



TECHNICAL REPORT #38:

Examining Technical Features of Progress Monitoring Measures Across Grade Levels in Writing

*Kristen L. McMaster, Miya Miura Wayman, Stanley L. Deno,
Christine A. Espin, and Seungsoo Yeo*

RIPM Year 3: 2005 – 2006

Dates of Study: January 2006 – May 2006

December 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.

Examining Technical Features of Progress Monitoring Measures Across Grade Levels in Writing

Background

Writing is critical for students' successful performance in school and later life. To determine whether students are making sufficient progress in writing, measures are needed that produce both reliable and valid performance information, as well as reliable pictures of growth over time. Curriculum-based measurement (CBM; Deno, 1985) is one approach that has historically yielded technically sound information in core academic areas (reading, writing, spelling, mathematics). Whereas some research has established the technical adequacy of scores from CBM in writing (CBM-W; see McMaster & Espin, 2007 for review), little research has been conducted to determine the technical adequacy of slopes produced from frequently collected CBM-W progress monitoring data. Recent research has been conducted with beginning (first-grade) writers (McMaster et al., 2011); in this study, we replicated procedures used in that study with children in Grades 2-5.

In a previous study examining technical adequacy of scores produced from CBM-W prompts administered to students in Grades 3, 5, and 7 (McMaster & Campbell, 2008), narrative prompts administered for 5 min and scored for correct word sequences (CWS, a measure that takes into account both spelling and grammar; Videen, Deno, & Marston, 1982) and correct minus incorrect word sequences (Espin, Scierka, Skare, & Halverson, 1999) yielded reliable and valid scores across grade levels. For students in Grades 5 and 7, expository (descriptive) prompts scored for CWS and CIWS also yielded sufficiently reliable and valid information about student writing. Expository prompts were not administered to students in Grade 3 in the previous study. In this study, we replicate and extend McMaster & Campbell's findings by administering both

narrative and expository prompts to students in Grades 2-5, and examining technical adequacy of both performance level scores and slopes from progress monitoring data.

Research Questions

- (1) For each grade level (Grades 2-3 and 4-5), do alternate-form reliability and criterion validity coefficients obtained from 5-min narrative and expository prompts scored for CWS and CIWS replicate those found in previous research? Do reliability and criterion validity coefficients increase when scores are aggregated (by calculating the mean of two adjacent weeks)? We expected alternate-form reliabilities to be $r > .70$, and criterion-related reliability to be $r > .50$, as in previous research (McMaster & Campbell, 2008). We expected aggregated scores to yield slightly stronger correlation coefficients.
- (2) Does student performance level differ significantly by grade level (Grades 2-3 and 4-5)? We expected students in Grades 4-5 to outperform students in Grades 2-3 on both tasks (narrative and expository prompts) and both scoring procedures (CWS and CIWS).
- (3) For each grade level (Grades 2-3 and 4-5), do the measures yield reliable, stable slopes, and how many data points are needed to obtain these reliable/stable slopes? Does the reliability and stability of slopes increase when scores are aggregated (by calculating the mean of two adjacent weeks)? We expected that the more data points, the greater the reliability and stability of slopes (cf. Christ, 2006; Hintze & Christ, 2004; McMaster et al., 2011).
- (4) For each grade level (Grades 2-3 and 4-5), are the measures sensitive to growth (i.e., do scores produce slopes that are significantly greater than 0) in a short time period? Does sensitivity to growth increase when scores are aggregated (by calculating the mean of two adjacent weeks)?

Method

Participants and Setting

The study took place in two urban elementary schools in a large urban Midwestern school district. Five classroom teachers participated. One teacher was a reading specialist who worked with students in Grade 3 in the reading resource room. One teacher taught a combined Grades 2-3 class, one taught Grade 3 only, one taught a combined Grades 4-5 class, and one taught Grade 5 only. Four teachers had master's degrees and elementary education teaching licenses; one also had an educational specialist degree and a principal's license. Years of teaching experience ranged from 20 to 27 years (mean of 23.25 years). One teacher did not provide demographic data despite repeated attempts to collect this information from her.

Participants in the study were recruited from the classrooms of the 5 participating teachers. There were 89 participants (43 male and 49 female), including 51 students in Grades 2/3 and 38 students in Grades 4/5. Seventy-one percent of the students were White, 3% Hispanic, 7% Asian, and 19% African American. Approximately 92% of the students spoke English as their home language. Other languages spoken in the home included Hmong (1%), Spanish (1%), Somali (4%), and Ewe (1%). Seven percent of the students were receiving English as a Second Language services, and 3% were receiving special education services. Approximately 36% of the students qualified for free or reduced lunch. All students with permission to participate were included in the study.

Measures

CBM tasks. Each student responded to narrative and expository writing prompts (all prompts are provided in Appendix A). To identify narrative and expository prompts for this study, four members of the research team (the first author and three special education doctoral

students) conducted the following steps. First, we gathered writing prompts used in previous CBM writing research. As part of previous RIPM activities, researchers had conducted a comprehensive literature review on CBM writing measures (see McMaster & Espin, 2007). All writing prompts reported in studies included in this review were compiled into a database. Second, we gathered additional prompts from internet resources (using a Google search for story starters and writing prompts). Third, we created additional prompts. Fourth, we compiled all prompts identified or created through the above process into one list. Fifth, we narrowed the list of prompts to those that were least ambiguous and reflected experiences to which most U.S. public school students would be able to relate. Sixth, we submitted the list of prompts to the larger research team (three senior-level faculty in special education, four special education doctoral students, and one school psychology student) to rank the prompts from their most to least favorite (based on what they thought would be most interesting to students). We also asked each team member to indicate which prompts they believed would elicit narrative vs. expository writing. Seventh, those prompts that received the highest rankings were selected, and the narrative and expository designations were discussed and confirmed.

Prompts were rewritten, when necessary, so that each followed a consistent format and were simple in terms of vocabulary and sentence structure. Narrative prompts were written so that a scenario was clearly introduced, the prompt was an incomplete sentence with an ellipsis to indicate that the writer should continue with a story. Expository prompts began with the word “Describe” (thus eliciting descriptive writing) followed by the specific prompt, and ending with “tell why” (to prompt students to explain their stance).

Sample duration. For each prompt, teachers were asked to instruct students to make a slash at the 3 min mark and continue writing for a total of 5 min; however classroom

observations revealed inconsistent implementation of this instruction. This inconsistent implementation, combined with previous research indicating that 5-min samples yielded stronger reliability and validity across grades (McMaster & Campbell, 2008) led to our decision to use only the data from 5-min samples.

Scoring procedures. Writing samples were scored using the following procedures, based on previous research (McMaster & Campbell, 2008; see also McMaster & Espin, 2007 for review of previous research) indicating that these scoring procedures yield reliable and valid data within and across the grade levels of interest:

1. 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.
2. Correct minus incorrect word sequences (CIWS; Espin et al., 1999).

Criterion measure. The criterion measure used in the present study consisted of pre- and posttest narrative writing samples scored using an analytic rubric based on that used for the Test of Written Language – Third Edition (TOWL-3; Hammill & Larsen, 1996) Spontaneous Writing subtest. Students were presented with a narrative writing prompt, told to think of a story, 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).

Fidelity observation. A fidelity checklist was developed to document accuracy of CBM administration. The checklist included 8 items (hands out student copies for writing prompt, reads standardized directions, provides students 30 sec to think about the prompt, starts stopwatch at correct time, tells students to mark progress after 3 min, tells students to stop and

put pencils down at 5 min, collects all materials, follows standardized directions to provide assistance to students). Items were marked as either observed or not observed; the fidelity score was a percentage of items observed out of the total items.

Writing Instruction

Teachers were asked to describe their writing instruction. They reported that, on average, they spent between 25% and 50% of their reading/language arts instructional time on writing. The majority of this instructional time was spent on the content of writing; emphasis was also given to vocabulary and writing conventions. Less emphasis was given to writing production and fluency. Writing activities included creating research reports and book reports, conducting interviews, partner writing, dictionary use, journaling, producing written responses to comprehension questions on reading material, using the writing process (write topic, sloppy copy, type final copy), writing workshops, writing about subjects being studied in science/social studies, and writing in different genres (personal narrative, story, persuasive essay, poetry styles). All teachers reported using a mixture of independent work, whole group instruction, pairs and small groups, and one-to-one instruction. Teachers reported that students typically wrote between 15 and 50 min per day, with time distributed across planning, composing, and revising, and that students wrote one to two paragraphs per day. Teachers differentiated instruction for struggling writers by providing one-to-one instruction, individual dictionaries, written models, ELL instruction, partner revisions and editing, and opportunities to dictate to an adult.

Procedures

Measure administration. CBM narrative writing prompts were group administered weekly by students' classroom teachers across 12 consecutive weeks starting in January. Two additional narrative prompts were administered at Week 17. At Week 1 and Week 17, one of the

narrative prompts was administered for 5 min to obtain CBM scores. Students were instructed to circle the last word they wrote (at the 5 min mark), and were then given an additional 10 min (for a total of 15 min) to continue writing, producing pre- and posttest writing samples that were scored using the analytic rubric. CBM expository writing prompts were group administered in Weeks 1, 6, 12, and 18 (one prompt each week except Week 18, in which two prompts were given) at each grade level. All classroom teachers were trained to administer the measures during a one-hour training session. Fidelity of administration was checked by trained graduate students on two to three separate occasions using the fidelity checklist. The range of fidelity was 75% to 100% and the average fidelity was 98%.

Scoring training and reliability. A doctoral student in special education experienced in the development, administration, and scoring of CBM-W was designated as the “expert” scorer. She and four other scorers (one doctoral student in special education and three master’s students in school psychology) met with a professor of special education for one 2-hour session to describe, demonstrate, and practice the scoring procedures. Each scorer then scored the same six writing samples to establish inter-rater agreement. The expert compared each scorer’s results with her own, and calculated the percent of agreement for each scoring procedure by dividing the smaller score by the larger score and multiplying by 100. If interrater agreement was at least 85%, the scorer was given additional writing samples to score independently. If agreement was below 85%, the scorer was given additional training and scored additional writing samples until 85% agreement was reached.

To determine interrater reliability, 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. Average inter-scorer agreement across the different scoring procedures ranged from 80% to 99%, with more complex scoring procedures (e.g., CIWS) yielding lower inter-scorer agreement.

To score the pre- and post-test writing samples using the analytic rubric from the TOWL-3, the first author and a graduate student in school psychology 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 80% agreement was obtained on 10% of the samples. We then divided and independently scored the remaining samples. Inter-scorer agreement across the different scoring procedures ranged from 84% to 92%, with a mean of 84% (with higher interrater reliability on Contextual Conventions and lower interrater reliability on Story Construction).

Results

Alternate-form Reliability and Criterion-related Validity

We calculated descriptive statistics of student performance level on each CBM probe, first for each single score (for narrative and expository prompts), and then for aggregated scores (for narrative prompts only; determined by computing the mean of scores from Weeks 1-2, 3-4, 5-6, 7-8, 9-10, 11-12, and 17a and 17b). Means and *SDs* by task, scoring procedure, and grade level are reported in Table 1. We also calculated means and *SDs* for the three analytic scores from the pre- and posttest writing measures (see Table 2).

Alternate-form reliability. We calculated alternate-form reliability coefficients using Pearson r bivariate correlations, along with 95% confidence intervals (CIs) using Fisher's r to z transformations. Correlation coefficients and CIs by task, scoring procedure, and grade level are reported in Table 3. Because of the large number of correlation coefficients calculated, we used Bonferroni corrections for each set of coefficients calculated (e.g., for the narrative single-score correlations, there were 44 comparisons, so we set the p -value at .001). For the narrative task, we first correlated scores from adjacent weeks, and then correlated adjacent aggregated scores. For Grades 2-3, single-score alternate-form reliability coefficients ranged from $r = .73$ to $.90$ for CWS, and from $r = .70$ to $.85$ for CIWS; aggregated-score coefficients ranged from $r = .82$ to $.95$ for CWS, and from $r = .83$ to $.92$ for CIWS. For Grades 4-5, single-score alternate-form reliability coefficients ranged from $r = .72$ to $.91$ for CWS, and from $r = .63$ to $.89$ for CIWS; aggregated-score coefficients ranged from $r = .86$ to $.92$ for CWS, and from $r = .79$ to $.91$ for CIWS. All p s were less than .001.

Because the expository prompts were given at 6-week intervals, correlating scores from adjacent weeks was not possible; however, students responded to two separate expository prompts at Week 18. For CWS, r s = $.85$ and $.75$ for Grades 2-3 and 4-5, respectively. For CIWS, r s = $.66$ and $.74$ for Grades 2-3 and 4-5, respectively. Again, all p s were less than .001.

Criterion-related validity. To determine criterion-related validity, we calculated concurrent correlation coefficients between CBM scores and pre- and posttest analytic writing scores using Pearson r bivariate correlations, along with 95% CIs using Fisher's r to z transformations. Correlation coefficients and CIs by task, scoring procedure, and grade level are reported in Table 4. Again, we used a Bonferroni correction for multiple comparison (72), setting the p -value at $< .001$. In Table 4, correlations that were significant at $p < .001$ are bolded.

For the narrative prompts, we correlated pretest scores with single scores from Week 1 and aggregated scores from Weeks 1-2, and posttest scores with aggregated scores from Week 17a and 17b, because these were the closest points in time to obtain concurrent validity coefficients. For Grades 2-3, single-score criterion-related validity coefficients with pretest scores ranged from $r = .29$ to $.63$ for CWS, and from $r = .24$ to $.73$ for CIWS. Aggregated-score coefficients with pretest scores ranged from $r = .40$ to $.66$ for CWS and from $r = .28$ to $.76$ for CIWS. Coefficients with posttest scores ranged from $r = .34$ to $.55$ for CWS and from $r = .55$ to $.74$ for CIWS. For Grades 4-5, single-score criterion-related validity coefficients with pretest scores ranged from $r = .46$ to $.61$ for CWS, and from $r = .41$ to $.76$ for CIWS. Aggregated-score coefficients with pretest scores ranged from $r = .38$ to $.69$ for CWS, and from $r = .39$ to $.77$ for CIWS. Coefficients with posttest scores were all nonsignificant for CWS and ranged from $r = .48$ to $.71$ for CIWS.

For the expository prompts, we correlated pretest scores with Week 1 scores, and posttest scores with the mean of the Week 18 scores, as these were the closest points in time to obtain concurrent validity coefficients. For Grades 2-3, criterion-related validity coefficients ranged from $r = .38$ to $.64$ for CWS and from $r = .42$ to $.70$ for CIWS. For Grades 4-5, coefficients ranged from $r = .36$ to $.62$ for CWS, and from $r = .43$ to $.73$ for CIWS.

Grade-Level Differences on CBM Performance Levels

To determine whether students' mean CBM performance levels differed by grade level, we ran multivariate analyses of variance (MANOVA) using grade level (Grades 2-3 vs. Grades 4-5) as a between-groups variable and weekly CBM scores as the dependent variable. We ran separate MANOVAs for each type of task and scoring procedure, using single scores for both narrative and expository prompts, and also using aggregated scores for narrative prompts.

Statistically significant MANOVAs were followed up with univariate ANOVAs. Effect sizes were calculated using Cohen's d (mean difference divided by the pooled SD).

For single narrative CWS, the MANOVA was statistically significant (Wilks' Lambda = .502, $F[14, 57] = 4.04$, $p < .001$). For aggregated narrative CWS, the MANOVA was statistically significant (Wilks' Lambda = .521, $F[7, 81] = 10.62$, $p < .001$). All follow-up ANOVAs were statistically significant at the $p < .001$ level (see Table 5); on average, students in Grades 4-5 outperformed students in Grades 2-3, with most effect sizes greater than $d = 1.00$.

For single narrative CIWS, the MANOVA was statistically significant (Wilks' Lambda = .452, $F[14, 55] = 4.76$, $p < .001$). For aggregated narrative CIWS, the MANOVA was statistically significant (Wilks' Lambda = .525, $F[7, 81] = 10.48$, $p < .001$). All follow-up ANOVAs were statistically significant at the $p < .001$ level (see Table 5). Again, on average, students in Grades 4-5 outperformed students in Grades 2-3, with most effect sizes greater than $d = 1.00$.

For expository CWS, the MANOVA was statistically significant (Wilks' Lambda = .593, $F[4, 82] = 14.04$, $p < .001$). All follow-up ANOVAs were statistically significant at the $p < .001$ level (see Table 5); on average, students in Grades 4-5 outperformed students in Grades 2-3, with all effect sizes greater than $d = 1.00$. For expository CIWS, the MANOVA was statistically significant (Wilks' Lambda = .608, $F[4, 82] = 13.22$, $p < .001$). All follow-up ANOVAs were statistically significant at the $p < .001$ level (see Table 5); on average, students in Grades 4-5 outperformed students in Grades 2-3, with all effect sizes greater than $d = 1.00$.

Reliability and Stability of CBM Slopes

Next, we calculated descriptive statistics of student slopes produced by weekly CBM probes, using slopes produced by (a) single consecutive scores (both narrative and expository), (b) scores from every other week (narrative only), and (c) aggregated scores (narrative only). For

narrative single scores and aggregated scores, we calculated slopes in increments, starting with the first three data points, and adding one data point at a time until reaching the full slope (using all 12 data points for single scores, and all 6 data points for aggregated scores). For expository scores, were limited to calculating just one slope, because there were only 4 data points available. Means and *SDs* of slopes by task, scoring procedure, and grade level are reported in Table 6.

Reliability of slopes. To determine reliability of slopes, we followed procedures described in McMaster et al. (2011). (This analysis was only possible for narrative slopes.) First, we calculated alternate-slope reliability, using slopes generated from data points from odd and even weeks. Alternate-slope reliability coefficients between odd and even CWS slopes were $r = .45$ ($p = .001$) for Grades 2-3 and $r = .39$ ($p = .390$) for Grades 4-5. Alternate-slope reliability coefficients between odd and even CIWS slopes were $r = .55$ ($p < .001$) for Grades 2-3 and $r = .12$ ($p = .490$) for Grades 4-5.

Next, we determined incremental slope reliability, by correlating slopes produced from incrementally-added data points with the overall slopes produced from both single scores (12 data points) and aggregated scores (6 data points). Results are shown in Figures 1-4. Using criteria from McMaster et al., slopes produced from single narrative CWS scores were “sufficiently” ($r > .70$) correlated with the overall slope when 8 or more data points were used for Grades 2-3, and 9 or more data points were used for Grades 4-5. Slopes produced from aggregated narrative CWS scores were sufficiently ($r > .70$) correlated with the overall slope when 4 or more data points were used for Grades 2-3 and for Grades 4-5.

Slopes produced from single narrative CIWS scores were sufficiently ($r > .70$) correlated with the overall slope when 8 or more data points were used for Grades 2-3, and 10 or more data

points were used for Grades 4-5. Slopes produced from aggregated narrative CWS scores were sufficiently ($r > .70$) correlated with the overall slope when 4 or more data points were used for Grades 2-3, but 5 points were needed for Grades 4-5.

Stability of slopes. To further examine the stability of slopes, we calculated the standard error of the estimate (SEE) and the standard error of the slope (SEb) of incremental slopes produced from single and aggregated scores. Descriptive data are presented in Table 7. As can be seen in Table 7, for narrative prompts, the SEb decreases substantially as data points are added to slopes. For expository prompts, SEbs are quite large in relation to the slope, indicating that these slopes are unstable.

Sensitivity to Growth

To gauge whether CBM slopes reflected student growth, we calculated t -values to determine whether slopes were statistically significantly greater than zero. Descriptive data are presented in Table 7. None of the slopes for narrative or expository prompts and scoring procedures were statistically significantly greater than zero (all $ps > .05$).

We also conducted a repeated-measures MANOVAs (RM-MANOVAs) using pre- and posttest analytic writing scores, to determine whether students made writing gains on an independent measure of writing. We included time as the within-subjects factor, and grade level as a between-groups factor. There were statistically significant main effects of both grade level (Wilks' Lambda = .727, $F[3, 83] = 10.40$, $p < .001$) and time (Wilks' Lambda = .819, $F[3, 83] = 6.10$, $p = .001$), but no grade level by time interaction (Wilks' Lambda = .947, $F[3, 83] = 1.55$, $p = .207$). Follow-up RM-ANOVAs (reported in Table 8) indicated main effects of time for both Contextual Conventions and Contextual Language, but not Story Construction, with students performing higher, on average, at posttest compared to pretest. Similarly, there were main effects

of group for both Contextual Conventions and Contextual Language, but not Story Construction, with students in Grades 4-5 outperforming students in Grades 2-3.

Next Steps

We have further plans to conduct longitudinal analyses with these data that will enable us to examine whether there is individual variation in growth rates and whether students' growth varies at different performance levels (e.g., based on lower- vs. higher- performing writers).

Discussion

Alternate-Form Reliability and Criterion-Related Validity

As expected, alternate-form reliability coefficients replicated those of previous research (McMaster & Campbell, 2008), with most coefficients of $r > .70$. In fact, many coefficients were greater than .80, suggesting that, over time, the alternate-form reliability of the narrative and expository prompts, scored for CWS and CWIS, yield scores with consistently strong reliability. Aggregating narrative scores appeared to strengthen reliability somewhat, with all coefficients near or above $r = .80$, and some above $r = .90$. This finding is consistent with earlier research (Fuchs, Deno, & Marston, 1982) that indicated that aggregating CBM writing scores increased reliability. The tradeoff is that more prompts must be administered and scored (which is time consuming) to obtain a single data point; given that single scores appear to have acceptable reliability, it may not be worth the extra time needed to collect enough data to aggregate scores.

Criterion-related validity coefficients were also similar to those found in previous research (McMaster & Campbell, 2008), with most coefficients of $r > .50$ between both narrative and expository prompts scored for CWS and CIWS and two of the analytic scoring subtests: Contextual Conventions and Contextual Language. However, fewer coefficients exceeded $r = .50$ between the CBM tasks and Story Construction. Thus, it appears that the measures may be more

strongly associated with writing conventions such as capitalization, punctuation, and spelling, and language conventions related to vocabulary, sentence construction, and grammar, than to the quality of compositional elements such as plot, prose, character development, and interest. Lower coefficients with Story Construction might also be related to the fact that it was more difficult to achieve interrater agreement on this subtest.

Validity coefficients for CIWS appeared to be slightly higher than were those for CWS, although confidence intervals overlapped. Aggregating narrative scores did not appear to alter the strength of correlation coefficients (CIs overlap considerably). This finding, along with the limited increase in alternate-form reliability obtained by aggregating scores, suggests that it might not be worth the time needed to collect enough data to aggregate scores.

Grade-Level Differences

Older students (Grades 4-5) reliably outperformed younger students (Grades 2-3) on both types of tasks (narrative and expository) and both types of scoring procedures (CWS and CIWS). These grade-level differences are consistent with previous research (see McMaster & Espin, 2007) and suggest that the measures are sensitive to differences in students' writing development.

Reliability and Stability of Slopes

We examined reliability and stability of slopes using several different approaches. First, alternate-slope reliability analyses conducted with narrative data revealed that slopes produced from data from odd and even weeks were relatively weak; for Grades 4-5, coefficients were nonsignificant. This result is likely due, at least in part, to the relatively few number of data points (6) contributing to each slope.

To determine how many data points are needed to achieve a slope with sufficient reliability ($r > .70$), we examined correlations of incremental slopes with the overall 12-week slope. For Grades 2-3, it took 8 weeks to achieve reliable slopes using single scores for both CWS and CIWS. When scores were aggregated, two data points were needed (but it is important to remember that these slopes were comprised of 8 weeks' worth of scores, so either way, it appears that 8 prompts must be given to obtain slopes that are sufficiently correlated with the overall slope). Also, this analysis assumes that the full slope represents the student's "true slope," which may not actually be the case given that we only collected 12 weeks of data (i.e., data collected over longer time periods may yield a different overall slope). It is logical that the more data points collected, the more strongly they are correlated with the overall slope, since they use much of the same data. Future research is needed to see if 8 weeks are sufficient when slopes are comprised of even more data (e.g., across an entire school year).

For Grades 4-5, it took 9 to 10 weeks to achieve reliable slopes using single scores for CWS and CIWS, respectively. When scores were aggregated, two and three data points were needed, respectively. The same limitations mentioned above apply here.

To determine stability of slopes produced from both narrative and expository data, we examined the SEb of incremental slopes (narrative only) and of the full slopes (expository). As expected, for narrative prompts, SEb decreased as data points were added.

Sensitivity to Growth

Finally, we examined whether the measures were sensitive to growth made in short time periods, by calculating t -values to determine whether slopes were statistically significantly greater than zero. None of the t -values were significant, indicating that slopes were not, on

average, sensitive to growth in the time period that writing was assessed. There are several possible explanations for this lack of growth on the CBM tasks:

1. Students might not actually have been making progress in writing. However, the repeated measures analysis of pre- and posttest analytic writing indicated that students did, on average, make gains in writing, at least on the Contextual Conventions and Contextual Language subtests of the TOWL. Gains were not as evident on the Story Construction subtest. Recall that this subtest was also not as strongly correlated with the CBM tasks (which is likely related, at least in part, to the lower reliability of this subtest).
2. The measures may truly not be sensitive to writing growth made in short time periods
3. Related to the above, it is possible that the tasks, duration, and/or scoring procedures used might not be appropriate for assessing growth. For example, it may be the case that, as students become more skilled in their writing, production (fluency) actually decreases (i.e., more-skilled writers may begin devoting more attention to planning and revising rather than simply knowledge-telling; McCutchen, 2006). This might not be the best approach to assessing typical writing development.
4. It is possible that the measures are more appropriate for struggling writers/students with LD. This question will be examined using longitudinal modeling techniques in future work.

References

- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232.
- Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*, 128–133.
- 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.
- Fuchs, L. S., Deno, S. L., & Marston, D. (1982). *Use of aggregation to improve the reliability of simple direct measures of academic performance* (Vol. IRLD-RR-94). University of Minnesota, Institute for Research on Learning Disabilities.
- Hammill, D. D., & Larsen, S. C. (1996). *Test of written language—Third edition*. Austin, TX: PRO-ED.
- Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*, 204–217.
- McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. MacArthur, S. Graham, & J. Fitzgerald, Eds. *Handbook of writing research*. NY: Guilford Press.
- McMaster, K. L., & Campbell, H. (2008). Technical features of new and existing measures of written expression: An examination within and across grade levels. *School Psychology Review, 37*, 550–566

McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children, 77*, 185-206.

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

Videen, J., Deno, S. L., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Vol. IRLD-RR-84). University of Minnesota, Institute for Research on Learning Disabilities.

Table 1
Descriptive Statistics for Weekly CBM Tasks

	Correct Word Sequences							Correct Minus Incorrect Word Sequences						
	Grade 2-3			Grade 4-5				Grade 2-3			Grade 4-5			
Narrative – Single Scores														
Week	Mean	(SD)	<i>n</i>	Mean	(SD)	<i>n</i>	<i>ES</i>	Mean	(SD)	<i>n</i>	Mean	(SD)	<i>n</i>	<i>ES</i>
1	33.06	(19.48)	51	60.87	(21.36)	38	1.37	6.90	(25.36)	51	44.29	(25.04)	38	1.48
2	33.55	(17.83)	51	57.21	(23.88)	38	1.15	12.25	(21.59)	51	42.18	(27.78)	38	1.23
3	36.35	(22.51)	49	62.58	(22.62)	38	1.16	15.38	(25.52)	48	42.21	(25.04)	38	1.06
4	31.33	(19.48)	51	61.18	(17.93)	38	1.58	7.78	(22.47)	51	43.50	(21.24)	38	1.63
5	37.02	(21.10)	51	69.55	(22.98)	38	1.48	14.47	(22.83)	51	53.89	(26.59)	38	1.61
6	35.94	(22.62)	49	68.03	(23.98)	38	1.38	14.71	(24.65)	49	49.92	(27.57)	38	1.36
7	36.62	(22.23)	50	69.37	(24.07)	38	1.42	14.80	(25.40)	50	52.21	(24.12)	38	1.50
8	39.58	(23.70)	48	73.84	(27.15)	37	1.36	16.67	(29.67)	48	58.95	(31.09)	37	1.40
9	37.62	(22.65)	50	71.11	(22.78)	38	1.47	13.10	(25.24)	50	51.03	(28.48)	38	1.42
10	39.33	(24.11)	51	74.65	(24.51)	34	1.46	16.98	(27.91)	51	58.82	(26.23)	34	1.54
11	42.02	(24.94)	50	68.58	(27.21)	38	1.02	15.64	(30.76)	50	47.13	(30.88)	38	1.02
12	39.82	(25.30)	51	66.95	(27.64)	38	1.03	15.18	(30.26)	51	49.18	(30.09)	38	1.13
17a	39.26	(22.01)	50	60.39	(25.91)	38	0.89	16.24	(24.19)	38	42.13	(31.21)	32	0.94
17b	36.07	(22.27)	46	61.49	(26.85)	37	1.04	10.74	(24.68)	38	40.88	(31.86)	32	1.07
Narrative – Aggregated Scores														
Weeks														
1-2	33.30	(17.37)	51	59.04	(21.02)	38	1.35	9.58	(22.07)	51	43.24	(24.25)	38	1.46
3-4	33.26	(20.39)	51	61.88	(18.87)	38	1.45	10.77	(22.73)	51	42.86	(20.91)	38	1.46
5-6	36.23	(21.17)	51	68.79	(21.97)	38	1.51	14.30	(22.44)	51	51.91	(25.11)	38	1.59
7-8	37.28	(22.28)	51	71.57	(24.31)	38	1.48	15.04	(25.94)	51	55.62	(26.25)	38	1.56
9-10	39.12	(23.56)	51	71.86	(22.79)	38	1.41	15.82	(26.36)	51	52.95	(27.45)	38	1.38
11-12	40.83	(24.18)	51	67.76	(26.39)	38	1.07	15.42	(29.19)	51	48.16	(29.24)	38	1.12
17a-17b	36.96	(21.46)	51	60.33	(26.06)	38	0.99	12.94	(21.75)	51	38.78	(28.87)	38	1.03
Expository														
Weeks														
1	28.96	(16.95)	51	59.29	(20.38)	38	1.64	11.24	(18.43)	51	44.95	(23.93)	38	1.61
6	33.02	(20.85)	49	60.08	(20.53)	38	1.31	14.18	(23.58)	49	44.03	(23.09)	38	1.28
12	36.39	(22.21)	51	65.53	(28.43)	38	1.16	16.22	(25.68)	51	52.26	(29.38)	38	1.32
18	31.70	(19.80)	51	53.99	(23.59)	38	1.04	8.45	(20.49)	51	39.51	(28.71)	38	1.28

Table 2
Descriptive Statistics for Pre- and Posttest Analytic Writing Scores

<i>Measure</i>	Grade 2-3			Grade 4-5			ES
	Mean	(SD)	<i>n</i>	Mean	(SD)	<i>n</i>	
Contextual Conventions							
Pre	3.92	(2.94)	51	6.03	(2.81)	38	0.73
Post	4.71	(2.49)	49	7.50	(3.27)	38	0.98
Contextual Language							
Pre	7.49	(3.00)	51	9.97	(3.31)	38	0.79
Post	7.47	(2.74)	49	11.24	(3.84)	38	1.15
Story Construction							
Pre	7.31	(2.72)	51	7.71	(2.85)	38	0.14
Post	7.08	(2.60)	49	8.34	(2.79)	38	0.47

Table 3
Alternate Form Reliability

	Correct Word Sequences								Correct Minus Incorrect Word Sequences							
	Grade 2-3				Grade 4-5				Grade 2-3				Grade 4-5			
	<i>r</i>	CI			<i>r</i>	CI			<i>r</i>	CI			<i>r</i>	CI		
Narrative: Single Scores																
1 with 2	0.73	0.57	to	0.84	0.73	0.54	to	0.85	0.77	0.63	to	0.86	0.69	0.48	to	0.83
2 with 3	0.76	0.61	to	0.86	0.72	0.52	to	0.85	0.7	0.56	to	0.82	0.70	0.56	to	0.83
3 with 4	0.85	0.75	to	0.91	0.73	0.54	to	0.85	0.73	0.57	to	0.84	0.63	0.39	to	0.79
4 with 5	0.84	0.74	to	0.91	0.76	0.58	to	0.87	0.81	0.69	to	0.89	0.67	0.45	to	0.81
5 with 6	0.88	0.80	to	0.93	0.75	0.57	to	0.86	0.81	0.69	to	0.89	0.72	0.52	to	0.85
6 with 7	0.89	0.81	to	0.94	0.74	0.55	to	0.86	0.82	0.70	to	0.89	0.70	0.49	to	0.83
7 with 8	0.88	0.80	to	0.93	0.83	0.70	to	0.91	0.81	0.69	to	0.89	0.84	0.71	to	0.91
8 with 9	0.90	0.83	to	0.94	0.91	0.83	to	0.95	0.84	0.74	to	0.91	0.89	0.80	to	0.94
9 with 10	0.89	0.81	to	0.94	0.88	0.78	to	0.94	0.79	0.66	to	0.87	0.87	0.76	to	0.93
10 with 11	0.90	0.83	to	0.94	0.78	0.61	to	0.88	0.84	0.74	to	0.91	0.77	0.60	to	0.87
11 with 12	0.87	0.78	to	0.93	0.85	0.73	to	0.92	0.85	0.75	to	0.91	0.84	0.71	to	0.91
17a with 17b	0.84	0.74	to	0.91	0.89	0.78	to	0.94	0.72	0.55	to	0.83	0.80	0.65	to	0.89
Narrative: Aggregated Scores																
1-2 with 3-4	0.82	0.70	to	0.89	0.83	0.70	to	0.91	0.86	0.77	to	0.92	0.81	0.66	to	0.90
3-4 with 5-6	0.89	0.81	to	0.94	0.85	0.73	to	0.92	0.91	0.85	to	0.95	0.79	0.63	to	0.89
5-6 with 7-8	0.92	0.86	to	0.95	0.89	0.80	to	0.94	0.89	0.81	to	0.94	0.87	0.76	to	0.93
7-8 with 9-10	0.95	0.91	to	0.97	0.92	0.85	to	0.96	0.92	0.86	to	0.95	0.91	0.83	to	0.95
9-10 with 11-12	0.95	0.91	to	0.97	0.83	0.70	to	0.91	0.89	0.81	to	0.94	0.82	0.68	to	0.90
Expository																
Week 18	0.85	0.75	to	0.91	0.75	0.57	to	0.86	0.66	0.47	to	0.79	0.74	0.55	to	0.86

All *ps* < .001

Table 4
Criterion-Related Validity Coefficients with Analytic Writing Scores

<i>Measure</i>	Correct Word Sequences						Correct Minus Incorrect Word Sequences					
	Grade 2-3			Grade 4-5			Grade 2-3			Grade 4-5		
	<i>r</i>	CI		<i>r</i>	CI		<i>r</i>	CI		<i>r</i>	CI	
Narrative												
Contextual Conventions												
Pre (Week 1)	.63***	0.43	to 0.77	.61***	0.36	to 0.78	.73***	0.57	to 0.84	.73***	0.54	to 0.85
Pre (Week 1-2 agg.)	.66***	0.47	to 0.79	.60***	0.35	to 0.77	.76***	0.61	to 0.86	.71***	0.51	to 0.84
Post (Week 17a-17b)	.55***	0.32	to 0.72	-.17	-0.46	to 0.16	.69***	0.52	to 0.81	.51***	0.23	to 0.71
Contextual Language												
Pre (Week 1)	.61***	0.41	to 0.76	.71***	0.51	to 0.84	.67***	0.49	to 0.80	.76***	0.58	to 0.87
Pre (Week 1-2 agg.)	.66***	0.47	to 0.79	.69***	0.48	to 0.83	.70***	0.52	to 0.82	.77***	0.60	to 0.87
Post (Week 17a-17b)	.34*	0.35	to 0.63	-.26	-0.54	to 0.07	.74***	0.59	to 0.84	.71***	0.51	to 0.84
Story Construction												
Pre (Week 1)	.29*	0.02	to 0.52	.46**	0.17	to 0.68	.24	-0.04	to 0.49	.53***	0.25	to 0.73
Pre (Week 1-2 agg.)	.40**	0.14	to 0.61	.38*	0.07	to 0.62	.28*	0.00	to 0.52	.49***	0.20	to 0.70
Post (Week 17a-17b)	.44**	0.19	to 0.64	.25	-0.08	to 0.53	.55***	0.33	to 0.72	.48**	0.19	to 0.69
Expository												
Contextual Conventions												
Pre	.57***	0.35	to 0.73	.42**	0.12	to 0.65	.68***	0.50	to 0.80	.52***	0.25	to 0.72
Post	.57***	0.35	to 0.73	.36*	0.05	to 0.61	.62***	0.41	to 0.77	.50**	0.22	to 0.71
Contextual Language												
Pre	.61***	0.41	to 0.76	.45**	0.15	to 0.68	.70***	0.52	to 0.82	.47**	0.18	to 0.69
Post	.64***	0.45	to 0.78	.62***	0.37	to 0.79	.69***	0.52	to 0.81	.73***	0.54	to 0.85
Story Construction												
Pre	.46***	0.21	to 0.65	.37*	0.06	to 0.62	.42**	0.16	to 0.62	.43**	0.13	to 0.66
Post	.38***	0.12	to 0.57	.41*	0.10	to 0.65	.48***	0.24	to 0.67	.52***	0.25	to 0.72

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

Table 5
Comparison of CBM Performance Levels by Grade Level

Week	Narrative CWS - Single			Narrative CIWS - Single		
	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
1	29.845	1, 70	<.001	33.875	1, 68	<.001
2	21.017	1, 70	<.001	28.421	1, 68	<.001
3	24.994	1, 70	<.001	21.756	1, 68	<.001
4	34.876	1, 70	<.001	39.405	1, 68	<.001
5	32.933	1, 70	<.001	40.379	1, 68	<.001
6	32.833	1, 70	<.001	30.885	1, 68	<.001
7	33.990	1, 70	<.001	39.063	1, 68	<.001
8	35.198	1, 70	<.001	39.443	1, 68	<.001
9	31.059	1, 70	<.001	32.289	1, 68	<.001
10	32.933	1, 70	<.001	37.198	1, 68	<.001
11	18.457	1, 70	<.001	15.851	1, 68	<.001
12	18.606	1, 70	<.001	20.312	1, 68	<.001
17a	18.510	1, 70	<.001	15.271	1, 68	<.001
17b	20.242	1, 70	<.001	19.869	1, 68	<.001
	Narrative CWS - Aggregated			Narr. CIWS - Aggregated		
Week	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
1-2	39.918	1, 87	<.001	46.533	1, 87	<.001
3-4	45.689	1, 87	<.001	46.413	1, 87	<.001
5-6	49.890	1, 87	<.001	55.215	1, 87	<.001
7-8	47.700	1, 87	<.001	52.750	1, 87	<.001
9-10	43.233	1, 87	<.001	41.680	1, 87	<.001
11-12	24.973	1, 87	<.001	27.347	1, 87	<.001
17a-b	21.478	1, 87	<.001	23.202	1, 87	<.001
	Expository CWS			Expository CIWS		
Week	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
1	55.496	1, 85	<.001	21.882	1, 85	<.001
6	36.538	1, 85	<.001	52.722	1, 85	<.001
12	27.541	1, 85	<.001	34.914	1, 85	<.001
18	21.882	1, 85	<.001	35.796	1, 85	<.001

Table 6
Descriptive Statistics for CBM Slopes

	Correct Word Sequences				Correct Minus Incorrect Word Sequences			
	Grade 2-3		Grade 4-5		Grade 2-3		Grade 4-5	
	Mean	SD	Mean	SD	Mean	(SD)	Mean	(SD)
Narrative Task								
Single scores								
Wks 1-3	1.12	(8.80)	0.86	(7.71)	3.59	(8.20)	-1.04	(9.87)
Wks 1-4	-0.34	(5.19)	0.63	(4.97)	0.42	(5.23)	-0.23	(6.14)
Wks 1-5	0.57	(3.98)	2.13	(3.76)	1.07	(4.23)	2.05	(4.55)
Wks 1-6	0.46	(3.14)	2.04	(3.28)	1.03	(3.42)	1.85	(3.76)
Wks 1-7	0.52	(2.31)	1.93	(2.58)	0.99	(2.67)	1.82	(2.78)
Wks 1-8	0.68	(1.86)	2.09	(2.11)	0.96	(2.25)	2.22	(2.24)
Wks 1-9	0.71	(1.70)	1.85	(1.78)	0.79	(2.15)	1.73	(2.02)
Wks 1-10	0.74	(1.48)	1.70	(1.63)	0.82	(1.96)	1.65	(1.83)
Wks 1-11	0.81	(1.36)	1.34	(1.34)	0.72	(1.79)	1.12	(1.47)
Wks 1-12	0.76	(1.26)	1.04	(1.23)	0.63	(1.72)	0.87	(1.27)
Odd Wks	0.71	(1.52)	0.91	(1.23)	0.63	(1.84)	0.46	(1.64)
Even Wks	0.84	(1.48)	1.23	(1.78)	0.63	(2.07)	1.17	(1.82)
Aggregated								
Wks 1-6	1.46	(6.48)	4.88	(6.99)	2.36	(7.06)	4.34	(7.84)
Wks 1-8	1.49	(3.81)	4.45	(4.49)	1.99	(4.59)	4.62	(4.62)
Wks 1-10	1.56	(3.09)	3.53	(3.34)	1.68	(3.99)	3.22	(3.87)
Wks 1-12	1.61	(2.57)	2.18	(2.48)	1.29	(3.45)	1.67	(2.65)
Expository Task								
Wks 1-18	1.22	(4.42)	-1.05	(6.30)	-0.56	(4.66)	-0.81	(6.91)

Table 7
Standard Errors of Estimate, Standard Errors of Slope, and t-values

<i>Data</i>	Grade 2-3						Grade 4-5					
	SEE		SEb		<i>t</i> -value*		SEE		SEb		<i>t</i> -value*	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Narrative CWS												
Single scores												
1-3	6.39	(6.22)	4.72	(4.40)	1.95	(8.96)	9.14	(9.18)	6.48	6.51	(0.44)	(3.95)
1-4	7.70	(5.18)	3.44	(2.31)	-0.43	(4.10)	10.35	(6.06)	4.62	2.70	1.06	(3.79)
1-5	8.08	(4.51)	2.56	(1.43)	0.60	(2.62)	10.87	(4.98)	3.44	1.57	0.85	(1.43)
1-6	8.33	(3.85)	1.99	(0.92)	0.51	(2.23)	10.99	(4.08)	2.63	0.98	0.89	(1.31)
1-7	8.34	(3.52)	1.58	(0.67)	0.61	(2.20)	11.02	(3.75)	2.08	0.71	0.99	(1.23)
1-8	8.44	(3.27)	1.30	(0.50)	0.73	(2.01)	11.01	(3.61)	1.70	0.56	1.31	(1.33)
1-9	8.38	(2.98)	1.08	(0.38)	0.78	(1.90)	10.70	(3.19)	1.38	0.41	1.38	(1.31)
1-10	8.39	(2.67)	0.92	(0.29)	0.94	(1.87)	10.64	(3.04)	1.17	0.34	1.51	(1.49)
1-11	8.27	(2.51)	0.79	(0.24)	1.03	(1.81)	11.16	(2.85)	1.06	0.27	1.21	(1.14)
1-12	8.38	(2.54)	0.70	(0.21)	1.12	(1.74)	11.43	(3.01)	0.96	0.25	1.05	(1.25)
Aggregated												
1-6	5.95	(3.96)	4.22	(2.81)	1.72	(6.30)	6.09	(4.77)	4.32	(3.38)	3.82	(7.63)
1-8	6.33	(3.08)	2.83	(1.38)	0.90	(1.84)	7.05	(3.68)	3.15	(1.64)	1.87	(2.54)
1-10	6.17	(2.62)	1.95	(0.83)	1.15	(2.08)	7.25	(3.07)	2.30	(0.97)	1.58	(1.71)
1-12	6.10	(2.25)	1.46	(0.54)	1.28	(1.85)	8.61	(3.52)	2.06	(0.84)	1.08	(1.22)
Narrative CIWS												
Single scores												
1-3	8.89	(8.52)	6.70	(6.02)	1.66	(4.27)	11.46	(9.65)	8.13	(6.84)	-2.48	(10.12)
1-4	10.09	(7.04)	4.51	(3.14)	0.39	(3.02)	13.08	(6.65)	5.84	(2.97)	0.38	(3.31)
1-5	9.81	(5.68)	3.10	(1.80)	0.54	(2.42)	13.78	(6.06)	4.36	(1.92)	0.93	(2.19)
1-6	10.06	(4.78)	2.41	(1.14)	0.65	(2.20)	14.03	(4.76)	3.36	(1.14)	0.67	(1.50)
1-7	10.16	(4.09)	2.02	(0.78)	0.57	(1.55)	13.83	(3.92)	2.61	(0.65)	0.72	(1.02)
1-8	10.69	(4.12)	1.65	(0.64)	0.68	(1.76)	13.80	(3.43)	2.13	(0.53)	1.05	(1.03)
1-9	10.62	(3.90)	1.37	(0.50)	0.60	(1.81)	13.55	(3.06)	1.75	(0.39)	0.98	(1.08)
1-10	10.76	(3.62)	1.19	(0.40)	0.74	(1.94)	13.27	(2.98)	1.46	(0.33)	1.12	(1.17)
1-11	10.77	(3.49)	1.03	(0.33)	0.69	(2.16)	13.83	(3.07)	1.32	(0.29)	0.78	(1.00)
1-12	10.97	(3.48)	0.92	(0.29)	0.65	(2.27)	13.95	(3.24)	1.17	(0.27)	0.68	(1.06)

Table 7
Continued
Standard Errors of Estimate, Standard Errors of Slope, and t-values

<i>Data</i>	Grade 2-3						Grade 4-5					
	SEE		SEb		<i>t</i> -value*		SEE		SEb		<i>t</i> -value*	
	Mean	(<i>SD</i>)	Mean	(<i>SD</i>)	Mean	(<i>SD</i>)	Mean	(<i>SD</i>)	Mean	(<i>SD</i>)	Mean	(<i>SD</i>)
Aggregated												
1-6	5.64	(3.79)	4.00	(2.69)	1.32	(3.87)	8.89	(6.30)	6.31	(4.47)	-0.12	(12.11)
1-8	6.87	(3.80)	3.07	(1.70)	0.89	(2.16)	9.18	(4.26)	4.10	(1.90)	1.66	(2.47)
1-10	7.02	(3.12)	2.22	(0.99)	0.99	(2.29)	9.35	(3.40)	2.96	(1.08)	1.51	(2.50)
1-12	7.52	(2.86)	1.80	(0.68)	0.75	(2.17)	10.64	(3.82)	2.55	(0.91)	0.66	(1.13)
Expository CWS	7.48	(5.24)	3.34	(2.34)	0.45	(2.21)	12.49	(7.84)	5.58	(3.50)	-0.20	(1.61)
Expository CIWS	9.39	(6.67)	4.19	(2.98)	0.53	(4.33)	13.10	(9.46)	5.85	(4.22)	-0.17	(2.64)

*No *t*-values were significant at the $p < .05$ level.

Table 8
Repeated-Measures Analyses of Pre- to Posttest Analytic Writing Scores

Main effect	<i>F</i>	<i>df</i>	<i>p</i>
Time			
Contextual Conventions	17.94	1, 85	< .001
Contextual Language	5.13	1, 85	.03
Story Construction	0.48	1, 85	.49
Grade Level			
Contextual Conventions	22.44	1, 85	<.001
Contextual Language	26.08	1, 85	<.001
Story Construction	2.79	1, 85	.10

Figure 1. Reliability of incremental slopes using single scores for narrative prompts, Grades 2-3.

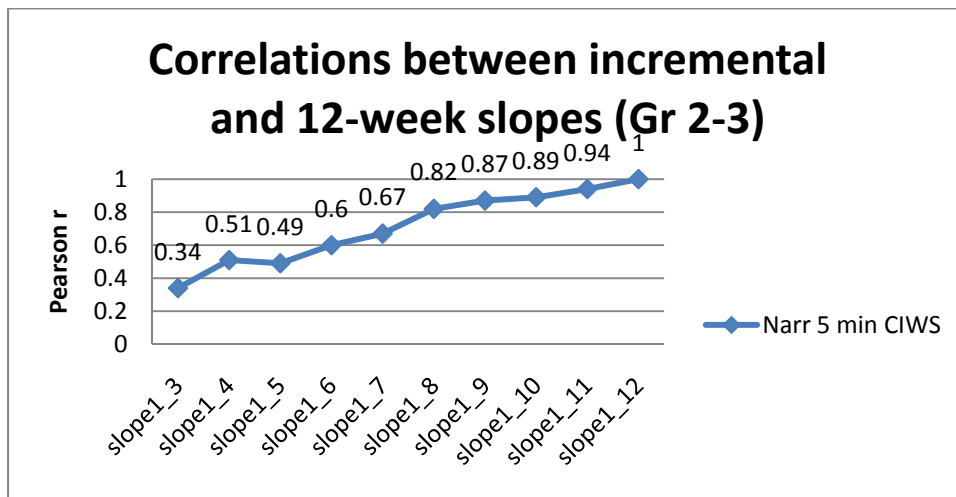
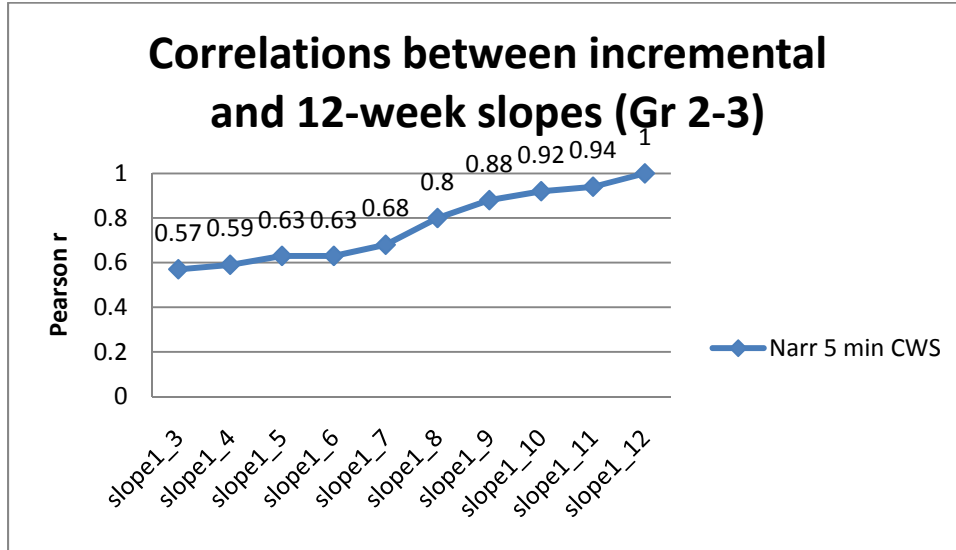


Figure 2. Reliability of incremental slopes using aggregated scores for narrative prompts, Grades 2-3

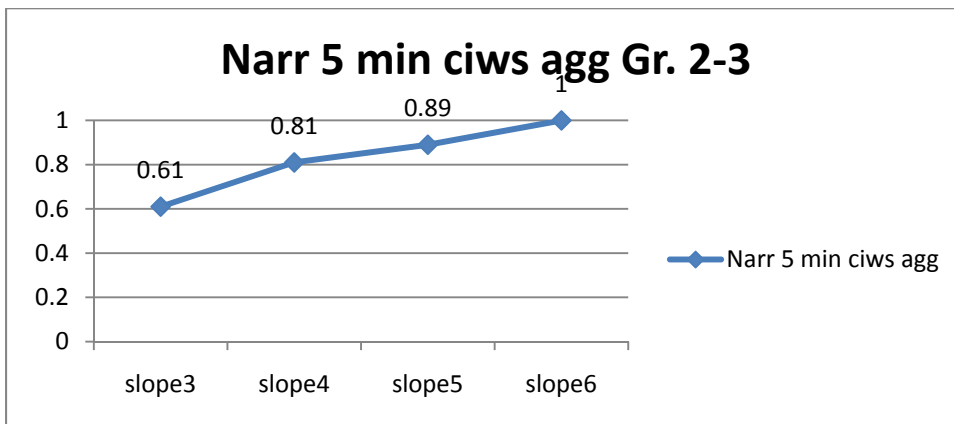
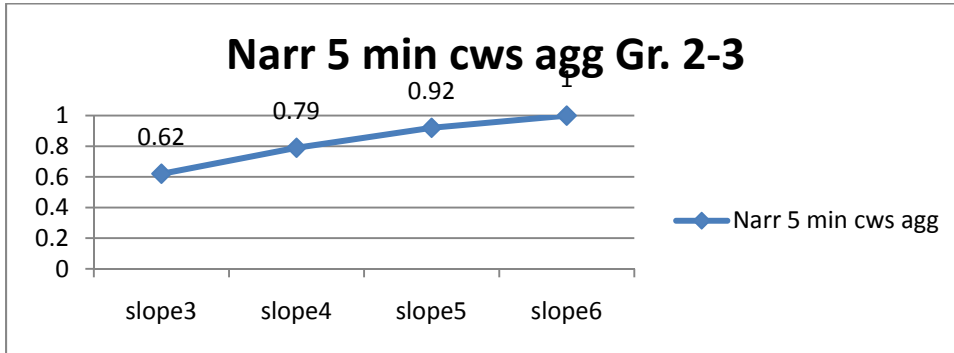


Figure 3. Reliability of incremental slopes using single scores for narrative prompts, Grades 4-5.

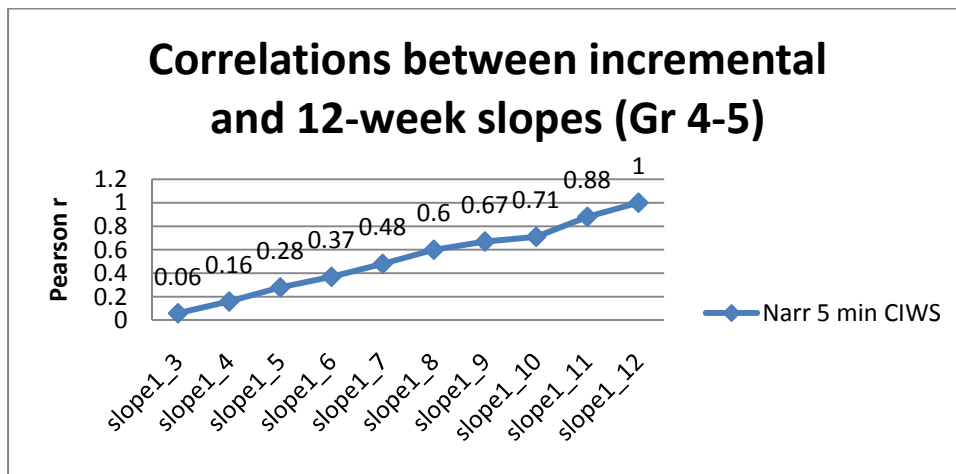
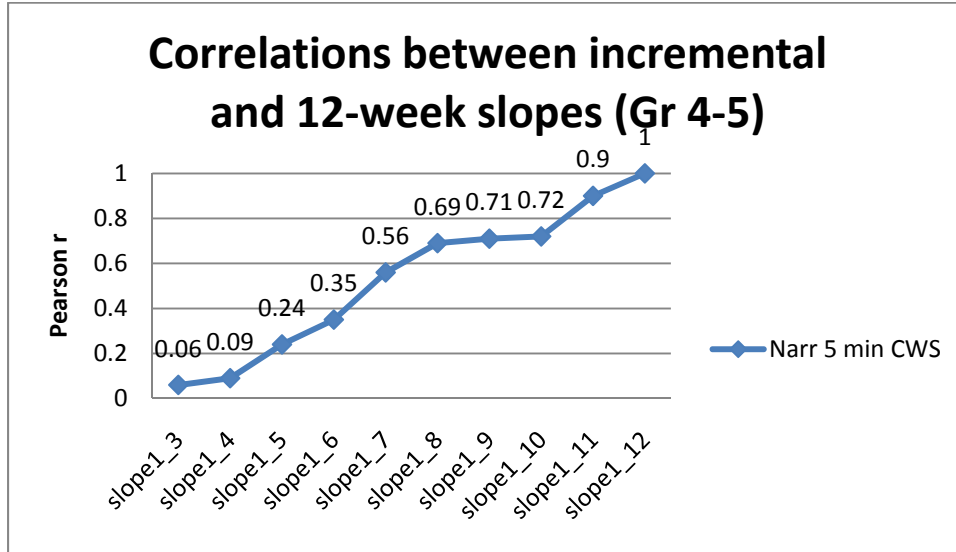
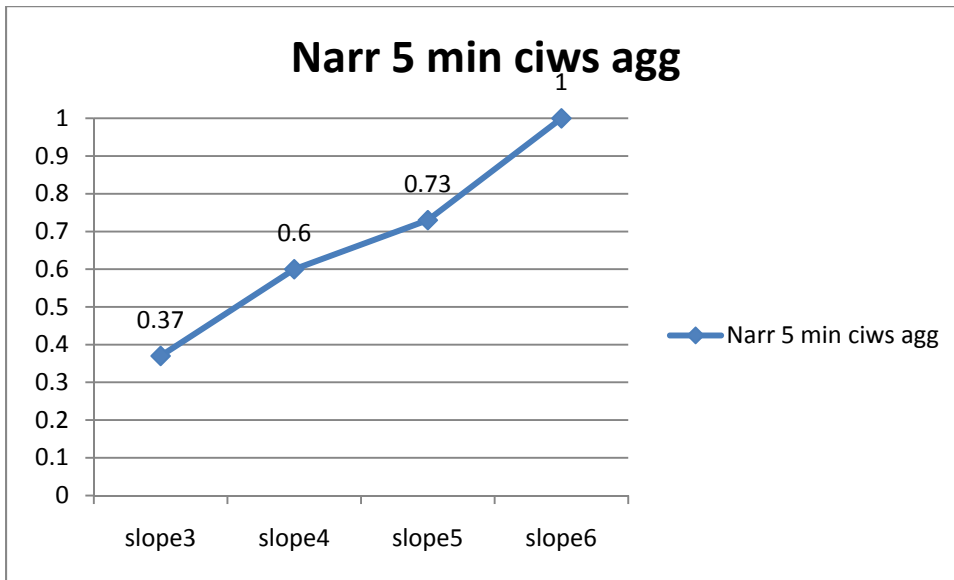
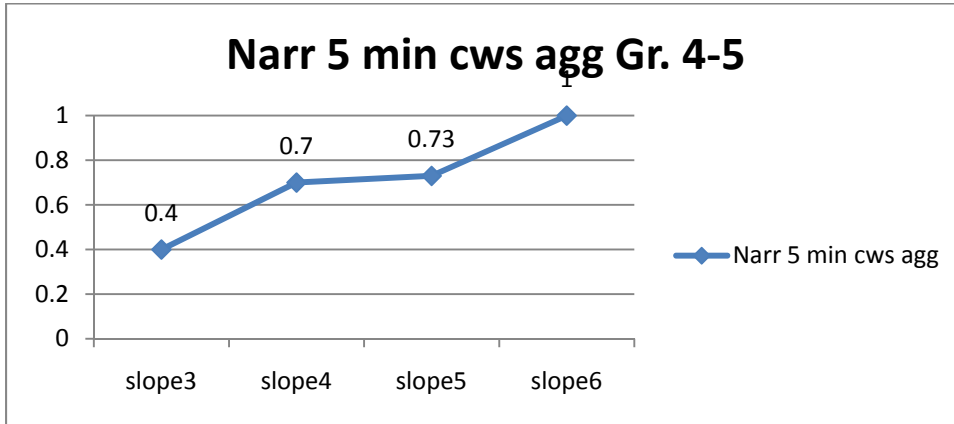


Figure 4. Reliability of incremental slopes using aggregated scores for narrative prompts, Grades 4-5.



APPENDIX A

Writing Prompts and order of administration.

Week	Narrative Prompts	Expository Prompts
1	One night I had a strange dream about...	Describe a game you like to play and tell why you like it.
2	It was the last day of school so I decided to...	
3	One day, when I got home from school...	Describe your favorite day of the week and tell why you like it.
4	One day my friend told me the strangest story...	
5	I was walking home when I found a \$100 bill on the sidewalk and...	
6	One morning I found a note under my pillow that said . . .	
7	One day I went to school but nobody was there except me, so I...	Describe your favorite time of the year and tell why you like it.
8	It was a dark and stormy night...	
9	I was on my way home from school and...	
10	It was the first day of school and...	Describe your favorite thing to do and tell why you like it.
11	I was watching TV when I heard a knock at the door and...	
12	I was talking to my friends when, all of a sudden...	Describe a place you like to go and tell why you like to go there.
15	One day I woke up and was invisible and...	
15	One day I found the most interesting thing and...	Describe the clothes that kids wear in your school and tell why kids wear them.
18	One summer I went on a trip and...	
18	I was walking down the street when I saw...	Describe the friends you have and tell why they are your friends.