .001$ (partial $\eta 2 = .109$, small effect size). Results of follow-up univariate analyses indicated that there were significant differences among the grade levels on all three scoring time scores (TWWtime, WSCtime, and CWStime), Fs(2,154) = 6.57 ($p < .01$), 5.54 ($p < .01$), and 15.42 ($p < .001$),

Table 1
Descriptive Data on Fall and Spring CBM Scores by Grade

		TWW *M* (*SD*)	WSC *M* (*SD*)	CWS *M* (*SD*)	%WSC *M* (*SD*)	%CWS *M* (*SD*)	CMIWS *M* (*SD*)
Grade 1	Fall	7.56	4.81	2.05	52.94	21.56	−3.84
	(n = 133)	(4.71)	(3.93)	(2.49)	(31.63)	(23.17)	(4.32)
	Spring	13.71	10.24	6.95	70.90	42.18	−1.23
	(n = 123)	(7.40)	(7.11)	(6.15)	(21.23)	(25.46)	(7.36)
Grade 2	Fall	23.98	19.74	14.81	80.88	53.62	3.14
	(n = 200)	(9.91)	(9.50)	(10.06)	(14.38)	(22.02)	(13.39)
	Spring	30.52	26.97	24.00	86.89	68.87	14.42
	(n = 156)	(12.18)	(12.06)	(12.79)	(10.78)	(18.13)	(14.34)
Grade 3	Fall	35.79	32.38	28.04	88.96	69.61	17.01
	(n = 168)	(12.59)	(12.85)	(14.13)	(10.56)	(20.52)	(18.09)
	Spring	35.90	32.62	30.79	90.92	78.11	22.44
	(n = 109)	(11.84)	(11.83)	(12.72)	(11.96)	(16.69)	(16.53)
Grade 4	Fall	40.59	38.16	37.61	93.63	83.72	30.69
	(n = 192)	(11.49)	(11.68)	(12.89)	(8.59)	(12.29)	(15.07)
	Spring	46.30	43.73	42.23	94.23	83.25	33.93
	(n = 182)	(15.86)	(15.35)	(15.90)	(5.15)	(12.09)	(16.92)
Grade 5	Fall	50.87	48.01	46.31	94.08	82.94	37.20
	(n = 127)	(17.18)	(16.62)	(17.54)	(5.69)	(14.58)	(20.29)
	Spring	67.15	64.94	63.02	96.76	86.46	52.78
	(n = 120)	(23.92)	(23.03)	(22.94)	(3.99)	(12.55)	(24.00)
Grade 6	Fall	43.67	41.86	40.79	94.21	84.19	33.70
	(n = 57)	(12.96)	(12.56)	(13.80)	(13.40)	(18.18)	(18.02)
	Spring	57.63	55.61	53.89	95.89	85.07	45.30
	(n = 54)	(14.44)	(15.17)	(16.51)	(6.27)	(12.86)	(19.87)
Grade 7	Fall	51.34	49.41	47.98	96.05	84.05	39.34
	(n = 44)	(14.53)	(14.60)	(16.69)	(4.70)	(13.11)	(19.75)
	Spring	57.82	55.73	54.68	95.86	85.66	46.41
	(n = 44)	(16.79)	(17.28)	(19.40)	(6.48)	(12.88)	(22.62)
Grade 8	Fall	73.64	70.24	70.20	95.71	87.57	60.96
	(n = 25)	(22.69)	(22.25)	(25.02)	(7.95)	(10.64)	(27.99)
	Spring	66.96	65.67	66.54	97.98	91.01	60.29
	(n = 24)	(20.56)	(20.91)	(23.25)	(2.49)	(9.78)	(26.07)

respectively (partial η^2s = .079, .067, .167, respectively, very small effect sizes). Follow-up Scheffe contrasts found that it took significantly less time to score early elementary samples on TWW than to score TWW for middle school students ($p < .001$). Likewise, regarding WSC, early elementary writing samples took significantly less time to score than middle school samples ($p < .01$). Finally, using CWS, early elementary students' samples took significantly less time to score than either elementary or middle school students' samples (ps $< .001$). Descriptive data for these analyses are presented in Table 4.

CBM Score Intercorrelations

To investigate the interrelationships among the six CBM written language scoring indices, correlational analyses were conducted. Table 5 presents the results. Most of the scores were highly

Table 2
Descriptive Data on Fall and Spring CBM Scores by Grade Level

		TWW *M* (*SD*)	WSC *M* (*SD*)	CWS *M* (*SD*)	%WSC *M* (*SD*)	%CWS *M* (*SD*)	CMIWS *M* (*SD*)
Early Elementary	Fall	17.42	13.77	9.71	69.72	40.81	0.35
	(*n* = 333)	(11.51)	(10.67)	(10.11)	(26.63)	(27.41)	(11.25)
	Spring	23.11	19.60	16.48	79.84	57.10	7.52
	(*n* = 279)	(13.29)	(13.13)	(13.41)	(18.05)	(25.37)	(14.11)
Elementary	Fall	41.61	38.74	36.58	92.14	78.65	27.67
	(*n* = 487)	(14.75)	(14.78)	(16.25)	(9.00)	(17.39)	(19.36)
	Spring	49.63	46.98	45.27	94.09	82.82	36.39
	(*n* = 411)	(21.42)	(21.18)	(21.42)	(7.67)	(13.91)	(22.33)
Middle School	Fall	52.29	50.13	49.13	95.15	84.81	41.08
	(*n* = 126)	(19.29)	(18.74)	(20.55)	(10.05)	(15.19)	(23.13)
	Spring	59.53	57.63	56.66	96.29	86.45	48.65
	(*n* = 122)	(16.90)	(17.49)	(19.50)	(5.83)	(12.44)	(22.75)

related to one another with two notable exceptions. The production-independent measures (%WSC and %CWS) were not significantly related to Total Words Written at the middle school level. Thus, with the older students, how much students wrote was not closely associated to the accuracy of their writing. The CMIWS scores, however, did relate well with both the production-dependent

Table 3
Descriptive Data on Fall CBM Writing Scores by Grade Level and Gender

		TWW *M* (*SD*)	WSC *M* (*SD*)	CWS *M* (*SD*)	%WSC *M* (*SD*)	%CWS *M* (*SD*)	CMIWS *M* (*SD*)
Early Elementary	Boys	15.12	11.65	8.07	63.52	37.30	−0.23
	n = 162	11.34	10.44	9.31	31.76	28.19	9.58
	Girls	19.61	15.79	11.27	75.60	44.14	0.90
	n = 171	11.27	10.52	10.62	18.94	26.31	12.63
Elementary	Boys	38.60	35.53	33.11	91.19	77.21	24.16
	n = 243	14.02	13.89	14.84	9.93	17.80	18.03
	Girls	44.72	42.05	40.16	93.14	80.14	31.30
	n = 238	14.97	15.07	16.98	7.90	16.98	20.17
Middle School	Boys	45.00	42.96	41.14	93.51	82.69	33.55
	n = 51	17.51	16.96	17.37	14.08	16.96	18.38
	Girls	57.25	55.00	54.57	96.27	86.26	46.20
	n = 75	18.96	18.42	20.87	5.78	13.79	24.69
Total Sample	Boys	30.97	27.88	25.11	81.62	63.64	16.54
	n = 456	18.06	17.98	18.64	24.75	29.45	20.15
	Girls	37.79	34.78	32.19	87.43	68.37	22.87
	n = 484	20.23	20.49	22.62	15.49	27.20	25.29

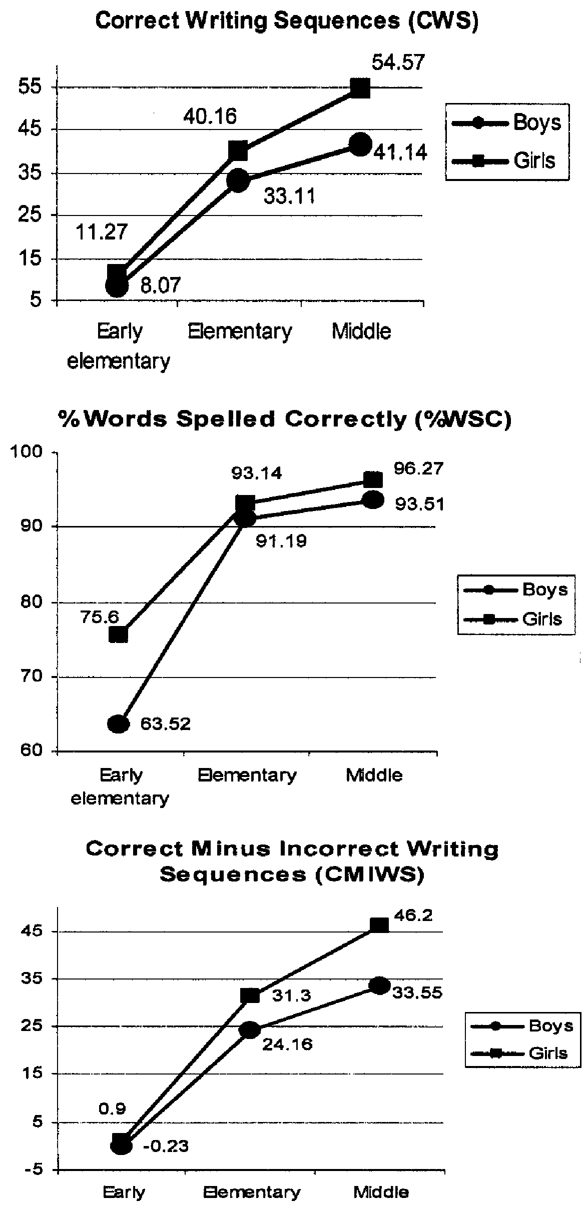

FIGURE 1. Significant gender × Grade level interactions on CWS, %WSC, and CMIWS.

and the production-independent indices at all grade levels, indicating that the CMIWS score is tapping aspects of both fluency and quality.

Change Over Time

A series of six repeated measures MANOVAs were conducted to determine if differences existed between students' fall and spring CBM writing scores by grade level. The results indicated

Table 4
Mean Scoring Time for TWW, WSC, and CWS by Grade Level

	TWW Seconds *M* (*SD*)	WSC Seconds *M* (*SD*)	CWS Seconds *M* (*SD*)	Total Scoring Time (Seconds) *M* (*SD*)
Early Elementary	21.69 (11.50)	25.39 (18.98)	46.25 (34.15)	93.33 (59.31)
Elementary	27.94 (13.24)	31.95 (15.39)	74.31 (27.11)	134.21 (51.69)
Middle	31.45 (13.83)	37.25 (17.18)	82.02 (34.68)	150.72 (61.67)

Note. Early Elementary $n = 40$, Elementary $n = 60$, Middle $n = 60$. Total Scoring Time Range (in seconds) for Early Elementary: 6.67 to 238.67; Elementary: 44.33 to 343.00; Middle: 2.00 to 358.67.

that there were significant differences between students' fall and spring scores in early elementary, elementary, and middle school on all fluency measures including TWW (Wilks' Lambda = .866, $F(1,809) = 124.79$, $p < .001$), WSC (Wilks' Lambda = .860, $F(1,809) = 131.23$, $p < .001$), and CWS (Wilks' Lambda = .854, $F(1,809) = 138.16$, $p < .001$). That is, at all grade levels and for all production-dependent indices, students' scores were higher in the spring than in the fall (partial η^2s = .134, .140, .146, respectively, small effect sizes). In addition, CMIWS scores were significantly higher from fall to spring for all grade levels (Wilks' Lambda = .887, $F(1,809) = 103.33$, $p < .001$, partial $\eta^2 = .113$, small effect size). No grade level interactions were present for any of these scores indicating that the change in scores from fall to spring was similar across grade levels.

Table 5
Intercorrelations Among the CBM Indices

	TWW	WSC	CWS	CMIWS	%WSC	%CWS
TWW						
Early Elementary		.96**	.85**	.40**	.53**	.54**
Elementary		.98**	.89**	.67**	.29**	.29**
Middle		.98**	.93**	.76**	.19	.21
WSC						
Early Elementary			.93**	.57**	.63**	.65**
Elementary			.94**	.77**	.46**	.42**
Middle			.95**	.81**	.29**	.29**
CWS						
Early Elementary				.82**	.58**	.77**
Elementary				.93**	.53**	.65**
Middle				.94**	.30**	.49**
CMIWS						
Early Elementary					.42**	.76**
Elementary					.64**	.85**
Middle					.35**	.68**
%WSC						
Early Elementary						.77**
Elementary						.75**
Middle						.71**

** $p < .01$.

However, regarding the percentage indices (%WSC and %CWS), a significant grade level interaction was present for %WSC (Wilks' Lambda = .939, $F(2,809) = 26.23$, $p < .001$) and for %CWS (Wilks' Lambda = .884, $F(2,809) = 53.03$, $p < .001$). In both cases, there was a significant difference between fall and spring scores only at the early elementary level (partial η^2s = .061 and .116, respectively, very small effect sizes). No significant differences were found over time on the percentage scores at the elementary or middle school grades.

Discussion

The current study adds to the literature by providing data on six CBM writing scores from students in first through eighth grade at two time points (fall and spring). As expected, elementary students scored better than early elementary students on all scoring indices (TWW, WSC, CWS, %WSC, %CWS, and CMIWS). Furthermore, middle school students scored better than elementary level students on almost all scoring indices (TWW, WSC, CWS, and CMIWS). At the early elementary and elementary grade levels the production-dependent (fluency), production-independent (accuracy), and accurate-production (accurate fluency) writing measures are all significantly related. At the middle school level, however, the measures of writing accuracy are not significantly related to how much students wrote. Thus, at the older grade levels, students' levels of writing fluency and accuracy are not closely associated. This is an important result that strengthens findings in recent research in this area (Espin et al., 2000; Jewell & Malecki, 2003). Many practitioners are using only TWW or WSC (fluency measures) at many stages of problem solving, even with older elementary and middle school students. The present study confirms previous findings that at all grade levels, and particularly with older grades, the accuracy measures (production-independent) or the accurate-production measure (CMIWS) should be strongly considered as the most appropriate scores to use. Of course, if pure fluency is the goal of an intervention or is of specific concern, then the fluency measures should be used.

Jewell and Malecki (2003) found that girls outperform boys on CBM writing fluency indices with students in second, fourth, and sixth grades. The present study found even more differences between first- through eighth-grade boys and girls. Girls were found to outperform boys on all six CBM writing scoring indices across grade levels. This finding is particularly important for school psychologists and educators to be aware of as many schools use normative data on the CBM scores for eligibility and screening purposes. These professionals need to be aware that girls may have an advantage on these indices, or that boys may be over-identified for difficulties in writing if normative data does not account for gender.

The present study found that at all grade levels, the measures of writing fluency and the accurate-production index increased significantly from fall to spring. In addition, at the early elementary level, the production-independent measures showed significant growth over time, suggesting that they would be appropriate for progress monitoring. Some researchers issue caution against using percentage indices to monitor student progress (e.g., Espin et al., 2000; Tindal & Parker, 1989); however, the present study found that these measures did show significant growth over a span of time for the early elementary students. This result needs to be interpreted cautiously because the percentage indices may not be as sensitive to student growth in the short-term as production-dependent indices. In addition, the present study contributes evidence that the percentage indices may not be as sensitive to growth over time as fluency measures at the older grade levels. As always, the assessment method chosen should be a good match with the target of the intervention (including the time span) and with the particular student.

On average, it took roughly 30 seconds to score a writing sample using either TWW or WSC (ranging from 22 to 37 seconds). It took slightly more time to score a sample using these two indices as the grade level of the student increased, but this increase was typically about 6 seconds between grade levels. Therefore, using only these two scoring indices, an early elementary student's writing sample could be scored in under a minute, and a middle school student's sample could be scored in just over a minute. When scoring a sample with CWS, the average time for the three grade levels ranged from 46 to 82 seconds, with the time increasing with grade level. Even at the oldest grade level, however, it took less than a minute and a half to score a sample using CWS, on average. Thus, even at the middle school level, the total time to score a writing sample using all three indices (TWW, WSC, and CWS) only averaged approximately $2\frac{1}{2}$ minutes. The total time to score an elementary student's sample was slightly under this time, and the average scoring time for an early elementary student's sample was only $1\frac{1}{2}$ minutes.

This study is the first to examine actual scoring times for CBM writing measures. This issue is of practical utility for school psychologists and educators because they can use this information to plan for scoring time. Knowing the amount of time required to score writing samples is particularly useful if a district or school is collecting normative data. Information on time can help them decide which scoring methods to use and how many scorers they may need to complete the scoring. Using the current study's timing data, if a district decided to collect 100 writing samples per grade in first through eighth grades and used all three core writing indices (TWW, WSC, and CWS), it would take approximately 30 hours to score all 800 writing samples. Thirty hours is a lot of time, however, schools can utilize the assistance of teacher's aides, parent volunteers, or paid substitute teachers to help score. In other words, although the amount of scoring may seem daunting, it is manageable.

Furthermore, the current study supports previous research that suggests that CWS is an important index of students' writing skills and this index is necessary to compute other valid indices such as %CWS and CMIWS (Espin et al., 2000; Jewell & Malecki, 2003; Videen et al., 1982). The timing data suggests that if practitioners are not collecting normative data on CWS due to anticipated scoring difficulties, their fears should be somewhat alleviated in that the increased time is not insurmountable. Furthermore, the extra time may be well worth it to increase the validity of the assessment.

One limitation of the current study is the lack of specific integrity data regarding how the writing samples were collected. The researchers examined the writing samples after they were collected to look for any problems that might have occurred in the administration of the probes. Although no problems (e.g., samples being too long or too short) were detected, because integrity data were not collected from teachers, it is not a guarantee that standardized procedures were followed in the collection of all writing samples. Another limitation is that other writing criterion measures were not available for all of the participants in the current sample. Had criterion data been available for all cases, a more comprehensive picture of how the scoring indices function could have been developed. Future research should take such a comprehensive approach and collect data from students across the elementary and secondary grades while collecting high stakes writing criterion measures to further examine the utility of CBM writing scores. Finally, although the overall sample size was adequate, a larger middle school sample would have been beneficial.

Summary

Along with previous supporting literature, results from the present study can inform school psychologists and educators with several practical guidelines. When conducting an assessment of a student, it is crucial to take into account the student's age and gender, and the purpose and duration of the assessment. Using these factors, a decision can be made about which of the three types of CBM writing measures (production-dependent, production-independent, or accurate-production) is the most appropriate to use when making educational decisions.

Production-dependent indices (Total Words Written, Words Spelled Correctly, and Correct Writing Sequences), or the fluency measures, are best to use with younger students, when writing fluency is the target of assessment or when monitoring student progress over time. These recommendations are based on the results of the present study and recent research in the field, indicating that fluency measures are significantly related to other writing criterion measures with younger students only; however, at all grade levels fluency measures showed significant growth over the course of the school year.

Production-independent measures (Percentage of Words Spelled Correctly and Percentage of Correct Writing Sequences), or measures of writing accuracy, should be the primary measures used when assessing the skills of older students. The writing accuracy measures, unlike the fluency measures, are significantly related to other writing criterion with older students. In addition, the percentage indices are an easy way to communicate the results of a writing assessment to parents and educators.

The relatively new accurate-production index (CMIWS) appears so far to be a good indicator of student performance at all grade levels, and for many purposes. The current study found students' scores on CMIWS to significantly increase from fall to spring, thus making it useful for monitoring student progress at all grade levels. Furthermore, it is significantly related to the measures of writing fluency and accuracy, and to other writing criterion.

In addition to the above recommendations, the student's gender needs to be considered due to the current study's findings that girls consistently outperform boys on the scoring measures. This result is important during an assessment such that boys do not become over-identified as having difficulty in the area of writing.

Lastly, the current study has demonstrated that the time required to score students' writing samples using a variety of scoring indices is manageable, ranging from $1\frac{1}{2}$ to $2\frac{1}{2}$ minutes per writing sample. The additional information gained from using measures of writing accuracy as well as measures of writing fluency insures that the time spent scoring the writing samples with the various types of indices is worthwhile.

References

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

Deno, S.L., Mirkin, P.K., & Marston, D. (1980). Relationships among simple measures of written expression and performance on standardized achievement tests (Rep. No. 22). Minneapolis, Minnesota.

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

Fuchs, L.S., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. Focus on Exceptional Children, 30, 1–16.

Jewell, J., & Malecki, C.K. (2003). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. Manuscript submitted for publication.

Marston, D., & Deno, S. (1981). The reliability of simple, direct measures of written expression (Rep. No. 50). Minneapolis, Minnesota.

Shinn, M.T. (1995). Identifying and defining academic problems: CBM screening and eligibility procedures. In M.R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 90–129). New York, NY: Guilford Press.

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 (Rep. No. 84). Minneapolis, Minnesota.

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.