



TECHNICAL REPORT #7:

Technical Features of New and Existing CBM Writing Measures Within and Across Grades

Kristen L. McMaster and Heather Campbell

RIPM Year 2: 2004 – 2005

Date of Study: November 2004 – May 2005

April 2007

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 examine technical features of new and existing curriculum-based measures of written expression (CBM-W) in terms of writing task, duration, and scoring procedures. Twenty-five 3rd-, 43 5th-, and 55 7th-graders completed passage copying tasks in 1.5 min and picture, narrative, and expository writing prompts in 3 to 7 min. 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 measures detected growth from fall to spring. Different types of measures yielded varying levels of technical adequacy at each grade, with longer durations and more complex scoring procedures generally having stronger technical adequacy for older students. Narrative writing appeared most promising in terms of its technical adequacy across grades. Implications for future research and practice are discussed.

Word Count: 142

Technical Features of New and Existing CBM Writing Measures Within and Across Grades

Progress monitoring has long been a hallmark of special education. Individualized Education Plan (IEP) teams use progress-monitoring data to establish students' present levels of performance, set goals, monitor progress toward those goals, and make instructional changes when progress is insufficient (Deno & Fuchs, 1987). Progress monitoring is also viewed as a way to uphold a major tenet of the Individuals with Disabilities Education Act (IDEA; 2004) by aligning individual goals and objectives with performance and progress in the general education curriculum (e.g., Nolet & McLaughlin, 2000). Relatively recently, progress monitoring has gained increased attention of educational policymakers and administrators. Current educational policies that emphasize standards and accountability (No Child Left Behind Act, 2002) have illuminated the need for assessment tools that can be used to track student progress and to quickly and accurately identify those at risk of failing to meet important yearly benchmarks.

Educators have also focused increasing attention on students' writing. This increased attention is, in part, in response to reports of high proportions of students who do not meet proficiency levels in writing. For example, in 2002, 72% of 4th-graders, 69% of 8th-graders, and 77% of 12-graders were performing below a proficient level in writing (National Center for Education Statistics, 2003). Thus, the National Commission on Writing (2003) urged educational policymakers and practitioners to focus on writing in its report: "The Neglected 'R': The need for a writing revolution." In a subsequent report, the National Commission on Writing (2006) promoted "an integrated system of standards, curriculum, instruction, and assessment" (p. 19) for ensuring that students achieve excellence in writing.

To document student progress within the curriculum and toward rigorous standards, identify those who are struggling, and inform instruction aimed at improving students' writing

proficiency, technically sound progress monitoring tools are needed. One well-researched progress monitoring approach is Curriculum-Based Measurement (CBM; Deno, 1985). A 30-year program of research has illustrated the capacity of CBM to provide reliable and valid indicators of student performance and progress in core content areas (Marston, 1989) and to effect substantial improvements in student achievement (Stecker, Fuchs, & Fuchs, 2005). Below, we briefly review research on the development of CBM in written expression (CBM-W).

Research at Elementary and Secondary Levels

CBM-W was first developed at the Institute for Research on Learning Disabilities (IRLD) at the University of Minnesota. IRLD researchers demonstrated that several simple, countable indices obtained from brief writing samples were valid ($r_s = .67$ to $.88$; Deno, Mirkin, & Marston, 1980; Deno, Mirkin, & Marston, 1982) and reliable ($r_s = .50$ to $.96$; Marston & Deno, 1981; Tindal, Marston, & Deno, 1983) indicators of writing proficiency for 3rd- to 5th-graders. Indices included the number of words written (WW) and words spelled correctly (WSC) when students responded to writing prompts for 3 to 5 min. Correct word sequences (CWS), which incorporates both correct spelling and grammar, also provided a valid index of writing (Videen, Deno, & Marston, 1982). The measures reliably differentiated among students at different skill levels and were sensitive to growth from fall to spring (e.g., Marston, Deno, & Tindal, 1983).

More recently, studies have produced relatively modest technical data at the elementary level (see McMaster & Espin, in press for a review). Tindal and Parker (1991) found that criterion validity of WW, WSC, and CWS was weak to moderate for 2nd-5th graders ($r_s = -.02$ to $.63$). Gansle, Noell, VanDerHeyden, Naquin, and Slider (2002) and Gansle et al. (2004) found other writing dimensions, such as parts of speech, punctuation, and complete sentences, to have weak criterion validity ($r_s = -.05$ to $.42$) and weak to moderate alternate-form reliability (r_s

=.006 to .62) for 3rd- and 4th-graders. These findings raise questions about whether measures studied thus far are appropriate for indexing elementary students' writing proficiency.

Extensions of CBM-W to the secondary level indicate that simple indices such as WW and WSC may also not be sufficient for assessing older students' writing. Tindal and Parker (1989) and Parker, Tindal, and Hasbrouk (1991) found that, for middle school students, percentage measures (e.g., %WSC, %CWS) were better predictors of writing proficiency, and more reliably distinguished students with and without LD, than fluency measures (e.g., WSC, CWS). However, as Parker et al. noted, accuracy measures may be useful for screening, but might pose problems for progress monitoring. For example, if a student produced 10 WSC out of 20 WW in fall, and 50 WSC out of 100 WW in spring, %WSC would not reflect this growth.

Other researchers have shown that combinations of measures (Espin, Scierka, Skare, & Halverson, 1999) or correct minus incorrect word sequences (CIWS; Espin et al., 2000) are more appropriate indices of older students' writing than WW or WSC. Espin et al. (2000) demonstrated that reliability and validity of CIWS did not appear to depend on type of prompt (narrative vs. expository) or duration (3 min vs. 5 min) for middle schoolers. However, Espin, De La Paz, Scierka, and Roelofs (2005) found that CWS and CIWS from 35-min samples yielded stronger validity ($r_s = .58$ to $.90$) than 50-word samples ($r_s = .33$ to $.59$) for secondary students.

Results of elementary- and secondary-level research suggest that different scoring indices might be needed at different grades. Jewell and Malecki (2005) examined this issue directly by obtaining 3-min narrative samples from elementary and middle school students, and scoring them using percentage and fluency indices. Criterion validity was weak to moderate for elementary and middle school students ($r_s = .34$ to $.67$). Weissenburger and Espin (2005) administered 3-, 5-, and 10-min narrative prompts to 4th-, 8th-, and 10th-graders. Alternate-form

reliability was moderate to strong at each grade level for WW, CWS, and CIWS ($r_s = .55$ to $.84$). Correlations were stronger for longer writing samples and weaker at higher grades. Criterion validity between CWS and CIWS and a state writing test was moderate at grades 4 and 8 for CWS and CIWS ($r_s = .47$ to $.68$), and weak at grade 10 for CWS and CIWS ($r_s = .18$ to $.36$).

Extending CBM Research in Written Expression

Many questions remain if CBM-W is to be used within a system of accountability, whereby students at risk of failing to meet high standards are identified, intervention effectiveness is evaluated, and student progress within and across grades is monitored. Procedures developed thus far have yielded moderate technical adequacy at best. Continued research is needed to develop ways to accurately index students' writing proficiency and to determine which measures are most appropriate at which grades.

In the present study, our aim was to replicate and extend research designed to examine CBM-W both within and across elementary and secondary levels. Previous researchers who have examined CBM-W across grades (Jewell & Malecki, 2005; Weissenburger & Espin, 2005) administered only one type of writing task (narrative). These researchers have suggested that different scoring procedures might be needed at different grades, but it is not clear whether the type of task or sample duration should also change with grade level. We intended to add to the literature by examining three primary features of written expression measures administered to students at different grades: type of writing task, duration of writing sample, and scoring procedures. Below we discuss each of these features in detail.

Type of writing task. Researchers have suggested that reliability and validity do not vary substantially depending on type of prompt (i.e., narrative vs. expository) for elementary students (Deno et al., 1980) and middle school students (Espin et al., 2000). However, no direct

comparison has been made of these measures across grades. Given that (a) students are typically assigned to write in narrative formats at the elementary level and in expository formats at the secondary level (e.g., Deschler, Ellis, & Lenz, 1996), and (b) many low-performing writers struggle to adjust their writing to these two different approaches (e.g., Englert, Raphael, Fear, & Anderson, 1989; Miller & Lignugaris-Kraft, 2002), we considered narrative vs. expository writing to be worth further exploration.

In addition, given that writing prompts have yielded modest reliability and validity coefficients compared to other types of CBM measures (e.g., reading), we wondered whether there were *other* ways to obtain simple, efficient measures of written expression that are viable for progress monitoring. To begin a search for new measures, we turned to copying tasks. Our rationale for copying tasks is based on research demonstrating that (a) transcription skills such as spelling and handwriting speed and legibility are strongly related to writing composition (e.g., Graham, 1990; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Jones & Christenson, 1999), and (b) copying tasks can successfully discriminate across a wide age range (1st- through 9th-grades) at least in terms of handwriting speed and legibility (e.g., Graham, Berninger, Weintraub, & Schafer, 1998). We reasoned that passage copying might be robust enough to serve as an indicator of overall writing proficiency, yet retain administration and scoring ease that are characteristic of CBM. Passage copying could also be a logical measure to extend to other populations, such as younger writers (see Lembke, Deno, & Hall, 2003), English language learners (ELLs), and students with significant cognitive disabilities.

Sample duration. A second feature of writing measures that we examined was sample duration. Much of the research thus far has examined 3- to 5-min writing samples; however, Espin et al. (2005) found that, at least for older students, longer samples yielded stronger

reliability and validity coefficients. A goal of our research was to determine for which grades it is necessary to extend writing duration to obtain a technically sound index of writing. In this study, we examined 3-, 5-, and 7-min samples.

Scoring procedures. As illustrated above, WW, WSC, and CWS have shown some (albeit limited) promise for elementary students. Percentage measures (e.g., %WSC and %CWS) have, in some cases, yielded stronger reliability and validity coefficients for both elementary and secondary students (Jewell & Malecki, 2005; Parker et al., 1991; Tindal & Parker, 1989). Moreover, more complex scoring procedures such as CIWS appear to have stronger reliability and validity for secondary students (Espin et al., 1999; Espin et al., 2000). In our study, we included the most commonly examined scoring procedures, again with the purpose of determining which scoring procedures are appropriate at which grades.

Research questions. In the present study, specific research questions included: Which measures of writing performance (in terms of *task*, *duration*, and *scoring procedure*) (1) yield sufficient alternate-form reliability for students at grades 3, 5, and 7? (2) yield sufficient criterion validity for students at grades 3, 5, and 7? and (3) detect writing growth from fall to spring?

Method

Setting and Participants

This study took place in a school serving kindergartners through 8th-graders in a large urban Midwestern district. The school served approximately 750 students; 66.8% percent were from culturally or linguistically diverse backgrounds, 68.0% received free or reduced lunch, 17.3% received special education services, and 24.1% received ELL services. Participants were students from three 3rd-grade classrooms, three 5th-grade classrooms, and four 7th-grade classrooms. Informed written parental consent and complete fall data were obtained from 25 3rd-

43 5th-, and 55 7th-graders. Spring data were collected from 21 3rd-, 32 5th-, and 41 7th-graders (the remaining students moved or had excessive absences during spring testing). Demographic information for participants at each grade level is provided in Table 1.

CBM Writing Measures

Tasks. Students completed two “easy” and two “hard” copying tasks and responded to two picture, two narrative, and two expository prompts. Each prompt was printed at the top of a sheet of paper, followed by lines printed on the same sheet. Additional sheets of lined paper were available if needed. Students wrote in pencil and were encouraged to cross out rather than erase mistakes. We did not specify whether the student should print or write in cursive.

We used both easy and hard copying tasks because we anticipated that students’ accuracy and fluency might vary based on the difficulty of the passages they copied. One easy and one hard copying task were administered at the beginning of the first two testing sessions. Difficulty level was determined based on the readability of the passage. “Easy” passages were simple in sentence structure and had lower readability levels. They were structured after the format of the passage copying task in the Process Assessment of the Learner (Berninger, 2001), which consists of a simple set of directions to a gas station, and were approximately at a 4.3 grade level according to the Flesch-Kincaid readability formula. “Hard” passages were drawn from the district’s 3rd-grade reading curriculum (Cooper & Pikulski, 1999). The passages were written at a 5.8 to 6.0 grade level according to the Flesch-Kincaid formula. Students were instructed to copy a practice sentence (e.g., “The quick brown fox jumped over the lazy dog.”), and then to copy the passage exactly as it appeared, including capitalization and punctuation, for 1.5 min.

Each student also responded to two sets of picture, narrative, and expository prompts. Prompts were intended to reflect experiences to which most U.S. public school students would

be able to relate, and to be simple in terms of vocabulary and sentence structure. Prompts within a set were designed to tap similar background knowledge, so that background knowledge would be unlikely to interact with students' responses to different prompts. At the same time, the prompts were intended to be sufficiently different so that students would not write the same thing in response to each one. Set 1 prompts were designed to tap background knowledge of school-related travel. The picture prompt consisted of students boarding a bus outside of a school. The narrative prompt was, "On my way home from school, a very exciting thing happened...." The expository prompt was, "Write about a trip you would like to take with the students in your class." Set 2 prompts were intended to tap background knowledge related to games or free time. The picture prompt was of students playing ball outside a school. The narrative prompt was, "One day, we were playing outside the school and...." The expository prompt was "Write about a game that you would like to play."

Sample duration. For each copying task, all students wrote for 1.5 min. On the remaining prompts, 3rd- and 5th-graders were instructed to make a slash at the 3 min mark and continue writing for a total of 5 min. Seventh graders made a slash at the 3- and 5-min marks, and continued to write for a total of 7 min.

Scoring procedures. Writing samples were scored using the following procedures:

1. Words written (WW): The total number of words written in the passage.
2. Legible words (LW; Tindal & Parker, 1989): Words in which letters are identifiable, and letter groupings approximate known words.
3. Words spelled correctly (WSC): Words spelled correctly in the context of the sentence.
4. Correct word sequences (CWS; Videen et al., 1982): Any two adjacent, correctly spelled words that are syntactically and semantically correct within the context of the sample.

5. Correct minus incorrect word sequences (CIWS; Espin et al., 1999).
6. Percent WSC (Tindal & Parker, 1989): WSC divided by WW.
7. Percent CWS (Tindal & Parker): CWS divided by total word sequences.

Criterion Variables

Test of Written Language. The Test of Written Language – Third Edition (TOWL-3; Hammill & Larsen, 1996) Spontaneous Writing subtest (Form A) was group-administered to all participants. Students were presented with a picture depicting a futuristic scene of astronauts, space ships, and construction activity; told to think of a story about the picture; and then asked to write for 15 min. Writing samples were scored using analytic rubrics for Contextual Conventions (capitalization, punctuation, and spelling), Contextual Language (quality of vocabulary, sentence construction, and grammar), and Story Construction (quality of plot, prose, character development, interest, and other compositional elements). Alternate-form reliability for 8- to 13-year-olds is reported as .80 to .85. The average validity coefficient with the Writing Scale of the *Comprehensive Scales of Student Abilities* (Hammill & Hresko, 1994) is reported as .50.

Minnesota Comprehensive Assessment. The writing subtest of the Minnesota Comprehensive Assessment (MCA; Minnesota Department of Children, Families, and Learning and NCS Pearson, 2002) is a state standards test that only 5th-graders were required to take at the time of this study. The test requires students to respond to one of four prompts designed to elicit either problem/solution, narrative, descriptive, or clarification writing formats. Samples are scored holistically based on composition, style, sentence formation, grammar, and spelling/mechanics. The technical manual did not include reliability and validity data.

Language arts GPA. In Grade 7 only, students' end-of-year language arts grade point averages (GPAs) were available from district records. GPA is reported based on a 4-point scale.

Procedures

CBM administration. The CBM-W measures were administered in November and May, 2004. The second author group-administered measures in three sessions, each one week apart. In Sessions 1 and 2, students completed two copying tasks followed by two prompts. In Session 3, students responded to the remaining two prompts. We counterbalanced easy and hard passages such that students in different classes copied Passages 1 and 2 of each type in different sessions. We counterbalanced writing prompts such that students in different classes responded to two picture, two narrative, or two expository prompts, with one from Set 1 and one from Set 2 in counterbalanced order in each session.

Scoring training and agreement. A special education doctoral student experienced in the development, administration, and scoring of CBM-W was designated as the “expert” scorer. She met with four other scorers for a two-hour session to describe, demonstrate, and practice the scoring procedures. Each scorer then scored a writing packet that included all prompts and copying tasks produced by one student. The expert compared each scorer’s results with her own, and calculated the percent of agreement for each scoring procedure. Any scorer who achieved less than 80% agreement with the expert on any scoring procedure received further training.

For each scorer, the expert randomly selected one of every 10 packets, scored them independently, and compared the scorer’s results with her own. If agreement for each score was not at least 80%, the expert 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% agreement with the expert again. In only one case did a scorer have to rescore an entire packet.

To score the TOWL-3 Spontaneous Writing samples, the first author and a doctoral student in special education met for one hour to review and practice scoring procedures. We then scored 10% of the writing samples. The number of agreements was divided by the number of agreements plus disagreements and multiplied by 100 to obtain percent agreement. All discrepancies were discussed until we agreed on the appropriate score. We scored common samples until at least 85% agreement was obtained on 10% of the samples. We then divided and independently scored the remaining samples.

Date analysis. Fall writing data were analyzed to determine a subset of measures (those with sufficient alternate-form reliability) to be administered again in spring. Because of the large amount of data collected, we conducted the first set of analyses on all measures, and then narrowed the focus of each subsequent set of analyses to measures deemed “sufficient” in the previous analyses. Specifically, we began by examining distributions of each writing task, scoring procedure, and duration. Measures with relatively normal distributions were examined to determine which had sufficient alternate-form reliability (defined in the Results section) by calculating Pearson r correlation coefficients between forms. Those with sufficient reliability were then examined to determine which had sufficient criterion validity (also defined in the Results) by calculating Pearson’s r or Spearman’s Rho coefficients with criterion measures. Measures with sufficient reliability *and* criterion validity were examined to identify whether they reliably detected fall to spring growth. Repeated measures MANOVAs were conducted within each grade. Reliable main effects of time were followed up with repeated measures ANOVAs.

Results

Before addressing the three research questions, we created histograms for each measure administered in fall at each grade. For the copying tasks, all fluency measures (WW, LW, WSC,

CWS, and CIWS) were sufficiently normally distributed. Percentage measures (%LW, %WSC, and %CWS) were negatively skewed and thus dropped from further analyses. On the picture, narrative, and expository prompts, all measures except %LW were sufficiently normally distributed; %LW was negatively skewed and dropped from further analyses. Table 2 provides fall means and *SDs* on the two alternate forms of each measure at each grade.

Alternate-Form Reliability

Pearson *r* correlation coefficients were calculated for each type of writing task, duration, and scoring procedure. A Bonferroni adjustment was made to reduce the risk of Type I error. At each grade, 50 to 77 correlation coefficients were computed; thus, a significant *p*-value of .001 was set (.05 was divided by the number of coefficients computed).

Because there is not a consensus on criteria by which to judge reliability of measures, we report the strength of reliability coefficients in relative terms. For example, we can compare coefficients to those found for other types of CBM and to other types of writing measures. In reading, reliability coefficients have generally been reported as $r > .85$ (Wayman et al., in press). For standardized writing measures, reliability estimates have ranged from .70 to above .90 (Taylor, 2003). With this information in mind, we consider reliability coefficients of $r > .80$ to be strong, $r = .70$ to $.80$ moderately strong, $r = .60$ to $.70$ moderate, and $r < .60$ weak. Based on these criteria, we determined which measures had sufficient (moderately strong) alternate-form reliability within each grade. Table 3 presents fall and spring alternate-form reliability coefficients for each grade; those with “sufficient” reliability are bolded.

Fall. As shown in Table 3, for 3rd-graders most scoring procedures and writing tasks had strong reliability ($r_s > .80$). Coefficients for the expository prompts were moderate to moderately strong ($r_s = .60$ to $.78$) except for LW and WSC written in 5 min ($r_s = .55$ and $.52$, respectively).

At grade 5, all writing tasks and most scoring procedures had moderately strong to strong reliability ($r_s > .70$). At Grade 7, measures with moderately strong to strong reliability ($r_s > .70$) included hard copying (LW, CWS); picture (3 min: %CWS; 5 min: CIWS, %CWS; 7 min: all scoring procedures); narrative (3 min: LW, WW, %CWS; 5 and 7 min: all scoring procedures); and expository (5 and 7 min; all except %WSC which was weak at 5 min, $r = .45$).

Spring. Based on the above findings, we selected a subset of measures to administer in spring—specifically, those that consistently appeared to have moderate to strong alternate-form reliability ($r_s > .60$) across task, duration, and scoring procedure—within at least two grades. Because of limited time and resources, not all measures could be administered at all grades in spring. Thus, care was taken to ensure that each type of task was administered in at least two grades. Spring measures included hard copying for all grades (easy copying was not included because grade 7 reliability coefficients were weak; thus, hard copying was selected to represent the copying task); picture prompts for grades 3 and 7; narrative prompts for grades 3, 5, and 7; and expository prompts for grades 5 and 7. The sample size at each grade decreased from fall to spring due to students moving or repeated absences during spring testing and make-ups. As shown in Table 3, spring correlations are generally similar to those observed in fall.

Criterion Validity

We examined criterion validity of measures using those that had sufficient alternate-form reliability ($r_s > .60$). Spring CBM-W scores were correlated with spring TOWL-3 raw scores, the MCA Writing Test (grade 5 only), and end-of-year language arts GPA (grade 7 only). Means and SDs for these measures are provided in Table 4. Table 5 presents validity coefficients (measures for which there were no statistically significant correlations are not included). For all analyses, we set a p -value of .01 due to the large number of correlations calculated. Again, correlations of

$r > .80$ were considered strong, $r = .70$ to $.80$ moderately strong, and $r = .60$ to $.70$ moderate. Given that writing measures have historically yielded modest criterion validity coefficients (Taylor, 2003), and because we wished to be as inclusive as possible in identifying promising measures, we also considered correlations above $.50$ to be sufficient for inclusion in further analyses. “Sufficient” correlations are bolded in Table 5.

Grade 3. Measures that yielded moderate to moderately strong validity coefficients with the TOWL-3 included hard copying (WSC, CWS, CIWS; $r_s = .55$ to $.66$); picture (3 min: CWS, CIWS; $r_s = .62$ to $.63$; 5 min: WSC, CWS, %CWS; $r_s = .60$ to $.70$); and narrative (3 min: CWS, CIWS, %CWS; 5 min: CWS, CIWS, %CWS; $r_s = .56$ to $.70$).

Grade 5. Moderate validity coefficients with the TOWL-3 were obtained for CIWS and %CWS on the 3- and 5-min narrative prompts and the 5-min expository prompts ($r_s = .57$ to $.65$). Coefficients with the MCA Writing Test were moderate for CWS, CIWS, and %CWS on the 3- and 5-min narrative prompt ($r_s = .54$ to $.68$); for CIWS on the 3-min expository prompt ($r = .54$); and for CWS, CIWS, and %CWS on the 5-min expository prompt ($r_s = .51$ to $.60$).

Grade 7. None of the validity correlation coefficients for the TOWL-3 were above $r = .60$, and few were above $r = .50$. Spearman’s Rho was calculated between CBM–W scores and language arts grades, which ranged from 1 to 4. Coefficients were weak to moderate for the 3- and 5-min picture prompt ($r_s = .45$ to $.60$). Moderate correlations were found for the other prompts, particularly with respect to more complex fluency-based scores (CWS, CIWS) and longer samples (5 to 7 min), with r_s ranging from $.55$ to $.72$.

Growth from Fall to Spring

To examine which indicators of writing performance detected students’ progress in writing from fall to spring, we identified measures that met the above criteria for “sufficient”

alternate-form reliability and criterion validity *within* grades. Repeated-measures MANOVAs were conducted for each set of measures at each grade, with time as the within-subjects factor (unless only one measure was included in the analysis, in which case a paired-samples *t*-test was conducted). Reliable main effects were followed up with repeated measures ANOVAs. Table 6 displays means, *SDs*, and the average weekly rate of growth for each measure that showed reliable growth. Below we summarize results of MANOVAs and highlight findings of measures showing reliable growth within each grade. Table 7 displays complete ANOVA results.

Grade 3. Hard copying WSC, CWS, and CIWS were identified as having sufficient reliability and criterion validity, and thus were examined for growth. There was no reliable main effect of time, Wilks' Lambda = .84, $F(3, 18) = 1.11, p = .37$. For the 3-min picture prompt, CWS and CIWS were examined. There was a reliable main effect of time, Wilks' Lambda = .65, $F(2, 15) = 4.13, p = .04$. Follow-up tests showed reliable gains for CWS. Given that the measures were administered approximately 20 weeks apart, they showed average gains of .37 CWS per week. For the 5-min picture prompt, WSC, CWS, and %CWS were examined. There was no reliable main effect of time, Wilks' Lambda = .74, $F(3, 15) = 1.73, p = .20$.

For the 3-min narrative prompt, CWS, CIWS, and %CWS were examined. There was no reliable fall to spring gain on these measures, Wilks' Lambda = .94, $F(3, 18) = .40, p = .76$. For the 5-min narrative prompt CWS, CIWS, and %CWS were examined. However, they also did not produce reliable fall to spring differences, Wilks' Lambda = .78, $F(3, 18) = 1.67, p = .21$.

Grade 5. For the 3-min narrative prompt, CWS, CIWS, and %CWS were examined. There was a reliable main effect of time, Wilks' Lambda = .69, $F(3, 27) = 4.09, p = .02$. Students increased reliably on CWS, with an average gain of .54 CWS per week; and on CIWS, with an average gain of .51 CIWS per week. For the 5-min narrative prompt, CWS, CIWS, and %CWS

were examined. There was a reliable main effect of time, Wilks' Lambda = .73, $F(3, 29) = 3.57$, $p = .03$. Follow-up tests indicated fall to spring gains for CWS and CIWS. On average, students increased .72 CWS and .70 CIWS per week.

For the 3-min expository prompt, CIWS were examined. A paired-samples t -test revealed reliable fall-to-spring growth, $t = -2.80$, $p = .009$; students increased, on average, .48 CIWS per week. For the 5-min expository prompt, CWS, CIWS, and %CWS were examined. There was a reliable main effect of time, Wilks' Lambda = .50, $F(3, 29) = 9.63$, $p < .001$. Follow-up tests indicated reliable fall to spring gains for CWS, CIWS, and %CWS. Students gained, on average, .91 CWS, .89 CIWS; and .30 %CWS per week.

Grade 7. For the 3-min picture prompt, %CWS were examined. A paired-samples t -test indicated no fall to spring increase, $t = -1.03$, $p = .31$. For the 5-min picture prompt, %WSC and %CWS were examined. Again, there was no fall to spring increase, Wilks' Lambda = .99, $F(2, 38) = .26$, $p = .77$. For the 7-min picture prompt, CWS, CIWS, and %CWS were examined. There was no reliable main effect of time, Wilks' Lambda = .84, $F(3, 41) = 2.62$, $p = .06$.

For the 3-min narrative prompt, CWS, CIWS, and %CWS were examined. There was no reliable main effect of time, Wilks' Lambda = .86, $F(3, 40) = 1.61$, $p = .20$. For the 5- and 7-min narrative prompts, CWS, CIWS, %WSC, and %CWS were sufficient. For the 5-min prompt, there was no reliable main effect of time, Wilks' Lambda = .83, $F(4, 41) = 2.05$, $p = .11$. For the 7-min narrative prompt, there was a reliable main effect of time, Wilks' Lambda = .78, $F(4, 40) = 2.79$, $p = .04$. Follow-up tests indicated reliable fall to spring gains for CIWS, %WSC, and %CWS. Students gained, on average, .59 CIWS, .16 %WSC, and .20 %CWS per week.

For the 3-min expository prompt, WSC, CWS, CIWS, and %CWS were examined. There was no reliable main effect of time, Wilks' Lambda = .81, $F(4, 35) = 2.02$, $p = .113$. For the 5-

min sample, CWS and CIWS were examined. There was a reliable main effect of time, Wilks' Lambda = .80, $F(2, 37) = 4.81$, $p = .02$. Follow-up tests revealed a reliable fall to spring increase for CIWS. Students gained, on average, .43 CIWS per week. Finally, for the 7-min expository prompt, WSC, CWS, CIWS, and %CWS were examined. There was a reliable main effect of time, Wilks' Lambda = .75, $F(4, 37) = 3.12$, $p = .03$. Follow-up ANOVAs indicated a reliable fall to spring gain for %CWS. Students gained, on average, .29 %CWS per week.

Discussion

In this study, we examined technical features of “new” (easy and hard copying) and “existing” (picture, narrative, and expository) measures of written expression within and across grades 3, 5, and 7. A large number of measures was examined in this study; Table 8 presents a comprehensive (and hopefully digestible) summary of the measures examined and whether they met our criteria for sufficient alternate-form reliability, criterion validity, and capacity to show fall-to-spring growth. Those measures that were sufficient in all three areas are shaded. Below, we highlight measures that appeared most promising and discuss implications for further research and practice using CBM-W to monitor progress within and across grades.

Promising Measures Within and Across Grades 3, 5, and 7

Passage copying. The easy copying task produced sufficient alternate-form reliability for grades 3 and 5. The hard copying task produced sufficiently reliable scores on alternate forms for students in all three grades. Because our aim was to reduce the number of measures to be administered in spring by selecting measures that showed promise across all three grades (for practical as well as conceptual reasons), we only administered hard copying in spring. For grade 3, hard copying produced *valid* scores in relation to the TOWL-3.

These results provide further evidence of the relation between transcription skills and overall writing proficiency, at least for younger students, as described by researchers such as Graham et al. (1997), and Jones and Christenson (1999). This is a promising finding because copying tasks are relatively easy to administer and score compared to other CBM-W measures. Unfortunately, hard copying did not show reliable fall to spring growth within grade 3, raising questions about its utility for progress-monitoring. Also, because we did not administer easy copying in the spring, we cannot draw many conclusions about this measure. It is possible that easy copying would be more sensitive to growth of 3rd-graders than hard copying. There is other preliminary evidence of the utility of copying tasks for beginning writers (Lembke et al., 2003). Such tasks should be examined further for young students.

Picture prompts. Generally, our findings suggest that picture prompts are promising for progress monitoring in grade 3. CWS and CIWS produced in 3 min yielded sufficient alternate-form reliability and criterion validity; as did WSC, CWS, and %CWS produced in 5-min. Moreover, CWS written in 3 min showed reliable fall to spring growth. Further research should examine whether picture prompts can be used to monitor 3rd-graders' progress on an ongoing and frequent basis (e.g., weekly) and whether such progress data can be used to enhance instructional decision-making. A challenge that will likely arise in using pictures to monitor progress on a frequent basis is that it is difficult to find appropriate pictures to use as equivalent probes; this may be why picture prompts have not appeared more frequently in the CBM-W literature.

Picture prompts were also administered to 7th-graders in spring, but only a handful of scoring procedures showed sufficient criterion validity, and none reflected fall to spring growth. To better understand the role of picture prompts in a seamless and flexible progress monitoring system, further research should examine how far beyond grade 3 picture prompts can be

extended, as well as whether they work with younger students and students with disabilities.

Picture prompts may also be useful for students who encounter difficulties with language, such as those who are deaf/hard of hearing or ELLs.

Narrative prompts. CBM-W researchers have frequently examined narrative prompts, and they are commonly administered in practice. However, a recent review of research has indicated weak to moderate reliability and validity for narrative prompts, especially when relatively simple scoring procedures such as WW and WSC are used (McMaster & Espin, in press). Our findings suggest that more complex fluency-based (CWS and CIWS) and percentage-based (%WSC, %CWS) scoring procedures might yield more accurate results, even for elementary students. CWS, CIWS, %WSC, and %CWS yielded sufficient reliability and criterion validity for students at all three grades on 3- and 5-min samples; WW and WSC did not. These findings are consistent with previous research supporting the use of percentage measures and more complex scoring procedures for both elementary and secondary level students (e.g., Jewell & Malecki, 2005; Parker et al., 1991; Tindal & Parker, 1989).

Narrative prompts did not reflect fall to spring growth for 3rd-graders, but reliable growth was detected for CWS and CIWS on 5th-graders' 3- and 5-min samples, and for CIWS, %WSC, and %CWS on 7th-graders' 7-min samples. The finding that percentage indices reflected growth is interesting, because they were expected to be problematic for monitoring progress (e.g., Parker et al., 1991). Whether growth on percentage measures consistently provides meaningful information warrants further investigation.

Expository prompts. Expository prompts did not appear to be sufficiently reliable for 3rd-graders, but several scoring procedures obtained from 3- and 5-min samples had sufficient reliability and validity for 5th- and 7th-graders. CIWS obtained from 3-min samples, and CWS,

CIWS, and %CWS from 5-min samples reflected reliable fall to spring growth for 5th-graders. CIWS and %CWS on 5- and 7-min samples reflected reliable growth for 7th-graders. These findings are consistent with those of previous researchers who have found technical features of narrative and expository prompts to be similar for secondary students (e.g., Espin et al., 2000). Again, progress based on percentage measures should be viewed with caution.

Limitations

The following study limitations should be considered. First, we had a relatively small sample, especially at grade 3. The small sample size could have affected the strength of correlations. Second, our sample included only 3rd-, 5th-, and 7th-graders to represent mid- to late elementary and middle school; we cannot generalize findings beyond these grades. Third, there are limitations to the prompts that we used. For example, whereas we referred to our copying prompts as “easy” and “hard,” these distinctions are only based on the readability of the prompts. Whether they were truly easy or hard for students is likely influenced by individuals’ reading and writing skill, their grade level, and other factors. It is also not clear whether our findings would generalize to prompts that tap different background knowledge (i.e., different pictures or prompts with different content), or other kinds of expository writing (e.g., analytic or persuasive writing).

A fourth limitation has to do with our process of reducing measures from fall to spring. We eliminated measures in the fall for two reasons: (a) they did not demonstrate sufficient reliability, which is necessary for examining validity and (b) we had limited time and resources to administer and score a set of measures in spring, and so selected those that appeared most promising. However, our elimination of some of the measures may have been premature. For example, easy copying showed sufficient reliability for 3rd and 5th-graders in fall, as did the picture prompts for 5th-graders. Had we administered these measures in spring, we would have

additional information regarding their utility for monitoring progress within and across grades. Future research should include further examination of these measures.

Implications for Research and Practice

Using CBM-W Across Grades

What we know. Across grades 3, 5, and 7, our results support the use of 5-min narrative writing prompts for screening. Educators should consider using more complex scoring procedures (CWS and CIWS as opposed to WW or WSC) to ensure reliability and validity. Whereas administration and scoring of writing prompts are more time consuming than copying tasks, it appears that narrative prompts have the capacity to extend across a wider range of grades. Finally, at least across grades 5 to 7, CWS and CIWS obtained from expository writing prompts appear promising for screening and connecting progress across grades.

What we still need to learn. Further research should examine the utility of the above measures to connect progress to grades that we skipped in this study (e.g., 3rd to 4th, 4th to 5th, etc.) as well as to younger and older writers. More research is needed to examine measures that were eliminated from the study after the fall administration (easy copying and picture prompts). These measures could prove useful for younger students and for students with disabilities.

Using CBM-W Within Grades

What we know. Within grade 3, our results support the use of 1.5 min “hard” passage copying tasks for screening. The advantage to using copying tasks is that they are quick and easy to administer and score. Three-min picture prompts scored for CWS, or 5-min picture prompts scored for CWS or CIWS, seem most viable in terms of reliability, criterion validity, and capacity to show growth across the school year. Within grade 5, CWS and CIWS for both 3- and 5-min narrative and expository prompts seem viable in terms of reliability, validity, and capacity

to show growth. Within grade 7, 7-min narratives scored for CIWS, and 5- to 7-min expository prompts scored for CIWS, seem most promising. In other words, within each grade level, measures appear to vary in their utility for progress monitoring in terms of *type* of prompt (picture prompt in mid-elementary, narrative or expository prompts in late elementary to middle school); *duration* (i.e., longer durations at higher grade levels); and *scoring procedures* (i.e., more complex scoring procedures at higher grade levels).

What we still need to learn. Within each grade level, research is needed to further examine the utility of measures identified as promising for progress monitoring. For example, do the measures show growth over relatively brief periods (e.g., weekly or monthly) such that teachers can use them for instructional decision-making? How many data points are needed, and how often must measures be administered, to establish stable growth trajectories? When teachers use CBM-W for instructional decision-making, does students' writing performance improve?

In addition, whereas some of the measures did not show promise in this study, we believe it would be premature to eliminate them from the universe of possible CBM-W tools. For example, the narrative prompts did not show substantial fall-to-spring growth for 3rd-graders in this study, but we do not suggest disregarding this measure as a possible progress-monitoring tool. Given its strong reliability and validity for 3rd-graders, and its widespread use, further examination of this measure is needed. It is possible that the 3rd-graders in our studies simply did not make much growth in writing from fall to spring. Future research could compare growth on CBM-W to growth on other writing measures. It is also possible that *other* narrative prompts, or aggregated scores of several narrative prompts, will yield a different picture of growth over time.

Conclusion

In this study, we identified measures that show promise for indexing students' writing proficiency within and across grades 3, 5, and 7. Whereas our results lead us closer to a set of measures for monitoring students' writing progress within and across grades, we should emphasize (as have others; e.g., Tindal & Hasbrouck, 1991), that writing is a complex and multidimensional process, and that few (if any) measures of written expression are likely to capture all of the critical dimensions of writing. Thus, for instructional decision-making, we encourage educators to consider qualitative as well as quantitative aspects of students' writing. Finally, educators should keep an eye toward the research for further development of progress monitoring approaches in written expression. It is our hope that continued research will lead to improvements in the technical soundness and instructional utility of CBM-W.

References

- Berninger, V. (2001). *Process Assessment of the Learner (PAL) Test Battery for Reading and Writing*. San Antonio, TX: The Psychological Corporation.
- Cooper, D., & Pikulski, J. (1999). *Invitations to Literacy*. Boston, MA: Houghton Mifflin.
- Deno, S. L., & Fuchs, L. S. (1987). Developing Curriculum-Based Measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- 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). Minnesota Univ, Minneapolis Inst for Research on Learning Disabilities.
- Deschler, D. D., Ellis, E., & Lenz, B. K. (1996). *Teaching adolescents with learning disabilities* (2nd ed.). Denver, CO: Love Publishing.
- Englert, C. S., Raphael, T. E., Fear, K. L., & Anderson, L. M. (1989). Students' metacognitive knowledge about how to write informational texts. *Learning Disability Quarterly, 11*, 18-46.
- 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. 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.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to responsiveness-to-intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 92-99.
- Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternative measures for curriculum-based measurement in writing. *School Psychology Review, 31*, 477-497.
- Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Slider, N. J., Hoffpauir, L. D., & Whitmarsh, E. L. et al. (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.
- Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82*, 781-791.
- Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*, 170-182.
- Graham, S., Berninger, V., Weintraub, N., & Schafer, W. (1998). Development of handwriting speed and legibility in grades 1-9. *Journal of Educational Research, 92*, 42-52.

Hammill, D. D. & Hresko, W. P. (1994). *Comprehensive scales of student abilities*. Austin, TX: PRO-ED.

Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language-Third Edition*. Austin, TX: PRO-ED, Inc.

Individuals with Disabilities Education Improvement Act, P. L. 108-446 U.S.C. (2004).

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.

Jones, D., & Christensen, C. A. (1999). Relationship between automaticity in handwriting and students' ability to generate written text. *Journal of Educational Psychology, 91*, 44-49.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention, 28*, 23-35.

Marston, D, Deno, S. L., & Tindal, G. (1983). A Comparison of Standardized Achievement Tests and Direct Measurement Techniques in Measuring Pupil Progress (Vol. IRLD-RR-126, pp. 29). Minnesota Univ, Minneapolis Inst for Research on Learning Disabilities.

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

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., & Espin, C. A. (in press). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education*.

- Miller, T. L., & Lignugaris-Kraft, B. (2002). The effects of text structure discrimination training on the writing performance of students with learning disabilities. *Journal of Behavioral Education, 11*, 203-230.
- Minnesota Department of Children, Families and Learning and NCS Pearson. (2002). *Minnesota comprehensive assessments, Grades 3 & 5, Technical Manual*. Retrieved January 25, 2005 from <http://education.state.mn.us/>.
- National Center for Education Statistics (2003). *Nation's Report Card: Writing*. Retrieved April 23, 2006 from <http://nces.ed.gov/nationsreportcard/writing/>.
- National Commission on Writing (2003). *The neglected "R: The need for a writing revolution*. Retrieved March 16, 2007 from <http://www.writingcommission.org/>.
- National Commission on Writing (2006). *Writing and school reform*." Retrieved March 16, 2007 from <http://www.writingcommission.org/>.
- No Child Left Behind Act, Public Law No. 107-110, 115 Stat. 1425, 2002 U.S.C. (2002).
- Nolet, V., & McLaughlin, M. J. (2000). *Assessing the general curriculum: Including students with disabilities in standards-based reform*. Thousand Oaks, CA: Corwin Press, Inc.
- Parker, R. I., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality, 2*, 1-17.
- Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18*, 147-156.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.

- Taylor, R. L. (2003). *Assessment of exceptional students: Educational and psychological procedures*. (6th Ed.). Boston, MA: Allyn and Bacon.
- Tindal, G., & Hasbrouck, J. (1991). Analyzing student writing to develop instructional strategies. *Learning Disabilities Research and Practice, 6*, 237-245.
- Tindal, G., & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research and Practice, 6*, 211-218.
- 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.
- Tindal, G., Marston, D., & Deno, S. L. (1983). The reliability of direct and repeated measurement. (Vol. IRLD-RR-109, pp. 17). Minnesota Univ, Minneapolis Inst for Research on Learning Disabilities.
- 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). Minnesota Univ, Minneapolis Inst for Research on Learning Disabilities.
- Wayman, M.M., Wallace, T., Wiley, H.I., Ticha, R., & Espin, C.A. (in press). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*
- Weissenburger, J. W., & Espin, C. A. (2005). Curriculum-based measures of writing across grades. *Journal of School Psychology, 43*, 153-169.

Table 1

Student Demographics

| | Grade 3 | | Grade 5 | | Grade 7 | |
|---------------------------------|----------|------|----------|------|----------|------|
| | <i>n</i> | (%) | <i>n</i> | (%) | <i>n</i> | (%) |
| Male | 10 | (40) | 13 | (30) | 29 | (53) |
| Ethnicity | | | | | | |
| American Indian | 3 | (12) | 5 | (12) | 3 | (5) |
| African American | 14 | (56) | 20 | (47) | 19 | (35) |
| Asian | 0 | 0 | 2 | (5) | 5 | (9) |
| Hispanic | 0 | 0 | 4 | (9) | 5 | (9) |
| White | 8 | (32) | 12 | (28) | 23 | (42) |
| Eligible for free/reduced lunch | 15 | (60) | 28 | (65) | 32 | (58) |
| Eligible for ELL services | 7 | (28) | 12 | (28) | 7 | (13) |
| Receiving Special Education | 2 | (8) | 6 | (14) | 5 | (9) |

Note. ELL = English Language Learner.

Table 2

Descriptive Statistics for Fall Writing Measures

| Measure | Grade | Probe | LW | | WW | | WSC | | CWS | | CIWS | | %WSC | | %CWS | |
|-----------------|-------|-------|-------|---------|--------|---------|-------|---------|-------|---------|-------|---------|-------|---------|-------|---------|
| | | | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) |
| Easy Copy | 3 | 1 | 15.33 | (6.25) | 15.79 | (6.54) | 13.83 | (5.60) | 14.25 | (5.47) | 10.46 | (6.30) | na | na | na | na |
| | | 2 | 17.52 | (7.55) | 17.84 | (7.37) | 16.44 | (8.04) | 17.24 | (8.17) | 15.20 | (9.12) | na | na | na | na |
| | 5 | 1 | 23.81 | (4.80) | 24.28 | (4.59) | 22.79 | (4.62) | 23.70 | (5.64) | 19.21 | (12.57) | na | na | na | na |
| | | 2 | 25.63 | (5.45) | 26.05 | (5.28) | 24.81 | (5.60) | 25.65 | (6.64) | 23.02 | (8.26) | na | na | na | na |
| | 7 | 1 | 30.89 | (8.33) | 31.20 | (8.36) | 28.20 | (7.91) | 30.53 | (9.33) | 25.36 | (10.26) | na | na | na | na |
| | | 2 | 34.05 | (7.82) | 34.73 | (7.70) | 32.36 | (7.58) | 34.38 | (8.97) | 29.98 | (11.77) | na | na | na | na |
| Hard Copy | 3 | 1 | 16.80 | (7.31) | 17.04 | (7.27) | 15.96 | (6.62) | 17.12 | (8.25) | 13.60 | (9.67) | na | na | na | na |
| | | 2 | 14.16 | (8.33) | 14.64 | (8.07) | 13.28 | (7.69) | 13.56 | (8.78) | 10.16 | (9.87) | na | na | na | na |
| | 5 | 1 | 26.37 | (5.99) | 27.44 | (6.18) | 25.05 | (6.17) | 28.09 | (7.32) | 22.65 | (14.03) | na | na | na | na |
| | | 2 | 24.53 | (5.21) | 25.00 | (5.24) | 22.95 | (6.36) | 24.33 | (6.10) | 20.12 | (10.01) | na | na | na | na |
| | 7 | 1 | 33.09 | (9.18) | 33.69 | (9.14) | 31.24 | (9.46) | 35.40 | (11.49) | 30.20 | (14.10) | na | na | na | na |
| | | 2 | 31.73 | (8.15) | 32.55 | (7.93) | 30.45 | (8.17) | 32.15 | (9.78) | 27.65 | (12.79) | na | na | na | na |
| Picture (3 min) | 3 | 1 | 24.84 | (14.70) | 28.24 | (14.64) | 20.36 | (13.13) | 14.88 | (13.01) | 0.64 | (18.86) | 72.59 | (19.77) | 49.12 | (24.98) |
| | | 2 | 25.64 | (13.62) | 29.32 | (14.15) | 20.96 | (12.62) | 16.72 | (14.24) | 1.56 | (21.57) | 72.83 | (18.81) | 52.38 | (25.92) |
| | 5 | 1 | 36.98 | (14.75) | 38.47 | (14.72) | 31.56 | (13.20) | 26.98 | (13.13) | 11.86 | (16.47) | 81.83 | (11.85) | 63.99 | (18.60) |
| | | 2 | 36.16 | (15.19) | 38.19 | (15.10) | 29.63 | (13.03) | 23.77 | (12.70) | 5.58 | (18.23) | 76.44 | (16.35) | 56.10 | (21.24) |
| | 7 | 1 | 47.65 | (14.63) | 48.64 | (14.32) | 43.02 | (14.61) | 39.95 | (16.36) | 27.29 | (19.96) | 87.25 | (10.11) | 73.92 | (16.94) |
| | | 2 | 43.53 | (12.40) | 44.65 | (12.51) | 38.53 | (11.80) | 35.18 | (13.88) | 21.87 | (19.21) | 85.90 | (9.77) | 71.54 | (18.65) |
| Picture (5 min) | 3 | 1 | 38.40 | (23.19) | 43.44 | (22.01) | 31.12 | (20.85) | 22.76 | (21.19) | 1.32 | (29.22) | 71.65 | (19.85) | 48.33 | (25.53) |
| | | 2 | 39.48 | (24.54) | 45.32 | (25.04) | 32.08 | (22.71) | 25.36 | (24.98) | 1.48 | (35.03) | 71.40 | (18.56) | 49.98 | (24.52) |
| | 5 | 1 | 59.02 | (21.15) | 61.49 | (20.84) | 50.26 | (19.36) | 42.72 | (19.85) | 18.23 | (25.71) | 81.10 | (11.41) | 62.80 | (17.73) |
| | | 2 | 60.63 | (24.38) | 63.47 | (24.12) | 50.12 | (21.28) | 42.67 | (23.24) | 14.81 | (31.39) | 77.92 | (15.63) | 59.15 | (21.18) |
| | 7 | 1 | 72.63 | (22.48) | 74.09 | (22.11) | 65.33 | (22.45) | 60.81 | (25.15) | 41.13 | (29.96) | 86.79 | (9.38) | 73.08 | (16.56) |
| | | 2 | 71.60 | (19.97) | 73.31 | (20.20) | 63.45 | (20.32) | 58.22 | (23.22) | 36.65 | (30.85) | 85.55 | (10.90) | 71.32 | (18.34) |
| Picture (7 min) | 7 | 1 | 97.44 | (32.02) | 99.58 | (32.08) | 87.51 | (31.56) | 81.76 | (34.02) | 54.91 | (38.94) | 86.69 | (8.61) | 73.20 | (15.22) |
| | | 2 | 99.45 | (30.11) | 102.02 | (30.27) | 88.60 | (31.12) | 82.84 | (35.64) | 53.51 | (45.53) | 85.63 | (10.49) | 71.77 | (17.74) |

| | | | | | | | | | | | | | | | | |
|-----------------------|---|---|--------|---------|--------|---------|--------|---------|-------|---------|-------|---------|-------|---------|-------|---------|
| Narrative (3 min) | 3 | 1 | 29.20 | (15.94) | 32.84 | (15.78) | 23.48 | (14.64) | 19.08 | (14.86) | 1.80 | (20.94) | 69.18 | (20.57) | 50.14 | (26.67) |
| | | 2 | 27.08 | (14.86) | 30.20 | (15.04) | 21.92 | (13.62) | 17.04 | (14.52) | 1.32 | (20.66) | 71.24 | (17.49) | 50.11 | (24.45) |
| | 5 | 1 | 39.95 | (16.52) | 41.23 | (16.49) | 33.93 | (14.53) | 29.33 | (14.80) | 14.12 | (19.32) | 81.64 | (12.59) | 64.84 | (18.17) |
| | | 2 | 40.81 | (15.48) | 42.53 | (15.80) | 34.79 | (13.84) | 29.49 | (14.57) | 13.26 | (18.49) | 81.15 | (11.73) | 63.73 | (18.36) |
| | 7 | 1 | 50.62 | (14.71) | 51.71 | (14.65) | 44.78 | (14.38) | 41.36 | (15.83) | 26.35 | (21.10) | 86.33 | (9.64) | 72.87 | (16.66) |
| | | 2 | 51.93 | (14.88) | 53.45 | (15.16) | 45.42 | (14.39) | 41.04 | (16.13) | 24.18 | (21.80) | 84.94 | (10.11) | 70.50 | (17.00) |
| Narrative (5 min) | 3 | 1 | 45.40 | (25.27) | 51.04 | (24.81) | 36.60 | (23.03) | 29.56 | (23.65) | 2.92 | (32.80) | 70.71 | (18.59) | 51.02 | (24.52) |
| | | 2 | 43.40 | (21.08) | 48.56 | (21.16) | 35.20 | (20.31) | 28.00 | (22.60) | 3.00 | (32.36) | 71.32 | (17.26) | 50.93 | (24.11) |
| | 5 | 1 | 64.47 | (26.36) | 66.81 | (26.23) | 54.65 | (23.43) | 47.02 | (23.53) | 21.70 | (29.95) | 80.88 | (12.56) | 63.74 | (17.51) |
| | | 2 | 66.47 | (25.70) | 69.30 | (26.09) | 56.67 | (22.91) | 49.12 | (24.41) | 22.60 | (29.94) | 80.89 | (11.50) | 63.79 | (17.58) |
| | 7 | 1 | 79.96 | (21.89) | 81.93 | (21.72) | 70.69 | (20.62) | 64.78 | (22.67) | 40.04 | (30.56) | 86.18 | (8.73) | 72.04 | (15.71) |
| | | 2 | 83.27 | (23.20) | 85.65 | (23.24) | 73.33 | (22.21) | 67.11 | (25.14) | 41.33 | (33.82) | 85.43 | (9.41) | 71.65 | (16.22) |
| Narrative (7 min) | 7 | 1 | 111.09 | (29.19) | 113.89 | (28.53) | 98.31 | (27.91) | 90.36 | (31.25) | 56.18 | (42.28) | 86.09 | (8.75) | 72.03 | (15.09) |
| | | 2 | 113.15 | (31.13) | 116.60 | (31.43) | 99.95 | (29.91) | 92.02 | (33.42) | 56.78 | (44.91) | 85.58 | (9.11) | 71.80 | (15.70) |
| Expository (3 min) | 3 | 1 | 30.16 | (10.95) | 33.44 | (10.71) | 24.88 | (10.71) | 18.80 | (11.40) | 1.76 | (17.43) | 73.70 | (15.29) | 51.96 | (22.63) |
| | | 2 | 28.48 | (11.49) | 32.12 | (12.34) | 24.20 | (10.12) | 19.80 | (12.12) | 4.52 | (19.93) | 77.11 | (15.29) | 57.76 | (25.67) |
| | 5 | 1 | 42.12 | (14.69) | 43.88 | (14.46) | 36.30 | (12.91) | 31.42 | (13.39) | 16.23 | (17.03) | 82.80 | (9.84) | 67.29 | (16.42) |
| | | 2 | 43.09 | (12.75) | 44.49 | (12.42) | 37.88 | (12.45) | 34.30 | (13.78) | 20.84 | (19.60) | 84.65 | (12.38) | 71.15 | (18.86) |
| | 7 | 1 | 47.24 | (16.99) | 48.18 | (17.01) | 41.98 | (16.48) | 38.93 | (17.71) | 26.27 | (20.95) | 85.02 | (14.54) | 73.24 | (18.32) |
| | | 2 | 56.71 | (14.61) | 58.29 | (15.54) | 50.84 | (14.47) | 47.05 | (15.15) | 31.24 | (19.68) | 87.16 | (8.65) | 74.80 | (14.37) |
| Expository (5 min) | 3 | 1 | 44.60 | (17.60) | 49.04 | (17.23) | 35.60 | (16.64) | 26.76 | (17.63) | 0.72 | (25.86) | 71.65 | (15.47) | 49.68 | (21.74) |
| | | 2 | 44.08 | (18.28) | 50.16 | (20.08) | 36.68 | (16.88) | 28.56 | (19.15) | 2.56 | (30.30) | 73.81 | (14.38) | 52.35 | (23.36) |
| | 5 | 1 | 64.93 | (23.08) | 67.91 | (22.47) | 55.35 | (20.40) | 47.51 | (21.54) | 22.51 | (28.02) | 81.38 | (10.19) | 64.87 | (16.92) |
| | | 2 | 69.05 | (19.78) | 71.53 | (18.92) | 60.33 | (19.13) | 54.42 | (20.84) | 31.86 | (29.15) | 83.98 | (12.24) | 70.20 | (17.86) |
| | 7 | 1 | 75.27 | (26.95) | 77.04 | (27.24) | 66.55 | (25.62) | 61.73 | (27.29) | 40.44 | (30.62) | 84.85 | (11.73) | 72.56 | (16.29) |
| | | 2 | 87.53 | (21.95) | 90.13 | (22.46) | 78.60 | (21.36) | 72.73 | (23.06) | 48.09 | (29.98) | 86.99 | (7.52) | 74.47 | (13.47) |
| Expository (7 min) | 7 | 1 | 98.67 | (37.53) | 101.45 | (37.97) | 87.09 | (35.45) | 80.51 | (36.91) | 51.44 | (40.77) | 84.88 | (9.37) | 72.27 | (15.06) |
| | | 2 | 112.71 | (31.85) | 116.20 | (31.95) | 101.13 | (30.53) | 93.47 | (32.60) | 61.05 | (40.38) | 86.65 | (7.06) | 73.83 | (13.14) |

Note. LW = legible words; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.

Table 3

Alternate-Form Reliability of Writing Tasks and Scoring Procedures by Grade Level (Fall and Spring)

| | LW | | WW | | WSC | | CWS | | CIWS | | %WSC | | %CWS | |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Fall | Spring | Fall | Spring | Fall | Spring | Fall | Spring | Fall | Spring | Fall | Spring | Fall | Spring |
| Easy Copy | | | | | | | | | | | | | | |
| Grade 3 | 0.84 | na | 0.85 | na | 0.84 | na | 0.88 | na | 0.72 | na | na | na | na | na |
| Grade 5 | 0.76 | na | 0.73 | na | 0.74 | na | 0.80 | na | 0.56 | na | na | na | na | na |
| Grade 7 | 0.54 | na | 0.50 | na | 0.49 | na | 0.60 | na | 0.59 | na | na | na | na | na |
| Total | 0.80 | na | 0.78 | na | 0.77 | na | 0.80 | na | 0.67 | na | na | na | na | na |
| Hard Copy | | | | | | | | | | | | | | |
| Grade 3 | 0.94 | 0.82 | 0.95 | 0.85 | 0.91 | 0.79 | 0.86 | 0.84 | 0.60 | 0.67 | na | na | na | na |
| Grade 5 | 0.77 | 0.73 | 0.65 | 0.68 | 0.45 | 0.69 | 0.69 | 0.63 | 0.71 | 0.61 | na | na | na | na |
| Grade 7 | 0.70 | 0.58 | 0.65 | 0.58 | 0.63 | 0.61 | 0.71 | 0.65 | 0.64 | 0.58 | na | na | na | na |
| Total | 0.86 | 0.74 | 0.83 | 0.74 | 0.77 | 0.76 | 0.83 | 0.78 | 0.73 | 0.71 | na | na | na | na |
| Picture (3 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.80 | 0.91 | 0.82 | 0.91 | 0.83 | 0.88 | 0.87 | 0.93 | 0.88 | 0.85 | 0.86 | 0.80 | 0.85 | 0.73 |
| Grade 5 | 0.84 | na | 0.86 | na | 0.87 | na | 0.86 | na | 0.74 | na | 0.68 | na | 0.74 | na |
| Grade 7 | 0.55 | 0.52 | 0.51 | 0.53 | 0.56 | 0.55 | 0.62 | 0.54 | 0.68 | 0.55 | 0.62 | 0.60 | 0.70 | 0.59 |
| Total | 0.78 | 0.69 | 0.76 | 0.69 | 0.79 | 0.70 | 0.80 | 0.72 | 0.79 | 0.72 | 0.75 | 0.79 | 0.78 | 0.74 |
| Picture (5 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.84 | 0.91 | 0.85 | 0.90 | 0.88 | 0.89 | 0.90 | 0.91 | 0.91 | ns | 0.83 | 0.72 | 0.86 | 0.64 |
| Grade 5 | 0.87 | na | 0.87 | na | 0.90 | na | 0.88 | na | 0.83 | na | 0.78 | na | 0.79 | na |
| Grade 7 | 0.63 | 0.66 | 0.60 | 0.67 | 0.63 | 0.65 | 0.68 | 0.65 | 0.74 | ns | 0.67 | 0.68 | 0.78 | 0.68 |
| Total | 0.82 | 0.80 | 0.80 | 0.80 | 0.83 | 0.81 | 0.84 | 0.82 | 0.84 | 0.43 | 0.80 | 0.80 | 0.83 | 0.77 |
| Picture (7 min) | | | | | | | | | | | | | | |
| Grade 7 | 0.75 | 0.66 | 0.73 | 0.65 | 0.75 | 0.67 | 0.80 | 0.75 | 0.81 | 0.77 | 0.76 | 0.59 | 0.81 | 0.74 |
| Narrative (3 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.73 | 0.63 | 0.73 | 0.70 | 0.78 | 0.70 | 0.86 | 0.76 | 0.86 | 0.72 | 0.84 | ns | 0.85 | ns |
| Grade 5 | 0.60 | 0.76 | 0.60 | 0.76 | 0.54 | 0.78 | 0.58 | 0.80 | 0.69 | 0.84 | 0.70 | 0.80 | 0.74 | 0.80 |
| Grade 7 | 0.71 | ns | 0.70 | ns | 0.69 | ns | 0.69 | 0.57 | 0.67 | 0.69 | 0.64 | 0.67 | 0.72 | 0.76 |
| Total | 0.75 | 0.69 | 0.74 | 0.68 | 0.75 | 0.72 | 0.77 | 0.77 | 0.76 | 0.82 | 0.79 | 0.55 | 0.80 | 0.71 |
| Narrative (5 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.80 | 0.74 | 0.79 | 0.74 | 0.86 | 0.73 | 0.90 | 0.78 | 0.89 | 0.80 | 0.86 | ns | 0.86 | ns |
| Grade 5 | 0.73 | 0.80 | 0.74 | 0.76 | 0.71 | 0.84 | 0.71 | 0.86 | 0.77 | 0.88 | 0.81 | 0.83 | 0.80 | 0.91 |
| Grade 7 | 0.79 | 0.64 | 0.78 | 0.64 | 0.79 | 0.67 | 0.80 | 0.70 | 0.78 | 0.77 | 0.75 | 0.74 | 0.81 | 0.81 |

| | | | | | | | | | | | | | | |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Total | 0.82 | 0.78 | 0.81 | 0.76 | 0.83 | 0.81 | 0.84 | 0.84 | 0.84 | 0.87 | 0.84 | 0.69 | 0.85 | 0.77 |
| Narrative (7 min) | | | | | | | | | | | | | | |
| Grade 7 | 0.75 | 0.63 | 0.73 | 0.63 | 0.76 | 0.66 | 0.81 | 0.75 | 0.85 | 0.82 | 0.80 | 0.72 | 0.88 | 0.79 |
| Expository (3 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.64 | ns | 0.71 | ns | 0.62 | ns | 0.67 | ns | 0.70 | ns | 0.62 | ns | 0.77 | ns |
| Grade 5 | 0.74 | 0.73 | 0.70 | 0.74 | 0.75 | 0.77 | 0.74 | 0.88 | 0.63 | 0.84 | 0.54 | 0.41 | 0.55 | 0.83 |
| Grade 7 | 0.59 | 0.76 | 0.55 | 0.74 | 0.57 | 0.74 | 0.62 | 0.79 | 0.69 | 0.82 | ns | 0.65 | 0.49 | 0.69 |
| Total | 0.70 | 0.69 | 0.66 | 0.67 | 0.70 | 0.70 | 0.74 | 0.79 | 0.74 | 0.82 | 0.49 | 0.46 | 0.65 | 0.79 |
| Expository (5 min) | | | | | | | | | | | | | | |
| Grade 3 | 0.55 | ns | 0.60 | ns | 0.52 | ns | 0.62 | ns | 0.75 | ns | 0.64 | ns | 0.78 | ns |
| Grade 5 | 0.76 | 0.76 | 0.73 | 0.76 | 0.79 | 0.78 | 0.82 | 0.89 | 0.77 | 0.85 | 0.69 | ns | 0.69 | 0.83 |
| Grade 7 | 0.67 | 0.72 | 0.66 | 0.72 | 0.69 | 0.74 | 0.75 | 0.79 | 0.82 | 0.81 | 0.45 | 0.73 | 0.67 | 0.70 |
| Total | 0.75 | 0.70 | 0.73 | 0.68 | 0.77 | 0.71 | 0.82 | 0.81 | 0.83 | 0.83 | 0.65 | 0.43 | 0.76 | 0.80 |
| Expository (7 min) | | | | | | | | | | | | | | |
| Grade 7 | 0.75 | 0.70 | 0.74 | 0.70 | 0.78 | 0.72 | 0.82 | 0.76 | 0.87 | 0.78 | 0.69 | 0.64 | 0.79 | 0.64 |

Note. All correlations are significant, $p < .001$, unless indicated as ns (not significant) or na (not applicable; these correlations were not computed due to the non-normality of the score distributions or because they were not administered in spring). Grade 3 Fall $n = 25$, Spring $n = 21$; Grade 5 Fall $n = 43$, Spring $n = 32$; Grade 7 Fall $n = 55$, Spring $n = 41$. LW = legible words; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.

Table 4

Descriptive Statistics for Criterion Measures

| | Grade 3 | | Grade 5 | | Grade 7 | | Total | |
|---------------------|---------|---------|---------|---------|---------|---------|-------|---------|
| | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) |
| TOWL-3 | 21.65 | (11.73) | 24.12 | (12.38) | 32.98 | (11.94) | 27.97 | (12.94) |
| MCA Writing Subtest | -- | -- | 33.88 | (29.64) | -- | -- | -- | -- |
| Language Arts GPA | -- | -- | -- | -- | 2.40 | (1.42) | -- | -- |

Note. TOWL-3 = Test of Written Language - Third Edition; MCA = Minnesota Comprehensive Assessment; GPA = grade point average.

Table 5

Criterion Validity Correlations Between Spring Writing Measures and TOWL-3, MCA, and Language Arts GPA

| | | LW | WW | WSC | CWS | CIWS | %WSC | %CWS |
|--------------------|--------|-------|------|--------------|--------------|--------------|--------------|--------------|
| Hard Copy | | | | | | | | |
| Grade 3 | TOWL-3 | ns | ns | .55* | .66** | .62** | na | na |
| Grade 7 | TOWL-3 | .43* | .39* | .41* | .42* | .41* | na | na |
| Picture (3 min) | | | | | | | | |
| Grade 3 | TOWL-3 | ns | ns | ns | .63* | .62* | ns | ns |
| Grade 7 | TOWL-3 | ns | ns | ns | ns | ns | ns | .49** |
| | GPA | ns | ns | ns | ns | .45* | .46* | .60** |
| Picture (5 min) | | | | | | | | |
| Grade 3 | TOWL-3 | ns | ns | .60* | .70** | ns | ns | .60* |
| Grade 7 | TOWL-3 | ns | ns | ns | ns | ns | .45* | .58** |
| | GPA | ns | ns | ns | ns | ns | .51** | .60** |
| Picture (7 min) | | | | | | | | |
| Grade 7 | TOWL-3 | ns | ns | ns | .40* | .55** | .47* | .59** |
| | GPA | .40* | ns | .47* | .57** | .67** | .49** | .62** |
| Narrative (3 min) | | | | | | | | |
| Grade 3 | TOWL-3 | ns | ns | ns | .63* | .70** | ns | .56* |
| Grade 5 | TOWL-3 | ns | ns | ns | ns | .62* | ns | .61* |
| | MCA | ns | ns | ns | .54* | .62** | ns | .59** |
| Grade 7 | GPA | .42* | ns | .44* | .55** | .65** | .46* | .61** |
| Narrative (5 min) | | | | | | | | |
| Grade 3 | TOWL-3 | ns | ns | ns | .66** | .68** | ns | .58* |
| Grade 5 | TOWL-3 | ns | ns | ns | ns | .65* | ns | .63* |
| | MCA | ns | ns | ns | .56** | .68** | ns | .56** |
| Grade 7 | GPA | ns | ns | .43* | .59** | .71** | .55** | .71** |
| Narrative (7 min) | | | | | | | | |
| Grade 7 | GPA | .40* | ns | .47** | .57** | .72** | .61** | .70** |
| Expository (3 min) | | | | | | | | |
| Grade 5 | MCA | ns | ns | ns | .45* | .54** | .45* | .48* |
| Grade 7 | TOWL-3 | ns | ns | ns | ns | .44* | ns | ns |
| | GPA | .44* | .42* | .53* | .61** | .67** | .41** | .51** |
| Expository (5 min) | | | | | | | | |
| Grade 5 | TOWL-3 | ns | ns | ns | ns | .57* | ns | .57* |
| | MCA | ns | ns | .45* | .55** | .60** | .46* | .51* |
| Grade 7 | TOWL-3 | ns | ns | ns | .42* | .48* | .43* | .48* |
| | GPA | .45* | ns | .48* | .59** | .66** | .48* | .48* |
| Expository (7 min) | | | | | | | | |
| Grade 7 | TOWL-3 | ns | ns | ns | .45* | .52* | .43* | .51** |
| | GPA | .49** | .45* | .52** | .62** | .68** | .49** | .51** |

Note. LW = legible words; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; CIWS = correct minus incorrect word sequences; TOWL-3 = Test of Written Language - 3rd Edition; MCA = Minnesota Comprehensive Assessment; GPA = Language arts grade point average.

Table 6

Descriptive Statistics for Measures Showing Reliable Fall-to-Spring Growth

| | Fall | | Spring | | Mean Difference | Gain per week |
|------------------|-------|---------|--------|---------|--------------------|---------------------|
| | Mean | (SD) | Mean | (SD) | | |
| Grade 3 | | | | | | |
| Picture 3 min | | | | | | |
| CWS | 11.59 | (9.48) | 19.00 | (18.02) | 7.41 ** | 0.37 |
| Grade 5 | | | | | | |
| Narrative 3 min | | | | | | |
| CWS | 29.67 | (15.55) | 40.43 | (19.73) | 10.76 ** | 0.54 |
| CIWS | 14.47 | (21.85) | 24.67 | (25.11) | 10.20 ** | 0.51 |
| Narrative 5 min | | | | | | |
| CWS | 46.22 | (25.81) | 60.59 | (31.73) | 14.37 ** | 0.72 |
| CIWS | 22.38 | (33.00) | 36.28 | (37.53) | 13.91 ** | 0.70 |
| Expository 3 min | | | | | | |
| CIWS | 16.50 | (18.01) | 26.03 | (25.33) | 9.53 *** | 0.48 |
| Expository 5 min | | | | | | |
| CWS | 46.63 | (21.32) | 64.88 | (29.94) | 18.25 *** | 0.91 |
| CIWS | 20.91 | (29.27) | 38.75 | (38.69) | 17.84 *** | 0.89 |
| %CWS | 63.91 | (18.32) | 69.82 | (18.21) | 5.90 ** | 0.30 |
| Grade 7 | | | | | | |
| Narrative 7 min | | | | | | |
| CIWS | 59.55 | (41.32) | 71.36 | (42.93) | 11.82 ** | 0.59 |
| %WSC | 86.43 | (8.39) | 89.53 | (7.38) | 3.10 * | 0.16 |
| %CWS | 72.67 | (14.58) | 76.63 | (13.99) | 3.96 * | 0.20 |
| Expository 3 min | | | | | | |
| WSC | 44.15 | 16.65 | 45.77 | 16.75 | 1.62 | 0.08 |
| Expository 5 min | | | | | | |
| CIWS | 41.69 | (30.99) | 50.21 | (28.47) | 8.51 * | 0.43 |
| Expository 7 min | | | | | | |
| %CWS | 71.70 | (14.95) | 77.41 | (13.19) | 5.71 ** | 0.29 |

Note. CWS = correct word sequences; CIWS = correct minus incorrect word sequences; WSC = words spelled correctly.

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7

Repeated Measures Analyses of Variance for Sufficiently Reliable and Valid Measures

| Measure | <i>df</i> | <i>F</i> | | <i>p</i> |
|---------------------------|-----------|----------|-----|----------|
| Grade 3 | | | | |
| Picture Prompt (3 min) | | | | |
| CWS | 1, 16 | 8.72 | ** | .009 |
| CIWS | 1, 16 | 3.08 | | .099 |
| Grade 5 | | | | |
| Narrative Prompt (3 min) | | | | |
| CWS | 1, 29 | 12.86 | ** | .001 |
| CIWS | 1, 29 | 8.61 | ** | .006 |
| %CWS | 1, 29 | 2.97 | | .096 |
| Narrative Prompt (5 min) | | | | |
| CWS | 1, 31 | 10.81 | ** | .003 |
| CIWS | 1, 31 | 10.01 | ** | .003 |
| %CWS | 1, 31 | 3.31 | | .079 |
| Expository Prompt (5 min) | | | | |
| CWS | 1, 29 | 28.44 | *** | .000 |
| CIWS | 1, 29 | 17.24 | *** | .000 |
| %CWS | 1, 29 | 8.51 | ** | .007 |
| Grade 7 | | | | |
| Narrative Prompt (7 min) | | | | |
| CWS | 1, 42 | 43.28 | ** | .000 |
| CIWS | 1, 42 | 5.63 | * | .022 |
| %WSC | 1, 42 | 2.85 | * | .033 |
| %CWS | 1, 42 | 3.22 | | .080 |
| Expository Prompt (5 min) | | | | |
| CWS | 1, 38 | 0.68 | | .420 |
| CIWS | 1, 38 | 4.45 | * | .040 |
| Expository Prompt (7 min) | | | | |
| WSC | 1, 40 | 0.00 | | .997 |
| CWS | 1, 40 | 0.07 | | .795 |
| CIWS | 1, 40 | 3.12 | | .085 |
| %CWS | 1, 40 | 9.97 | ** | .003 |

Note. LW = legible words; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; CIWS = correct minus incorrect word sequences.

* $p < .05$

** $p < .01$

*** $p < .001$

Table 8

Measures With Sufficient Alternate-Form Reliability and Criterion Validity, and Reliable Fall-to-Spring Growth

| Task/Duration | Scoring | Grade 3 | | | Grade 5 | | | Grade 7 | | |
|-------------------|---------|-----------------------------|---------------------|------------------------|-----------------------------|---------------------|------------------------|-----------------------------|---------------------|------------------------|
| | | Alternate-Form Reliability? | Criterion Validity? | Fall to Spring Growth? | Alternate-Form Reliability? | Criterion Validity? | Fall to Spring Growth? | Alternate-Form Reliability? | Criterion Validity? | Fall to Spring Growth? |
| Easy Copy 1.5 min | LW | Yes | -- | -- | Yes | -- | -- | No | -- | -- |
| | WW | Yes | -- | -- | Yes | -- | -- | No | -- | -- |
| | WSC | Yes | -- | -- | Yes | -- | -- | No | -- | -- |
| | CWS | Yes | -- | -- | Yes | -- | -- | No | -- | -- |
| | CIWS | Yes | -- | -- | No | -- | -- | No | -- | -- |
| Hard Copy 1.5 min | LW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WSC | Yes | TOWL-3 | No | Yes | No | -- | Yes | No | -- |
| | CWS | Yes | TOWL-3 | No | Yes | No | -- | Yes | No | -- |
| | CIWS | Yes | TOWL-3 | No | Yes | No | -- | Yes | No | -- |
| Picture 3 min | LW | Yes | No | -- | Yes | -- | -- | No | No | -- |
| | WW | Yes | No | -- | Yes | -- | -- | No | No | -- |
| | WSC | Yes | No | -- | Yes | -- | -- | No | No | -- |
| | CWS | Yes | TOWL-3 | Yes | Yes | -- | -- | Yes | No | -- |
| | CIWS | Yes | TOWL-3 | No | Yes | -- | -- | Yes | No | -- |
| | %WSC | Yes | No | -- | Yes | -- | -- | Yes | No | -- |
| | %CWS | Yes | No | -- | Yes | -- | -- | Yes | GPA | No |
| Picture 5 min | LW | Yes | No | -- | Yes | -- | -- | Yes | No | -- |
| | WW | Yes | No | -- | Yes | -- | -- | Yes | No | -- |
| | WSC | Yes | TOWL-3 | No | Yes | -- | -- | Yes | No | -- |
| | CWS | Yes | TOWL-3 | No | Yes | -- | -- | Yes | No | -- |

| | | | | | | | | | | |
|-----------------|------|-----|--------|----|-----|-------------|-----|-----|-------------|----|
| | CIWS | Yes | No | -- | Yes | -- | -- | Yes | No | -- |
| | %WSC | Yes | No | -- | Yes | -- | -- | Yes | GPA | No |
| | %CWS | Yes | TOWL-3 | No | Yes | -- | -- | Yes | TOWL-3, GPA | No |
| Picture 7 min | LW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WSC | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | CWS | -- | -- | -- | -- | -- | -- | Yes | GPA | No |
| | CIWS | -- | -- | -- | -- | -- | -- | Yes | TOWL-3, GPA | No |
| | %WSC | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | %CWS | -- | -- | -- | -- | -- | -- | Yes | TOWL-3, GPA | No |
| Narrative 3 min | LW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WSC | Yes | TOWL-3 | No | Yes | No | -- | Yes | No | -- |
| | CWS | Yes | TOWL-3 | No | Yes | MCA | Yes | Yes | GPA | No |
| | CIWS | Yes | TOWL-3 | No | Yes | MCA, TOWL-3 | Yes | Yes | GPA | No |
| | %WSC | Yes | No | -- | Yes | No | -- | Yes | No | No |
| | %CWS | Yes | TOWL-3 | No | Yes | MCA, TOWL-3 | No | Yes | GPA | No |
| Narrative 5 min | LW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WW | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | WSC | Yes | No | -- | Yes | No | -- | Yes | No | -- |
| | CWS | Yes | TOWL-3 | No | Yes | MCA | Yes | Yes | GPA | No |
| | CIWS | Yes | TOWL-3 | No | Yes | MCA, TOWL-3 | Yes | Yes | GPA | No |
| | %WSC | Yes | No | No | Yes | No | -- | Yes | GPA | No |
| | %CWS | Yes | TOWL-3 | No | Yes | MCA, TOWL-3 | No | Yes | GPA | No |
| Narrative 7 min | LW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WSC | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | CWS | -- | -- | -- | -- | -- | -- | Yes | GPA | No |

| | | | | | | | | | | |
|------------------|------|------------|----|----|------------|--------------------|------------|------------|--------------------|------------|
| | CIWS | -- | -- | -- | -- | -- | -- | Yes | GPA | Yes |
| | %WSC | -- | -- | -- | -- | -- | -- | Yes | GPA | Yes |
| | %CWS | -- | -- | -- | -- | -- | -- | Yes | GPA | Yes |
| Expository 3 min | LW | Yes | -- | -- | Yes | No | -- | Yes | No | -- |
| | WW | Yes | -- | -- | Yes | No | -- | Yes | No | -- |
| | WSC | Yes | -- | -- | Yes | No | -- | Yes | GPA | No |
| | CWS | Yes | -- | -- | Yes | No | -- | Yes | GPA | No |
| | CIWS | Yes | -- | -- | Yes | MCA | Yes | Yes | GPA | No |
| | %WSC | Yes | -- | -- | Yes | No | -- | Yes | No | -- |
| | %CWS | Yes | -- | -- | Yes | No | -- | Yes | GPA | No |
| Expository 5 min | LW | No | -- | -- | Yes | No | -- | Yes | No | -- |
| | WW | No | -- | -- | Yes | No | -- | Yes | No | -- |
| | WSC | No | -- | -- | Yes | No | -- | Yes | No | -- |
| | CWS | Yes | -- | -- | Yes | MCA | Yes | Yes | GPA | No |
| | CIWS | Yes | -- | -- | Yes | TOWL-3 | Yes | Yes | GPA | Yes |
| | %WSC | Yes | -- | -- | Yes | No | -- | Yes | No | No |
| | %CWS | Yes | -- | -- | Yes | MCA, TOWL-3 | Yes | Yes | No | No |
| Expository 7 min | LW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WW | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | WSC | -- | -- | -- | -- | -- | -- | Yes | GPA | No |
| | CWS | -- | -- | -- | -- | -- | -- | Yes | GPA | No |
| | CIWS | -- | -- | -- | -- | -- | -- | Yes | TOWL-3, GPA | No |
| | %WSC | -- | -- | -- | -- | -- | -- | Yes | No | -- |
| | %CWS | -- | -- | -- | -- | -- | -- | Yes | TOWL-3, GPA | Yes |

Note. LW = legible words; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; CIWS = correct minus incorrect word sequences; TOWL-3 = Test of Written Language - 3rd Edition; MCA = Minnesota Comprehensive Assessment; GPA = Language arts grade point average. Blanks indicate that conclusions were not determined in this study.