



TECHNICAL REPORT #8:

Technical Features of Narrative vs. Expository and
Handwritten vs. Word-Processed Measures of Written
Expression for Secondary-Level Students

*Heather Campbell, Kristen L. McMaster, Sharon Lane-Getaz,
Kelsey Harr, Chris Miller, and Elizabeth Zahn*

RIPM Year 2: 2004 – 2005

Dates of Study: March 2005

October 2010

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA



Produced by the Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

Abstract

The purpose of this study was to extend current research on progress monitoring of written expression (CBM-W) for secondary-level students by comparing the technical features of several measures based on type of writing tasks, response mode, duration, and scoring procedure. One-hundred twenty-nine 10th-graders produced 3 to 7 min narrative and expository writing prompts using word processing or handwriting. Samples were scored quantitatively. Measures that yielded sufficient alternate-form reliability were examined to determine which had sufficient criterion validity, and those with sufficient criterion validity were examined to determine which best predicted results on a statewide, high-stakes writing test. Based on the results of this study, type of prompt (narrative vs. expository) seemed to matter less for technical adequacy than type of scoring procedure or response mode (handwritten vs. word-processing). For the handwriting response mode, calculating the percent of correct word sequences in 5 or 7 min was the most promising indicator for predicting failure on the Minnesota Basic Standards writing test (MBST). For the word-processed response mode, calculating correct minus incorrect word sequences for 5 or 7 min appeared most promising in terms of its ability to predict failure on the MBST.

Word Count: 191

Technical Features of Narrative vs. Expository and Handwritten vs. Word-Processed Measures of Written Expression for Secondary-Level Students

Curriculum-based measurement procedures, such as gathering oral reading fluency (ORF) scores several times per year, are widely used by teachers and district assessment coordinators to provide data for response-to-intervention models. A (slowly) growing body of research points to the reliability and validity of the use of curriculum-based writing measures (CBM-W) for both elementary and secondary students (McMaster & Espin, 2007) as well. However, CBM-W has not been as widely adopted. Why?

Perhaps one key lies in the time commitment. Collecting ORF data takes 1 minute per student, so a classroom teacher with 30 students may spend perhaps 40 minutes weekly collecting and scoring ORF measures. While administering CBM-W prompts is easily done in large groups and is time-efficient (perhaps up to 10 min total), scoring may not be. Early research showed that simple scoring procedures, such as counting the number of words written, were valid and reliable tools to use as writing indicators (Deno, Mirkin, & Marston, 1980; Deno, Mirkin, & Marston, 1982; Marston & Deno, 1981; Tindal, Marston, & Deno, 1983). However, later researchers could not replicate these findings, noting that more complex scoring procedures, such as counting the number of correct word sequences (CWS) were required to produce technically adequate measures, particularly for older students (see McMaster & Espin, 2007 for a discussion of these results).

Gansle and her colleagues (Gansle et al., 2002) investigated the use of CBM-W measures with suburban 3rd and 4th graders and found that it took an average of 1 min for an *expert* to score a 3-min writing sample for CWS. A new scorer may take up to 5 min to score a 7-min writing sample from a native English speaker. Scoring generally takes longer for the writing of students

with certain learning disabilities as well as English language learners, as their writing is frequently difficult to both decipher and understand and may lack basic grammatical structure. The teacher with 30 students in a class may spend 10 min giving the prompts and maybe more than 3 hours scoring; how many teachers would commit to 3-4 hours weekly of scoring and tracking writing prompts? If the technical adequacy was convincing and there was a better way to score and track CBM-W measures, teachers may embrace CBM-W as a way to determine whether or not their students are making gains in writing.

The purpose of this study was to extend current research on progress monitoring of written expression for secondary-level students by comparing the technical features of several measures based on types of writing prompt (narrative vs. expository), response modes (handwritten vs. word-processed), duration (3 to 7 min), and scoring procedures. Two research questions were addressed: (1) What is the alternate-form reliability and criterion validity for writing samples produced across varying time frames in response to different types of prompts using different response modes? (2) Which scoring procedures produce the most technically sound information about students' written expression performance *and* best predict performance on a high-stakes writing test?

Method

Setting and Participants

This study took place in an urban high school (serving students in grades 9-12) in a large Midwestern metropolitan school district. The school served approximately 1276 students. Of these students, 66.8% percent were from culturally or linguistically diverse backgrounds, 54.9% received free or reduced lunch, 11.0% received special education services, and 18.9 % received English language learner services (English as a second language: ESL). Participants in the study

were 129 students in eleven 10th-grade English classrooms and included 15.5% ESL-eligible students and an additional 13.2% of students who spoke a language other than English at home but were not receiving ESL services. Home languages included: English (71.3%), Somali (9.3%), Spanish (7.8%), Hmong (7.0%), and Arabic, Cambodian, Latvian, Shon, and Turkmanese (.8% each). Students were European-American (45%), African-American (32.6%), Asian-American (10.9%), Latino (9.4%), and American Indian (1.6%). 43% of participants were receiving free or reduced lunch. Students receiving special education services comprised 6.2% of the participants. 39.5% of the participants were female and 61.5% were male. 15% of the students were in a Pre-Advanced Placement English Class.

Independent Variables

Narrative vs. expository prompts. Because a large number of prompts (four narrative and four expository) were to be used in the study, it was important to ensure that the prompts would produce the appropriate types of writing. Researchers defined narrative prompts as writing prompts that would elicit creative story writing that is temporally organized (i.e., has a clear beginning, middle, and end). Expository prompts were defined as writing prompts that would elicit descriptive or informational writing that is not necessarily temporally organized. To help classify and choose prompts, 23 native English-speaking, graduate level students in special education were shown a list of ten possible narrative and expository prompts. They were asked to use the definitions to classify each prompt as narrative or expository. They were also asked to rank order the prompts from 1 to 10, with 1 being the prompt most likely to elicit the most writing from high school students, and 10 being the least likely to do so. The prompts that were most consistently identified as narrative or expository and were most frequently identified at the top of the rank ordered lists were selected for inclusion in the study. Narrative prompts included

the following (the percent of graduate student agreement that they were narrative is in parentheses):

P5. I was the first one to smell smoke and... (95%)

P6. The strangest thing happened to me when I was on my way home... (100%)

P7. I was talking with my friends when all of a sudden... (91%)

P8. I looked out the window and could not believe what I saw... (91%)

Expository prompts included the following (the percent of graduate student agreement that they were expository is in parentheses):

P1. I would like to take a trip to... (95%)

P2. You know a person is a good friend because... (91%)

P3. In my free time, I like to... (95%)

P4. If I were to describe my high school, I would say... (95%)

Handwritten vs. word-processed responses. Participants produced a total of eight writing samples: four handwritten and four word-processed. The handwritten responses were group-administered either in the students' classrooms or in the media center. Each prompt was printed at the top of a sheet of paper followed by lines printed on the same sheet. Students wrote their responses on the lines and were given extra sheets of lined paper if their writing took up more than one page. All students were instructed to write for 7 min and to make slashes at the 3- and 5-min marks. Appendix A includes directions that were given to the students for the handwriting task.

Word-processed prompts were completed in the school's media center. Each student typed their response on a Macintosh iBook, iMac, or G4 computer; all computers were running either OS9 or OS10. Students used the computers' Simple Text (OS9) or Text Edit (OS10)

programs to word process their responses; all editing preferences (including spelling and grammar check) were disabled for the duration of the testing. Responses were saved to a secure school server in folders designated for each English class. At the beginning of the first word-processed prompt response, students were instructed to type the phrase, “The quick brown fox jumped over the lazy dog” as many times as they could in one minute. The number of words typed per min was recorded to control for typing speed in statistical analyses. Appendix B includes directions that were given to the students for the word-processing task.

Scoring procedures. Writing samples were scored based on scoring procedures that have shown promise in previous progress monitoring research for secondary school students (see McMaster & Espin (2007) for a review of the literature). The writing samples were only scored for correct word sequences (CWS) and incorrect word sequences (IWS); additionally, correct minus incorrect word sequences (CIWS) and the percent of correct word sequences (%CWS) were calculated.

CWS (Videen, Deno, & Marston, 1982) are any two adjacent, correctly spelled words that are acceptable within the context of the sample to a native speaker of the English language (meaning that a native speaker would judge the word sequences as syntactically and semantically correct). CIWS (Espin, Scierka, Skare, & Halverson, 1999) are the number of correct word sequences minus the number of incorrect word sequences. %CWS (Tindal & Parker, 1989) is calculated by dividing the CWS by the total number of word sequences in the writing sample.

Criterion Variables

Classroom writing performance. Two measures were used to indirectly assess students’ writing performance in the classroom. First, teachers were asked to rate each student’s writing skills on a 5 point Likert-type scale (1 = less skilled, 5 = more skilled). The teachers were

encouraged to use the full range of ratings to describe their students. Second, students' end-of-year grade point averages and English grades were obtained from district records and reported based on a 4-point grade point average scale.

MBST Writing Subtest. The Minnesota Basic Standards Test (MBST; Minnesota Department of Education & Beck Evaluation and Testing Associates, 1997) in written expression is a high-stakes test required for graduation. It is untimed, administered annually in April to all 10th graders. The prompts used in the test are designed to elicit expository writing that does not require specialized knowledge. Past prompts have included: (1) What would be your dream job? Tell about that job and explain why it would be a good job for you. Include details so your readers will understand your choice; (2) Write about a time you learned something. Explain what you learned. Include enough details so your reader will understand the importance of this experience.

Student responses on the MBST are scored using a standardized rubric on a scale of 0 (low) to 6 (high); items on the rubric include clarity, focus, organization, detailed support or elaboration of ideas, and language conventions (Minnesota Basic Skills Technical Manual, 2001-2002). A score of 3 is necessary to pass; students who do not earn at least a 3 may retake the test until they pass. Each writing sample is scored independently by two raters (trained by NCS Pearson) using a partial scale of 0 (low) to 4 (high). If scores are within one point, the sample receives the average of the two scores. If scores are not within one point of each other, an "expert" scores the writing sample, and it receives the score assigned by the expert. Once the initial scoring is completed, all samples with a score of 4 are rescored as a 4 (competent), 5 (skillful), or 6 (exceptionally skillful).

Procedures

Test administration. The written expression measures were administered in the spring of 2005. Two trained graduate assistants administered the measures during the students' English class periods, either in small groups or to the entire class. The measures were administered in four sessions, each of which was two or three days apart. There were eight prompts total (four narrative and four expository) to which each student responded. For each condition (handwriting vs. word processing), students responded to two different expository and two different narrative prompts. All eight prompts were "available" in both the handwriting and word-processing conditions. Half of the students completed two handwritten prompts first; half completed two word-processed prompts first. The method of text production alternated each testing period. For example, during the first session, Class A was given two narrative prompts (prompt 1 and 2) and was instructed to handwrite their responses. Prompts were counterbalanced within the class such that half of the class responded to prompt 1 first and half responded to prompt 2 first. The next session, the students were given expository prompts 3 and 4 to word-process. The following week, the students were given narrative prompts 5 and 6 to word-process, and in the last session were given expository prompts 7 and 8 to handwrite.

Training and scoring reliability. Three graduate students were trained to score the handwritten samples and one graduate student scored the word-processed samples. The trainer met with the scorers in one two-hour session to demonstrate scoring and practice the procedures. After a review of the scoring rules, the scorers completed a guided-scoring task following the trainer's model. Then, each scorer practiced scoring a second writing prompt with the trainer's help.

After training, each scorer then scored one writing packet (which included all handwritten writing samples produced by one student) without assistance. Scorers who had less

than 80% agreement with the trainer on any scoring procedure received further training. Once 80% agreement was met for each scoring procedure for each writing prompt, the scorer was given additional packets of writing samples to score.

Ten percent (10%) of each scorer's writing packets were randomly selected and scored by the trainer to check for reliability. If reliability was below 80%, the trainer and the scorer met to discuss the discrepancy. If there were only a few discrepancies, the two came to agreement on the correct score. If there were several discrepancies, the entire packet was rescored and the scorer had to reach 80% reliability with the trainer again in order to continue scoring. In only one case did the scorer have to rescore the entire packet; most discrepancies were due to writing samples containing few words where one disagreement resulted in lower reliabilities.

Word processed samples were scored using WordMark software developed at the University of Minnesota (Hooper, Miller & Rose, 2005). The word-processed texts were imported into the software, and scorers assigned a correct word sequence or incorrect word sequence between each word, similar to the handwritten samples. The software then "counted" the number of correct word sequences and incorrect word sequences for that sample.

Statistical Analysis

In order to determine the technical adequacy of the CBM-W writing tasks (narrative vs. expository), response modes (handwriting vs. word-processing), duration (3, 5, and 7 min), and scoring procedures (CWS, CIWS, %CWS), alternate-form reliability, criterion validity, and odds failure rates were calculated. In order to find the most promising measures threshold values (as defined in the Results section) for each analysis were set; values below these thresholds were not considered for further analyses. Measures with sufficient alternate-form reliability coefficients were then examined to determine which had sufficient criterion validity coefficients. Measures

with sufficient reliability and criterion coefficients were examined to identify whether they predicted failure on the MBST, which was given 2 months after the CBM-W writing tasks.

Results

Demographic Predictors

The sample was predominantly Caucasian (45.3%) and African-American (32.8%), although students of Asian and Hispanic descent were also represented. Only 11 students (8.5%) in the sample failed the MBST. While only 32.8% of the sample was African-American, they comprised 63.6% of the sample who failed the test. African-American students had 4.15 times higher odds of failing (95% CI = 1.14 – 15.08, $p = 0.03$) compared to children of other races. 47.8% of students in the sample were enrolled in either free or reduced lunch programs. Those who received free lunches had over 9 times the odds of failing than children not in the free lunch program (95% CI = 1.88 – 44.23, $p < 0.01$). GPA and gender were not significantly associated with failing the exam.

Alternate-form Reliability

Means and standard deviations for scoring procedures were calculated (see Table 1), and Pairwise correlation coefficients were calculated to determine the alternate-form reliability of each scoring procedure implemented for each prompt type written in 3, 5, and 7 min. A Bonferroni correction was applied to all correlation analyses to control the type-I error rate for multiple comparisons. A p -value ≤ 0.002 is considered statistically significant with this correction applied. Based on criteria set in previous CBM-W technical adequacy studies (McMaster & Campbell, 2008) in which reliability coefficients of $r > .85$ were considered strong, $r = .70$ to $.80$ moderately strong, $r = .60$ to $.70$ moderate, and $r < .60$ weak, a threshold

was set for alternate-form reliability coefficients at $r \geq .60$ for inclusion in criterion validity analyses. These coefficients are highlighted in red in Table 2. In general, handwritten responses ($r = .56-.72$) had lower coefficients than word-processed responses ($r = .43-.86$), but coefficients did not vary much across prompt type (narrative $r = .56-.86$ and expository $r = .43-.81$). Most reliability coefficients increased with time. All scoring procedures used across response method and prompt type, except for several at the 3-min mark, met the reliability threshold coefficient. Those meeting the criteria were then considered in the validity analyses.

Criterion Validity

Means and standard deviations for the criterion measures (teacher ratings, language arts/English grades, and MBST scores) are listed in Table 3. While the participants in the study represented a diverse population in many ways, their MBST scores are quite restricted (note that only 11 out of 129 students failed to pass the MBST, and scores ranged from 2.0 to 4.0 in .5 unit increments).

Pairwise correlation coefficients were calculated to determine the criterion validity of each scoring procedure implemented for each prompt type written in 3, 5, and 7 min with teacher ratings, language arts/English grades and MBST scores (see Table 4). While the reliability coefficient threshold was set at $r \geq .60$, here, $r \geq .40$ was determined to be “sufficient” for inclusion in further analyses due to the restricted nature of the data (teacher ratings were assigned on a whole number scale of 0-5; MBST scores for study participants ranged from 2.0-4.0) and the relatively unknown performance of CBM-W word-processed prompts.

The *handwritten* responses that met the threshold validity criteria with both the MBST and teacher ratings were 5- and 7-min narrative and expository prompts scored for %CWS. The *word-processed* responses that met the criteria were narrative prompts at 3 min and expository

prompts at 7 min scored for CWS and 5- and 7-min narrative and expository prompts scored for CIWS. Neither language arts/English course grades nor GPA were significantly correlated to any CBM-W scores.

Predicting Failure on MBST

A series of generalized linear models were used to explore the odds of failure on the MBST (score less than 3) with student performance on the writing measures as explanatory variables. Multivariate logistic regression models were used to evaluate the validity of the various scoring measures in predicting the odds of failing the MBST for handwritten measures, controlling for ethnicity and eligibility for the free lunch program, which was used as an indicator for economic status. Multivariate logistic regression models were also used to evaluate the validity of the same scoring measures in predicting the odds of failing the MBST administered by computer, controlling for ethnicity, free lunch eligibility, and the typing ability of the student (words per minute).

The regression coefficients of logistic regression are odds ratios, which are an effect size that indicates the strength of an association between a binary outcome variable (e.g., success, “1,” and failure, “0” on an exam) and predictor variables. A significant odds ratio less than 1.0 reflects a protective effect against failing the exam, while an odds ratio greater than 1.0 reflects increased odds of failure. For binary explanatory variables, the odds ratio can be interpreted as a change in odds from the comparison group. For example, if an explanatory variable for gender (females = 0, males = 1) generated an odds ratio of 1.5 in a logistic model, the odds of the outcome would be 50% greater for males compared to females. For continuous explanatory variables, the odds ratio should be interpreted as an exponential increase in the odds. For example, if an explanatory variable for correct word sequences was associated with an odds ratio

of 0.95, the decrease in odds would be calculated as a 0.95^n change in the odds, where n is the difference in the points separating the comparison group from the group of interest. For example, if one wanted to calculate the difference in the odds of passing for groups of children who wrote 40 and 50 words correctly in a given assessment, the children who wrote 10 more words correctly would have 40% greater odds of passing the test, since $0.95^{10} = 0.60$, and the difference is calculated as $1.0 - 0.60$, which is 0.40.

The effectiveness of each measure was evaluated using Akaike's Information Criterion (AIC), which are goodness of fit measures for statistical models. AIC is calculated with the formula $AIC = 2k - 2\ln(L)$, where k is the number of explanatory variables in the model, and L is the maximized log-likelihood of the model. The log-likelihood essentially reflects the amount of unexplained variability in a model. The formula is meant to "penalize" a model for unexplained variability as well as the number of parameters, so a lower value of AIC reflects a superior balance of fit and parsimony in a statistical model. An AIC for a single model is meaningless; it is meant to be a measure for which to compare competing "non-nested" models (i.e., models with different explanatory variables). Since all logistic models include the same number of parameters, a lower value of AIC in this case simply reflects a greater amount of variance accounted for by a particular scoring measure. Therefore, the models with the lowest AIC should be considered as superior predictors of MBST failure. Statistical analyses were conducted within the R statistical package, version 2.9.2 (R Development Core Team, 2009).

As noted in Table 5, of the measures meeting both reliability and validity criteria, 5- and 7- min handwritten prompts scored for %CWS were the best predictors of MBST failure, as were 5- and 7-min word-processed prompts scored for CIWS. Type of prompt seemed not to appreciably impact AIC scores.

Discussion and Implications for Further Research

The purpose of this study was to extend current research on technical adequacy of written expression (CBM-W) for secondary-level students by comparing the features of several measures based on type of writing tasks, response mode, duration, and scoring procedure. Several findings confirmed the results of previous studies, and some new promising findings related to response method (handwriting vs. word-processing) were noted.

The first research question addressed the reliability and validity of CBM-W measures when prompt type, response method, and scoring procedures varied across time. As has been noted in several studies (e.g., Weissenburger & Espin, 2005; Espin, et al., 2008; McMaster & Campbell, 2008), increasing time generally strengthens the reliability coefficients. Overall, there appeared to be little difference in reliability or validity for type of prompt (narrative vs. expository), a finding consistent with previous research (e.g., Espin et al., 2000, McMaster & Campbell, 2008).

Interestingly, there were significant differences noted for prompt type and scoring procedure *within* response method. The *handwritten* responses that met the threshold validity criteria with both the MBST and teacher ratings were 5- and 7-min narrative and expository prompts scored for %CWS. The *word-processed* responses that met the criteria were narrative prompts at 3 min and expository prompts at 7 min scored for CWS and 5- and 7-min narrative and expository prompts scored for CIWS. This finding about the use of CIWS scoring for 5- to 7-min prompts written by secondary students confirms previous research (e.g., Espin, et al., 2008), although in the current study, it was the word-processed prompts generating this finding and not the handwritten prompts.

The second research question investigated which valid and reliable scoring procedures

best predict performance on a high-stakes writing test. Based on analyses using AIC models, 5- and 7- min handwritten prompts scored for %CWS and 5- and 7-min word-processed prompts scored for CIWS were the best predictors for MBST failure. The criterion validity of %CWS has been noted in other research, but its utility as a progress-monitoring measure is questionable (McMaster & Campbell, 2008; Parker, Tindal & Hasbrouck, 1991). However, if %CWS can be used to screen for students who may have difficulty with the MBST or other similar high-stakes writing assessments, perhaps its utility as a progress-monitoring tool is less important.

The fact that word-processed responses produced scoring procedures that were reliable, valid, and predictive of MBST failure is good news for CBM-W scorers. One challenge in persuading teachers to use CBM-W is the time it takes to decipher the handwritten passage, score for CWS and IWS and then count the sequences. Using the WordMark software to score is much easier; the text can be quickly read, the computer counts the sequences, and it will even graph the CWS or CIWS growth over time.

Further research needs to be conducted across grade levels to determine whether or not word-processed CBM-W prompts are, indeed, technically adequate for students. Variables such as handwriting speed and keyboarding speed should be considered in any analysis, especially since younger students may have less experience with keyboarding. Additionally, understanding how students respond to word-processed progress monitoring prompts is an area for further research. Certainly, the utility of teachers scoring CBM-W prompts with computer technology should continue to be explored, especially related to progress-monitoring and the use of data for instructional decision-making.

Table 1
Descriptive Statistics for Writing Measures

	CWS		Handwritten CIWS		%CWS		CWS		Word-Processed CIWS		%CWS	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Narrative												
<i>3 min</i>												
P5	50.80	(19.77)	39.18	(21.53)	.81	(.13)	49.83	(22.90)	27.45	(29.48)	.72	(.13)
P6	49.90	(34.53)	34.53	(20.30)	.76	(.13)	48.44	(24.88)	27.06	(27.12)	.67	(.18)
P7	49.20	(18.72)	37.31	(21.52)	.80	(.14)	47.72	(29.25)	31.43	(28.63)	.73	(.19)
P8	47.03	(16.89)	34.76	(18.72)	.80	(.15)	50.07	(27.72)	30.77	(26.00)	.72	(.14)
<i>5 min</i>												
P5	79.36	(32.93)	59.76	(34.23)	.80	(.12)	90.59	(42.86)	56.74	(40.52)	.73	(.11)
P6	79.94	(27.09)	55.52	(32.40)	.77	(.14)	80.42	(40.74)	45.00	(42.84)	.68	(.16)
P7	80.88	(29.64)	61.20	(31.37)	.80	(.11)	80.68	(43.59)	48.74	(45.01)	.71	(.16)
P8	79.84	(24.31)	58.97	(28.20)	.80	(.15)	86.14	(41.23)	54.34	(38.68)	.73	(.12)
<i>7 min</i>												
P5	106.42	(48.93)	78.76	(51.09)	.79	(.13)	122.72	(58.02)	74.10	(54.08)	.72	(.11)
P6	110.71	(39.23)	77.76	(45.09)	.77	(.13)	112.85	(55.97)	63.09	(56.80)	.68	(.13)
P7	111.43	(41.22)	83.67	(43.02)	.80	(.11)	113.70	(57.49)	72.62	(54.99)	.73	(.13)
P8	110.21	(35.23)	82.39	(41.92)	.80	(.14)	121.70	(54.91)	76.20	(50.43)	.72	(.12)
Expository												
<i>3 min</i>												
P1	48.21	(17.07)	34.18	(17.92)	.77	(.12)	57.86	(26.96)	33.62	(26.10)	.71	(.12)
P2	46.04	(21.22)	32.04	(24.00)	.76	(.15)	53.17	(28.68)	36.69	(26.12)	.76	(.09)
P3	55.93	(17.71)	46.53	(20.31)	.85	(.12)	55.52	(29.28)	37.54	(27.99)	.75	(.15)
P4	51.60	(21.42)	39.74	(24.54)	.80	(.14)	48.96	(27.49)	31.50	(25.50)	.73	(.14)
<i>5 min</i>												
P1	74.73	(23.54)	51.31	(27.26)	.76	(.13)	96.63	(48.56)	60.31	(49.68)	.73	(.14)
P2	72.13	(33.75)	50.60	(37.09)	.76	(.14)	86.92	(45.47)	58.44	(41.68)	.74	(.10)
P3	89.69	(26.20)	72.44	(32.62)	.83	(.13)	91.45	(46.21)	62.76	(43.68)	.75	(.13)
P4	80.72	(33.02)	60.47	(37.55)	.78	(.14)	80.35	(43.35)	52.35	(38.49)	.74	(.11)
<i>7 min</i>												
P1	100.75	(31.04)	68.04	(35.44)	.76	(.12)	116.43	(57.97)	66.81	(58.04)	.70	(.14)
P2	95.15	(46.27)	66.21	(49.40)	.76	(.14)	117.47	(60.94)	77.44	(53.27)	.73	(.09)
P3	117.03	(39.13)	94.60	(45.72)	.83	(.13)	123.16	(67.04)	84.47	(63.88)	.75	(.13)
P4	106.40	(48.75)	79.77	(51.96)	.78	(.14)	107.07	(53.39)	70.52	(46.50)	.75	(.11)

Note: CWS=correct word sequences; CIWS = correct minus incorrect word sequences; %CWS = percent of total word sequences correctly written; P=prompt number.

Table 2
Alternate-form Reliability of Writing Task by Response Type and Scoring Procedure

	Handwritten			Word-Processed		
	CWS	CIWS	%CWS	CWS	CIWS	%CWS
Narrative						
3 min	.71***	.56***	.69***	.69***	.57***	.58***
5 min	.68***	.67***	.66***	.83***	.75***	.69***
7 min	.60***	.72***	.70***	.86***	.80***	.80***
Expository						
3 min	.55***	.54***	.57***	.77***	.63***	.43***
5 min	.64***	.66***	.68***	.80***	.78***	.69***
7 min	.68***	.68***	.66***	.81***	.78***	.74***

Note: Alternate form reliabilities are reported as pairwise correlation coefficients (r). CWS = correct word sequences; CIWS= correct minus incorrect word sequences; %CWS = percent of total word sequences correctly written. *** $p < 0.0001$. Correlations in red meet the reliability threshold ($r \geq .60$).

Table 3
Descriptive Statistics for Criterion Measures

	Mean	(Standard Deviation)	Range
Teacher Rating (0-5)	3.21	(1.15)	1-5
GPA (0-4.0)	2.57	(.86)	.51-4.0
English Grade (0-4.0)	2.30	(1.1)	0-4.0
MBST Writing Score (0-4)	2.98	(.39)	2.0-4.0

Note. Teacher Rating on a scale of 0-5; GPA=grade point average on a scale of 0-4.0; English Grade on a scale of 0-4.0; MBST=Minnesota Basic Skills Test on a scale of 0-4.

Table 4
Criterion Validity Correlations

	Handwritten				Word-Processed			
	Teacher Rating	English Grade	GPA	MBST	Teacher Rating	English Grade	GPA	MBST
Narrative								
CWS								
3 min	.34†	.12	.17	.19	.50**	.38†	.16	.41***
5 min	.36*	.16	.22	.16	.52**	.42†	.17	.37**
7 min	.35†	.15	.23	.14	.54***	.31	.20	.39**
CIWS								
3 min	.46***	.17	.23	.33*	.62***	.46**	.18	.42***
5 min	.47***	.18	.29*	.28†	.61***	.47**	.20	.42***
7 min	.45***	.19	.27†	.27†	.60***	.48**	.23	.45***
%CWS								
3 min	.49***	.19	.26†	.43***	.46**	.20	.15	.26
5 min	.47***	.18	.24	.42***	.53**	.33	.18	.36**
7 min	.47***	.18	.25	.42***	.54***	.31	.20	.38**
Expository								
CWS								
3 min	.54***	.31	.31*	.29†	.59***	.28	.19	.37**
5 min	.51***	.27	.32*	.27†	.48*	.35	.20	.38**
7 min	.52***	.26	.34**	.30†	.56**	.39†	.25	.40**
CIWS								
3 min	.53***	.27	.35**	.37**	.64***	.33	.21	.38**
5 min	.53***	.24	.34**	.37**	.61***	.41†	.25	.40**
7 min	.54***	.23	.35**	.38***	.63***	.42†	.29†	.41***
%CWS								
3 min	.40**	.09	.29†	.38***	.41†	.26	.18	.36**
5 min	.45**	.12	.27†	.41***	.53**	.29	.27	.34*
7 min	.46***	.12	.26†	.42***	.49*	.30	.30†	.34*

Note. Criterion validity estimates are reported as pairwise correlation coefficients. A Bonferroni correction was used to adjust type-I error for multiple comparisons. Therefore, correlations with $p \leq 0.002$ are considered significant. GPA = grade point average. MBST = Minnesota Basic Skills Test. CIWS = Correct minus incorrect word sequences. PWS = Percent of total word sequences correctly written. CWS = Correct word sequences. Correlations in red meet the reliability and criterion validity threshold ($r \geq .40$).

*** $p < 0.0001$; ** $p < 0.001$; * $p = 0.002$; † $p < 0.01$

Table 5
Logistic Regression Models Predicting Failure on the MBST

	Handwritten ^a		Word-Processed ^b	
	Odds Ratio	AIC	Odds Ratio	AIC
Narrative				
CWS				
3 min	0.98	59.96	1.01	48.27
5 min	0.99	60.98	1.00	47.81
7 min	0.99	61.04	1.00	47.87
CIWS				
3 min	0.96**	51.69	0.99	47.28
5 min	0.98**	51.29	0.99	46.56
7 min	0.98**	53.53	0.99	45.63
%CWS				
3 min	0.86***	44.47	0.97	47.45
5 min	0.85***	43.74	1.00	45.15
7 min	0.86***	45.53	0.93†	44.12
Expository				
CWS				
3 min	0.97*	53.53	0.97*	50.83
5 min	0.99†	54.75	0.99†	51.49
7 min	0.99*	54.15	0.99*	50.90
CIWS				
3 min	0.96**	47.57	0.95*	40.83
5 min	0.97**	47.20	0.98	45.26
7 min	0.98**	47.04	0.98†	44.55
%CWS				
3 min	0.89**	47.10	0.88*	41.07
5 min	0.88**	44.41	0.90*	43.11
7 min	0.88**	44.60	0.91*	43.60

Note. ^a Models control for ethnicity and economic status. ^b Models control for ethnicity, economic status, and typing ability (WPM). AIC = Akaike's Information Criterion. CWS = correct word sequences; CIWS= correct minus incorrect word sequences; %CWS = percent of total word sequences correctly written. AIC = Correlations in red meet the reliability and criterion validity threshold *and* have the lowest AICs.

*** p < .001, ** p < .01, * p < .05, † p < .10

References

- Deno, S. L., Mirkin, P., & Marston, D. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children Special Education and Pediatrics: A New Relationship*, 48, 368-371.
- Deno, S. L., Mirkin, P., & Marston, D. (1980). Relationships among simple measures of written expression and performance on standardized achievement tests (Vol. IRLD-RR-22, pp. 109). University of Minnesota, Minneapolis. Institute for Research on Learning Disabilities.
- Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 15, 5-27.
- 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.
- Espin, C. A., Wallace, T., Campbell, H., Lembke E. S., Long, J. D., Ticha, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children*, 74, 174-193.
- Hooper, S., Miller, C. & Rose, S. (2005). *WordMark*. Minneapolis, MN: Braingiants (www.braingiants.com)
- 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.
- Marston, D., & Deno, S. (1981). The reliability of simple, direct measures of written expression

- (Vol. IRLD-RR-50, pp. 25). U.S.; Minnesota.
- McMaster, K. L., & Campbell, H. (2008). New and existing curriculum-based writing measures: Technical features within and across grades. *School Psychology Review, 37*, 550-566.
- McMaster, K., & Espin, C.A. (2007). Literature synthesis on curriculum-based measurement in writing. *Journal of Special Education, 41*, 68-84.
- Minnesota Basic Skills Technical Manual*. (2001-2002). Minnesota Department of Children, Families, and Learning.
- Minnesota Department of Education & Beck Evaluation and Testing Associates. (1997). *Minnesota Basic Skills of Written Composition/Minnesota Comprehensive Assessments (BST/MCA)*. St. Paul, MN: Minnesota Department of Education.
- Parker, R. I., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality, 2*, 1-17.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved from <http://www.R-project.org>.
- Tindal, G., Marston, D., & Deno, S. L. (1983). The reliability of direct and repeated measurement. (Vol. IRLD-RR-109, pp. 17). University of Minnesota, Minneapolis. Institute for Research on Learning Disabilities.
- Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education, 23*(2), 169-183.
- Videen, J., Marston, D., & Deno, S. L. (1982). Correct word sequences: A valid indicator of proficiency in written expression (Vol. IRLD-RR-84, pp. 61). University of Minnesota, Minneapolis. Institute for Research on Learning Disabilities.

Appendix A

*Narrative and Expository Prompt Administration Instructions – Handwritten***MATERIALS**

Written expression packet for each student.
Stopwatch
Pencils

DIRECTIONS

Say to the students: “Please write your name, the hour/period you have this class, and today’s date on the top of your packet. You will write on two topics today.”

“Before you write, I want you to think about what you will write. First you will think, then you will write. You will have 30 seconds to think and 7 minutes to write. At 3 minutes and 5 minutes, I will ask you to make a slash mark wherever you are in your writing.”

Demonstrate drawing a slash mark between the two words “He/left” on the board or overhead.

“After making the mark, continue writing.”

“Do your best work. If you do not know how to spell a word, you should guess.”

“Keep your pencils down and listen. Turn to the next page. Your topic is written at the top of the page. You should begin your writing with what is written there.”

“You have 30 seconds to think of what to write.”

After 30 seconds: “You have 7 minutes to write. Keep writing until I tell you to stop. Are there any questions? You may begin.”

Start the stopwatch immediately. During the administration, if you see a student put their pencil down and stop writing, give the classwide prompt: “Please continue writing until I say stop.”

After 3 minutes, give the students the following prompt:

“Make a slash mark wherever you are on your paper. Continue writing.”

After 5 minutes, give the students the following prompt:

“Make a slash mark wherever you are on your paper. Continue writing.”

After 7 minutes, say: “Stop, Thank you; put your pencils down.”

Appendix B

Narrative and Expository Prompt Administration Instructions – Word-Processed

MATERIALS

Prompts for each student

Stopwatch

DIRECTIONS

Say to the students: “Double click the alias on your desktop. Please type your name, your ID number, and today’s date at the top of your computer screen. Go to “FILE” “SAVE AS” and save your document in UM...YOUR CLASS AND HOUR...Last Name_First Name_1.”

“Before you write, you will do a short typing task. You will have 2 minutes to type “The quick brown fox jumped over the lazy dog” (pass out cards). Are there any questions? Ready? Type.”

After two minutes: “Stop typing. Hit RETURN until you can no longer see your typing. Please do not go back to your previous typing. Go to FILE and SAVE.”

“You will write on two topics today. Before you write, I want you to think about what you will write. First you will think, then you will write. You will have 30 seconds to think and 7 min to write. At 3 min and 5 min, I will ask you to make an asterisk wherever you are in your writing.”

“Find the asterisk– it is on the 8 key. In order to type an asterisk, you need to hold the shift key and type 8. Please type 5 asterisks at the top of the page.”

“Do your best work. If you do not know how to spell a word, guess.”

“Your topic for writing is at your computer. Before you begin typing, you have 30 seconds to think of what to write.”

After 30 seconds: “You have 7 min to write on the topic in front of you. Before you begin typing, you should type the NUMBER and TOPIC at the top of your page. Keep typing until I tell you to stop. Are there any questions? Begin.”

Start the stopwatch immediately. During the administration, if you see a student stop typing, give the classwide prompt: “Please continue until I tell you to stop.”

After 3 minutes: “Make an asterisk wherever you are on your paper. Continue writing.”

After 5 minutes: “Make an asterisk wherever you are on your paper. Continue writing.”

After 7 minutes: “Stop, Thank you; hit the return key until you can’t see what you just typed. Go to FILE and SAVE.