Scoring Measures of Word Dictation Curriculum-Based Measurement in Writing: Effects of Incremental Administration

Apryl L. Poch

Duquesne University

Abigail A. Allen

Clemson University

Erica S. Lembke

University of Missouri

Published in *Psychology in the Schools*
Received: 14 June 2018
Accepted: 13 November 2018
Published: 04 April 2019

Author Note

Please direct all correspondence regarding the manuscript to Apryl L. Poch, Ph.D., Assistant Professor of Special Education, Duquesne University, Department of Counseling, Psychology, and Special Education, 600 Forbes Avenue, G9C Canevin Hall, Pittsburgh, PA 15282, PochA1@duq.edu , 412.396.4478 (phone), or 412.396.1340 (fax).

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A130144 to the University of Minnesota. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

Spelling has been identified as a key transcription skill that emerges during the elementary years as students learn how to write and subsequently develop fluency with writing (McCutchen, 1996), making the assessment of spelling a critical component of evaluation systems within schools. This includes the use of curriculum-based measures of writing (CBM-W). This study examined the extent to which word dictation CBM-W administered during the Fall, Winter, and Spring of an academic year maintained technical adequacy across 1-minute time intervals in grades 1–3. Results revealed moderate predictive and concurrent validity estimates with the Spelling subtest of the Weschler Individual Achievement Test – III. Statistically significant differences existed between and within grade levels across each minute of administration and across Fall, Winter, and Spring time points for all scoring procedures.

Keywords: curriculum-based measurement, spelling, writing

Scoring Measures of Word Dictation Curriculum-Based Measurement in Writing: Effects of Incremental Administration

Measures of students' writing progress that are technically adequate are a necessary component of evaluation systems within schools in order to ensure students attain standards of writing proficiency (McMaster & Espin, 2007). These writing measures are equally important to identify students who are at risk or identified with writing disabilities and for informing instruction and intervention. For early elementary writers, measures related to spelling ability have been suggested to be predictive of future writing proficiency (Berninger, et al., 2002). Indeed, spelling is a key transcription level skill that emerges during the elementary years as students learn how to write and subsequently develop fluency with writing (McCutchen, 1996), and as students begin to untangle phoneme-grapheme correspondences and master the alphabetic principle to decode and encode words (Weiser & Mathes, 2011).

The awareness of spelling as a critical element of writing is consistent with early theoretical models of writing, including the Simple View of Writing. The earliest representation of the Simple View of Writing, developed by Juel, Griffith, and Gough (1986), posited that writing was composed of a lower order skill (i.e., spelling) and a higher order skill (i.e., ideation). They found that spelling and ideation accounted for approximately 30% of the variance in writing quality in first and second grade after controlling for IQ and oral language ability. Later work by Berninger and colleagues (Berninger & Amtmann, 2003; Berninger et al., 2002), also advanced a Simple View of Writing model which included transcription level skills (e.g., spelling and handwriting), self-regulatory executive functions, and text generation, all situated within a working memory environment that accounted for the influence of working, short-term, and long-term memory required during the writing process. In 2006, Berninger and

Winn made slight modifications to the model, better detailing the self-regulatory executive functions and the relationship between working, short-term, and long-term memory. This new iteration became known as the Not So Simple View of Writing.

Spelling

Several studies have illustrated the influence of spelling ability on writing performance in the early grades. Graham and colleagues found that 66% of the variance in writing quality in first grade was accounted for by spelling and handwriting ability (Graham, Berninger, Abbott, Abbott, & Whitaker, 1997). Kim et al. (2011) found that spelling, along with oral language and letter writing fluency, uniquely predicted 33% of the variance in Kindergarten writing ability when controlling for reading ability. Abbott, Berninger, and Fayol (2010) discovered that spelling ability uniquely predicted text-level composition from first through seventh grade. Berninger et al. (2002) found that teaching spelling and composition in combination in third grade increased students' skills in each area and that transcription skills like spelling uniquely predicted writing fluency in the elementary grades. Most recently, Kim and Schatschneider (2017) revealed that spelling was one of three variables that fully mediated the relation of higher order cognitive skills like working memory to writing. These studies reinforce the importance of spelling in the writing process and may help support why spelling has maintained a prominent place in models such as the Simple View of Writing and the Not So Simple View of Writing. Indeed, spelling is critical for achieving fluency in writing, especially during the elementary grades, as difficulty with spelling can interfere with the writing process (e.g., planning and composing) and inhibit working memory (Graham, 1999). Furthermore, understanding students' spelling ability and patterns of spelling necessitates systems for evaluating and assessing students' spelling skills.

Spelling has been assessed in the research in several ways. One method is formal, standardized, norm-referenced tests such as the Test of Early Written Spelling – 4 (TWS-4; Larsen, Hammill, & Moats, 1999). While standardized tests are useful for making eligibility decisions, they are not always instructionally useful (Calfee & Miller, 2013). Informal teacher—or researcher—created criterion referenced assessments of spelling have been used to assess specific sets of words, but do not give an overall picture of general spelling ability (Hampton & Lembke, 2016). Furthermore, beginning spellers frequently spell words incorrectly, and standardized and criterion referenced spelling assessments that score test items as wholly correct or incorrect lead to many students receiving low total scores, which does not indicate specifically where a student is struggling in spelling (Clemens, Oslund, Simmons, & Simmons, 2014).

In response to some of the issues with traditional spelling assessments and scoring methods, researchers have attempted to validate alternate or supplemental scoring procedures. Ritchey, Coker, and McCraw (2010) found that on the TWS-4 (Larsen et al., 1999), scores obtained by counting sounds represented within words, letter pairs, and using a rubric to capture invented spellings, were highly correlated with each other and with measures of phonological awareness, letter naming, and writing in Kindergarten. Using a researcher-created criterion referenced spelling task, Masterson and Apel (2010) devised a new scoring system, the Spelling Sensitivity Score (SSS), and found that scoring phonological elements in each word as well as linguistic aspects (e.g., affixes) allowed them to better capture spelling development in Kindergarten and first grade. Clemens et al. (2014) used the same scoring methods as Ritchey et al. (2010) and the SSS (Masterson & Apel, 2010) to score the TWS-4 (Larsen et al., 1999) and found that all scoring methods were highly correlated to measures of early reading (e.g., phonological awareness, word reading). While these studies provide evidence of validity of various methods of spelling assessments and scoring, there is still a need to investigate how to appropriately capture student growth in spelling over time. Norm-referenced standardized assessments and criterion referenced assessments are not always technically adequate for progress monitoring (Hampton & Lembke, 2016). To address these issues, studies on Curriculum-Based Measures of Writing (CBM-W) seek to develop and refine global outcome measures for screening and progress monitoring (Deno, 2003); this includes CBM-W targeted at the word level to evaluate early writing skills specific to spelling.

CBM-W Tasks

Curriculum-based measurement (CBM) has been identified in the research as a valid and reliable means of tracking student progress (Deno, 1985). Originally developed in the mid-1970s by Dr. Stan Deno and colleagues at the University of Minnesota's Institutes for Research on Learning Disabilities, CBM is a process where students complete multiple forms of the same measure over a series of time. These forms, of equivalent difficulty, are scored and then graphed. CBM is intended to be simple, inexpensive, unobtrusive, and a quick check of student performance (Marston, 1989). CBM has focused on critical skills in academic areas such as reading, writing, mathematics, and science, and has been demonstrated to be highly predictive of students' educational outcomes (Deno, 2003). In the area of writing, the most common measure has been the story prompt CBM-W which requires a student to prepare a story in response to a sentence-starter. Students write for 3–5 minutes and are evaluated on the number of words written (WW), words spelled correctly (WSC), and correct word sequences (CWS; Videen, Deno, & Marston, 1982). See McMaster and Espin (2007) for a review. However, many young writers as well as struggling writers and writers with disabilities still struggle with lower-order writing (i.e., word writing, sentence construction) and cognitive (i.e., memory storage and retrieval, including long-term, short-term, and working memory) skills. For example, many writers in Kindergarten through grade 3 are still learning how to write, and are still developing phonological and orthographic knowledge, and phonemic awareness. Thus, researchers have created CBM for evaluating early writing fluency and production; this has included measures of word spelling, dictation, copying, and sentence writing (e.g., Hampton & Lembke, 2016; Lembke, Deno, & Hall, 2003; McMaster, Du, & Petursdottir, 2009; McMaster et al., 2011; Parker, McMaster, Medhanie, & Silberglitt, 2011). We briefly discuss the technical adequacy of many of these measures below.

Studies have indicated that tasks involving letter–, word–, and sentence-level copying, dictation, and novel writing, where participants generate their own sentences instead of copying or taking dictation, are more reliable and valid for Kindergarten through third grade as compared to the late elementary grades (Coker & Ritchey, 2013; Lembke, Deno & Hall 2003; McMaster & Campbell, 2008; McMaster, Du & Petursdottir, 2009; Ritchey, 2006; Ritchey & Coker, 2013, 2014). Ritchey (2006) found that letter writing, sound spelling, and word spelling were reliable (r = .89–.92) and valid (r = .27–.81) tasks to measure writing skills in Kindergarten at a single point in time. Later studies revealed that word spelling as measured by a word dictation (WD) task, was accurate at identifying students at risk for literacy problems in first grade (AUC = .780–.873; Ritchey & Coker, 2014). Lembke et al. (2003) found that WD had the strongest criterion validity in second grade compared to other copying and dictation tasks (r = .80–.92). However, these tasks must be given for enough time to yield accurate results while also being feasible for educators to administer in the classroom.

CBM-W Length of Administration

Most of the research on length of CBM-W administration has focused on sentence– and passage-level writing tasks. Deno, Mirkin, and Marston (1980b) administered various story prompts for 1–5 minutes in third through sixth grade. They found that the validity coefficients for story prompt and the Test of Written Language (TOWL; Hammill & Larsen, 1978), Stanford Achievement Test (Madden, Gardner, Rudman, Karlsen, & Merwin, 1978), and the Developmental Sentence Scoring System (Lee & Canter, 1971) appeared to increase across the 1, 2, and 3 minute administration levels and that the 3–minute time yielded scores with stronger validity coefficients (r = .65) than the shorter times (r = .60); however, beyond three minutes, the validity coefficients were similar across scores. Reliability was not assessed in this study. McMaster and Campbell (2008) extended this work by investigating whether 3, 5, and 7 minute administrations of various sentence– and passage-level CBM-W tasks in third, fifth, and seventh grades produced different results. They found that, in general, the administration time required to obtain reliable and valid scores increased with grade level, and for the elementary grades, a 3–5 minute administration time yielded the most reliable (r = .74–.93) and valid (r = .60–.70) results. McMaster et al. (2009) replicated these results in first grade and found that 3–5 minutes for sentence copying and sentence writing yielded the most technically sound results (reliability: r > .70; validity: r = .51–.70).

At the word–level, only a few studies have investigated student performance across varying administration times of WD tasks. Past research has found that in general, WD tasks administered for three minutes have demonstrated the strongest evidence of technical adequacy for assessing progress of struggling learners in the early grades (Deno, Mirkin, Lowry, & Kuehnle, 1980a; Deno et al., 1982; Hampton & Lembke, 2016; Lembke, et al., 2003), although some research has demonstrated strong validity coefficients at one and two minutes (Deno et al., 1980a; Deno et al., 1982).

Even though a 3 minute administration has demonstrated technical adequacy in the literature, it is important to question whether a shorter administration time might also yield evidence of technical adequacy and if there are differences in student performance between administration lengths and across scoring measures. In their study of spelling measures in grades 2–6, Deno et al. (1980a) found stronger criterion validity coefficients with WSC, correct letter sequences (CLS), and the Test of Written Spelling (Larsen & Hammill, 1976) at 3 minutes compared to 1 and 2 minutes of WD administration, although nearly all coefficients were above $r = .80$. Deno et al. (1980a) also found significant differences between grade levels for the WSC and CLS scoring methods at 3 minutes, although differences in time of administration were not differentiated by grade level. In a subsequent study, Deno et al. (1982) similarly found significant growth from Fall to Spring on WSC and CLS scoring methods with a 3 minute WD probe for students in first through sixth grade.

Consequently, with the available literature base being both scant and dated, it is currently unclear whether the administration time required to obtain reliable and valid scores remains the same, increases, or decreases by grade level. More research is needed to determine the timing interval with the most technical adequacy for administering WD CBM-W probes, and whether shorter assessment duration with WD probes can still produce reliable results in the early elementary grades (Espin, et al., 2008) and distinguish statistically significant differences between and within grade levels.

Purpose

Educators need technically adequate assessments for evaluating student writing, including those components of writing (e.g., spelling) that are known to be predictive of future writing proficiency (e.g., Berninger et al., 2002). When those measures are easy to administer and minimize the time taken away from instruction, educators are more likely to use such measures for understanding and tracking students' progress. Given that CBM-W can be quickly administered and scored, it is important that educators and researchers identify the amount of time required to get the best indication of a student's growth. This study examined technical adequacy for WD CBM-W across 1 minute time intervals, from 1 to 3 minutes, in grades 1–3. This study sought to answer the following two research questions: (a) To what extent does WD CBM-W maintain technical adequacy across 1 minute time intervals?, and (b) To what extent do statistically significant differences exist between and within grades across the various scoring procedures across 1 minute time intervals?

Methods

Participants and Setting

Participants were drawn from a larger, multi-site CBM-W benchmarking study conducted in the Midwest (Carlisle, Poch, & Lembke, 2015; Allen, Jung, et al., 2018). Participants included students in grades 1 (*n* = 96), 2 (*n* = 118), and 3 (*n* = 124) from two elementary schools within one Midwestern school district in a mid-sized city. The district served 17,905 students preschool through twelfth grade during the 2013-2014 academic year. Across the district, students were 61.6% White, 20.4% Black, 6.1% Hispanic, 5.3% Asian, and 5.9% Multi-racial. Approximately 39.6% of students were eligible for free/reduced price lunches district-wide, and 10.6% of students received special education services in the district.

Only participants from the larger study who had completed the Spelling and Sentence Composition sub-tests of the Weschler Individual Achievement Test – III (WIAT-III), the standardized criterion measure, were included in this study (*n* = 150; 50 students each at first, second, and third grades). Demographics of students in this sample (i.e., *n* = 150, 49% female, 62% White, 54% free/reduced price lunch, 0% English language learners, 5% receiving special education services) were comparable to the larger study population (i.e., *n* = 338, 51% female, 64% White, 54% free/reduced price lunch, 2% English language learners, 9% receiving special education services).

Measures

WD CBM-W. WD CBM-W is a measure designed to capture students' transcription skills at the word level. WD requires students to write words dictated twice by the examiner. Words are presented singularly; they are not used within the context of a sentence. WD probes were developed and initially researched by members of the larger research team. Words (*n* = 40) used in these probes were selected from high-frequency word lists and were designed to address students' knowledge of various spelling patterns (e.g. VC, CVC, VCe) appropriate for elementary writers. Four alternate WD forms were created and utilized in the larger screening study, using standardized administration directions.

CBM-W scoring. On the WD CBM-W measures, four standardized scoring methods were used. WD measures were scored for WW, WSC, CLS, and Correct Minus Incorrect Letter Sequences (C-ILS). An explanation of each scoring procedure follows.

Words written (WW). The total number of words written; a "word" was defined as a sequence of letters separated by a space from another sequence of letters (definition consistent with Deno et al., 1980a, 1980b; 1982; Hampton & Lembke, 2016; Lembke et al., 2003; Marston & Deno, 1981; and Parker, Tindal, & Hasbrouck, 1991).

Words spelled correctly (WSC). The number of correctly spelled words; a word spelled correctly had to match the form of the word dictated by the examiner, with the exception of homophones (definition consistent with Deno et al., 1980a, 1980b; 1982; Hampton & Lembke, 2016; Lembke et al., 2003; Marston & Deno, 1981; and Parker, Tindal, & Hasbrouck, 1991).

Correct letter sequences (CLS). Any two adjacent letters within a dictated word that are correctly placed when spelled (definition consistent with Deno et al., 1980a, 1980b; 1982; Hampton & Lembke, 2016; Lembke et al., 2003; and Marston & Deno, 1981). CLS are recorded if the first letter appropriately matches the initial sound of the dictated word, between all adjacent letters, and for correctly denoting the end sound of the dictated word. Therefore, each word has one more letter sequence than there are letters in the word. Take for example the word “mile” (five possible CLS). Should the student spell the word “myle,” letter sequences around the incorrect letter *y* would be incorrect; this student would have scored two incorrect letter sequences and three CLS.

Correct minus incorrect letter sequences (C-ILS). The number of CLS minus the number of incorrect letter sequences (Marston, 1989).

These scoring methods, especially WW, WSC, and CLS, have demonstrated strong correlation coefficients with standardized achievement measures, along with strong internal reliability and inter-rater reliability (Deno et al., 1980a, b; 1982; Marston & Deno, 1981).

WIAT-III. The WIAT-III is a standardized measure of students’ academic performance in grades Pre-K through 12. Average reliabilities across the subscales range from .83 to .97 (Pearson, 2009). Within the larger study, the Spelling and Sentence Composition subtests (which includes Sentence Combining and Sentence Building) were administered individually to participants in May of the academic year. For the purposes of this study, though, only the Spelling subtest was used. The Spelling subtest requires students to write the target letter sound or word that is presented by the examiner; letter sounds are presented within the context of a word, and words are presented within the context of a sentence (Pearson, 2009). All standardized administration and scoring procedures for the subtests were followed.

Procedures

Students completed two forms of WD CBM-W measures at each administration. In the larger screening study, six sets of CBM packets, each containing two alternate forms of WD, were counterbalanced across classrooms, stratified across grades, administered individually by trained members of the research team, and timed for three minutes at Fall (November/December), Winter (February), and Spring (April). Administrators marked the scoring copy at 1 and 2 minute intervals during administration. If a student paused on a word for more than five seconds, the administrator said to the student, “Let’s go on to the next word.” However, if a student had begun writing the word, he/she could take as much time as needed to finish spelling the word.

All WD measures were previously scored for WW, WSC, CLS, and C-ILS by trained members of the research team (i.e., professors, project coordinators, and advanced doctoral students in special education). Inter-rater reliability was a minimum of 85%. However, this scoring was initially only completed for the full 3 minute measure. The first and second author then scored students’ samples for 1 and 2 minutes, and rechecked the 3 minute scores. A first year doctoral student, who was trained by the first and second authors, also assisted in some of the minute level scoring. Twenty percent of the probes were re-scored for inter-rater reliability and all raters were 100% reliable with each other. The first and second authors then double-entered the data into Microsoft Excel prior to analysis and any inconsistencies were corrected.

Trained graduate students (not part of the original research team) and one of the project coordinators from the larger study administered and scored all WIAT-III assessments. Inter-rater reliability for scoring on the Spelling subtest ranged from 94–100%.

Data Analysis

Descriptive statistics (i.e., mean, *SD*, correlation) were calculated for all measures. Predictive and concurrent criterion validity coefficients with the WIAT-III Spelling subtest were calculated using Pearson correlations. To detect differences between grade levels across administration times, a between groups one-way analysis of variance (ANOVA) was conducted. LSD post-hoc testing was used to pinpoint between which grade levels significant differences existed. To detect differences within grade levels across administration times, a repeated measures ANOVA (RM-ANOVA) was conducted with simple and repeated contrasts to identify between which minutes and which time points (Fall, Winter, Spring) differences existed within grades. All data were scored and entered into SPSS (v. 22.0) for analysis.

Data analysis was conducted on the entire sample and then repeated with the 5% of participants receiving special education services removed from the analysis. Results were unchanged by removing the participants receiving special education services; therefore, the results of the full sample analysis are reported here.

Results

Descriptive Statistics

Across all grade levels and time periods, students’ scores gradually increased by minute for WW, WSC, CLS, and C-ILS. On the WIAT-III, first grade scores for Spelling were at the average score of 100, which falls within the average range of achievement. Scores for second and third grade also fell within the average range but were slightly lower (at 94) than the scores at first grade. Means and standard deviations of writing scores for early writers in grades 1–3 on the WD CBM measure (Form A) and the WIAT-III Spelling subtest can be seen in Table 1. Where available, standardized scores were reported, and data was disaggregated by grade, minute, and scoring method. On the WIAT-III Spelling subtest, standardized scores were converted from raw scores and are based on age norms. Examination of skewness, kurtosis, histograms, and P-P plots confirmed that the distribution of the data was approximately normal.

Pearson product moment correlations across time periods were also calculated and disaggregated by grade and scoring method. Due to the size of these tables, they were not included in this manuscript, but they are available from the first author upon request. In general, correlation coefficients across grade, time, and scoring method were moderate to strong and statistically significant ($p \leq .01$).

To provide a more detailed overview, correlation coefficients were examined to determine whether matched scoring methods at different time periods and grade levels were related (e.g., WW at Fall and Winter, Fall and Spring, and Winter and Spring) (see Table 2). At first grade, coefficients across matched scoring methods were moderately to strongly statistically correlated (Fall–Winter: r = .62–.89; Fall–Spring: r = .65–.86; Winter–Spring: r = .64–.89; $p \leq .01$), with most matched scoring methods having the strongest coefficient at three minutes.

At second grade, moderate to strong statistically significant coefficients ($p \leq .01$) were also found when examining matched scoring methods (Fall–Winter: r = .78–.95; Fall–Spring: r = .72–.92; Winter–Spring: r = .74–.94). Again, most matched scoring methods had the strongest coefficient at three minutes.

At third grade, all coefficients for matched scoring methods were moderate to strong and statistically significant ($p \leq .01$) (Fall–Winter: r = .67–.86; Fall–Spring: r = .58–.83; Winter–Spring: r = .77–.89). From Fall to Winter, matched scoring methods for WW, WSC, and CLS had the strongest coefficients at 2 minutes. For C-ILS, coefficients were equivalent at 3 minutes. However, coefficients of matched scoring methods at 3 minutes were strongest from Fall to Spring and Winter to Spring.

Predictive Validity

In past research, the generally accepted level of adequate validity for CBM-W has been $r \geq .50$ (McMaster & Campbell, 2008; McMaster et al., 2009) in order to identify promising measures and scoring methods and to account for historically modest validity coefficients of writing measures (Taylor, 2003). Coefficients meeting this criterion have been bolded in the results tables (see Table 3). Predictive validity of Fall WD scores with the WIAT-III Spelling subtest ranged from r = .09–.53 for first grade, .44–.74 for second grade, and .49–.76 for third grade (see Table 3). In first grade, WSC met the $r \geq .50$ criterion at one minute, and C-ILS met criterion at three minutes for the Spelling subtest. In second and third grades, validity coefficients for WSC, CLS, and C-ILS at 1–, 2–, and 3–minutes met the $r \geq .50$ criterion with the WIAT-III Spelling subtest. In general, validity coefficients increased slightly with each minute of time administration across grade levels. WW at first grade was the only scoring method that demonstrated exceptionally weak validity across time of administration.

Concurrent Validity

Again, past research has generally demonstrated that the accepted level of adequate validity for CBM-W has been $r \geq .50$ (McMaster & Campbell, 2008; McMaster et al., 2009). Coefficients meeting this criterion have been bolded in the results tables (see Table 3).

Concurrent validity of Spring WD scores with the WIAT-III Spelling subtest ranged from r = .11–.49 for first grade, .43–.77 for second grade, and .51–.66 for third grade. With the Spelling subtest in first grade, no scoring methods met criterion for any length of administration, although WSC and C-ILS came close (.45–.49). In second and third grade, all scoring procedures met the .50 criterion level for nearly every minute of administration. Overall, in second grade, validity coefficients increased with length of administration; this did not hold true in first or third grade. Moreover, concurrent validity correlations were weaker than predictive validity correlations.

Differences Between Grade Levels

To determine whether statistically significant differences existed between grades across minute intervals in Fall, Winter, and Spring on the various WD scoring procedures, a between-groups one-way ANOVA was run with an LSD post-hoc test to specify between which grade levels significant differences existed, if any. Results revealed statistically significant differences between grades at each minute interval in Fall, Winter, and Spring for all scoring procedures (see Table 4). Post-hoc comparisons revealed there were significant differences between first and second, second and third, and first and third grade at 1, 2, and 3 minutes ($p \leq .01$) for all scoring methods. This was true at the Fall, Winter, and Spring time points.

Differences Within Grade Levels

A repeated measures ANOVA (RM-ANOVA) was used to detect growth across time points at each minute of administration (e.g., was there growth at 1 minute in Fall to 1 minute in Winter to 1 minute in Spring) and across minutes at each time point (e.g., was there growth in Fall from 1 to 2 to 3 minutes, in Winter from 1 to 2 to 3 minutes, and in Spring from 1 to 2 to 3 minutes) within each grade level. For differences across time points at each minute of administration (1 minute Fall, 1 minute Winter, 1 minute Spring, etc.), Mauchly’s test indicated that the assumption of sphericity was violated for all scoring procedures at the 3 minute administration ($p \leq .04$); therefore, the degrees of freedom were corrected using the Greenhouse-Geiser estimates of sphericity (€ = .88). Results indicated a significant main effect for time for all scoring procedures for 1, 2, and 3 minutes at Fall, Winter, and Spring. This suggests that participants grew significantly from Fall to Winter, Winter to Spring, and Fall to Spring within each minute of administration (e.g., 1 minute Fall to 1 minute Winter to 1 minute Spring) using all scoring procedures. A significant time by grade interaction effect was found for CLS ($F(4, 244) = 2.81$, $p = .03$) and C-ILS ($F(2, 244) = 2.98$, $p = .02$) at the 2 minute administration (see Table 5). A series of simple and repeated contrasts indicated that for CLS there was no significant difference for the time by grade interaction effect from Fall to Spring but there was a significant difference from Fall to Winter ($F(2, 122) = 3.28$, $p = .04$) and Winter to Spring ($F(2, 122) = 3.37$, $p = .04$). For C-ILS, the only significant difference for the time by grade interaction effect was between Winter and Spring ($F(2, 122) = 3.37$, p = .04). This suggests that, at 2 minutes using the CLS and C-ILS scoring procedures, significant growth between time points was a function of grade level.

For differences across minutes within each time point (e.g., 1 minute Fall, 2 minute Fall, 3 minute Fall), Mauchly's test indicated that the assumption of sphericity was violated for all scoring procedures at each minute and time point ($p < .05$); therefore, the degrees of freedom were corrected using the Greenhouse-Geiser estimates of sphericity (€ = .88). Results indicated a significant main effect for minute of administration for all scoring procedures at 1, 2, and 3 minutes in Fall, Winter, and Spring (see Table 6). This indicates that each grade level demonstrated significant growth from 1 to 2 minutes, 2 to 3 minutes, and 1 to 3 minutes within each time point (Fall/Winter/Spring). A significant minute by grade interaction effect was also found for all scoring procedures across all minutes at each time point (see Table 6). A series of simple and repeated contrasts indicated that for all scoring procedures there were significant differences in scores obtained from 1 to 2 minutes, 2 to 3 minutes, and 1 to 3 minutes in Fall, Winter, and Spring and these differences were a function of grade level.

Discussion & Implications

The purpose of the current study was to answer the following research questions: (a) To what extent does WD CBM-W maintain technical adequacy across 1 minute time intervals?, and (b) To what extent do statistically significant differences exist between and within grades across the various scoring procedures across 1 minute time intervals? A series of Pearson product-moment correlations were calculated to determine the predictive and concurrent validity of WD with the Spelling subtest of the WIAT-III, a one-way ANOVA was run to explore statistically significant differences between grades, and a RM-ANOVA was run to explore statistically significant differences within grades.

Research Question 1: Evidence of Technical Adequacy

Predictive and concurrent validity results revealed that scoring at 3 minute intervals, particularly with the WSC and C-ILS methods, demonstrated the strongest evidence of validity across grade levels for the Spelling subtest of the WIAT-III. This result is consistent with the limited research available to date in this area (Deno et al., 1980a, 1980b; Deno et al., 1982; Hampton & Lembke, 2016). Additionally, when examining scoring methods, WSC, CLS, and C-ILS revealed stronger evidence of validity when compared to WW, which further reflects past research suggesting that WW is not as reflective of student spelling ability nor of as much instructional utility compared to the other scoring methods (Parker et al., 1991).

For comparisons to our criterion measure, the WIAT-III Spelling subtest, predictive and concurrent validity results demonstrated that although in second and third grade 3 minutes had the largest coefficients for WSC, CLS, and C-ILS (predictive: 2^{nd} grade r = .67–.74, 3^{rd} grade r = .69–.76; concurrent: 2^{nd} grade r = .76–.77, 3^{rd} grade r = .59–.66), scores obtained from one and 2 minute administrations for the same scoring indices also demonstrated adequate evidence of validity (r ≥ .50), where differences were usually within a few hundredths of a point. This indicates that it is possible to obtain a valid estimate and prediction of student spelling ability in second and third grade with a shorter administration of the WD task, which reflects past research investigating minute intervals (Deno et al., 1980b). This is encouraging given that educators' time in the classroom is at a premium and efficiency is vital.

In first grade, however, results were more mixed when measures were compared to the WIAT-III Spelling subtest, the criterion measure. In terms of predictive validity, only C-ILS at 3 minutes and WSC at 1 minute met the acceptable validity criterion level ($r \geq .50$), which suggests two things. One, it is possible to predict future spelling performance with a quick assessment of number of words spelled correctly in 1 minute. Two, a more in-depth scoring procedure for prediction of spelling ability (C-ILS) requires more time (3 minutes). For concurrent validity, no scoring methods met criterion; however, WSC and C-ILS at 3 minutes approached the criterion (r = .48 and .49, respectively). It appears that administering WD for 3 minutes provides valid information about potential future spelling ability but shows less evidence of technical adequacy for estimating current spelling ability in first grade. The moderate predictive validity in first grade aligns with past research showing that a single WD form scored with CLS and C-ILS reached moderate predictive validity with standardized tests of writing

(Hampton & Lembke, 2016). Future research is needed to explore the concurrent validity of WD in first grade with varying minute intervals of administration.

Moderate correlations between Fall WD scores and the Spelling subtest of the WIAT-III suggest average predictive validity. Low to moderate concurrent validity was found between Spring WD scores and the Spelling subtest of the WIAT-III, even though correlation coefficients for WSC and C-ILS were stronger. It may be that the Spelling subtest of the WIAT-III is measuring slightly different constructs than the WD CBM-W, suggesting that a different criterion variable is needed. Because of the moderate predictive validity with the WIAT-III, additional diagnostic assessment is recommended for describing students' strengths and weaknesses, and for making individual educational decisions.

Interestingly, stronger correlation coefficients for predictive and concurrent validity were found at second grade. This might indicate that second grade is a good grade at which to discriminate writing performance based on spelling and spelling patterns. This finding is also consistent with Lembke et al. (2003) who found that WD—when compared to other copying and dictation tasks—had the strongest criterion validity in second grade (r = .80–.92).

Research Question 2: Evidence of Between and Within Grade Level Differences

One-way ANOVA and RM-ANOVA results indicated that the WD task was capable of detecting significant differences between and within grade levels when given for 1, 2, and 3 minutes across an academic year, regardless of scoring procedure, which reflects past research. In terms of within-grade growth, previous studies have suggested that a 3 minute WD task can demonstrate growth across an academic year (Coker & Ritchey, 2013; Deno et al., 1982; Hampton & Lembke, 2016; Lembke et al., 2003), as well as 1 and 2 minute administrations (Hampton & Lembke, 2016). In terms of between-grade growth, previous studies suggested that a 3 minute administration of WD was sensitive to differences between elementary grade levels (Deno et al., 1980a; Deno et al., 1982), but no other studies have investigated or disaggregated results on between-grade differences as a function of minute of administration. The current study is the first study to systematically investigate whether different lengths of WD administration demonstrate statistically significant differences between first, second, and third grade performance. Overall, it appears that shorter 1 and 2 minute administrations of WD, regardless of scoring procedure, have the capacity to detect differences between and within grade levels and may have utility as a screening tool in the early elementary grades. However, when taken with the validity results and considering instructional utility of the scores obtained, a 3 minute administration of WD using the C-ILS scoring procedure showed evidence of being the strongest option for use as a screening tool and detecting early elementary grade level differences. It is worth noting that the C-ILS scoring procedure has been studied relatively less often than the other scoring methods used in this study and in past research. The current study lends further credibility to its use as a technically adequate scoring procedure with CBM-W and gives educators a valid and potentially instructionally useful way to assess their students' writing.

Limitations and Future Research

While this study helps to add to the literature on CBM-W and the technical adequacy of CBM-W, specifically WD, this study is not without its limitations. First, low to moderate predictive and concurrent validity shows that the WIAT-III Spelling subtest may be measuring a slightly different construct than the WD probes. This is striking given that the Spelling subtest of the WIAT-III is in some ways a mirror of the WD probes with the exception that spelling words increase in difficulty, that words are presented within the context of a sentence, and that this subtest is not timed. It may be that what is really being measured by the WD probes is not really spelling ability, but rather a proxy of writing fluency given the timed nature of these probes. Moreover, the reader might recall that spelling has remained a pinnacle component of models of early writing (e.g., Berninger et al., 2002). Though CBMs have been recognized for their strength as global indicators of students' performance, it may be that the general uniformity of these probes can only be used for more specific information about a student's spelling patterns based on vocabulary (i.e., spelling words) that is appropriate to age, rather than being a strong representation of students' abilities to work with words of increasing difficulty as found on standardized measures of spelling like the WIAT-III. Future research may include other administration times, more longitudinal data, different criterion variables (e.g., state assessment), predictive validity of performance across grade levels, as well as predictive power of WD to sentence-level and passage-level spelling in future grades.

Second, this study includes data from only a sample of that collected from a larger study given that these individuals had also completed the WIAT-III as a criterion measure. However, while it is preferred that all students within the larger study would have completed the subtests of the WIAT-III, our sample of students is representative of the larger study population.

Third, this study uses scoring methods that have been traditionally utilized within the CBM research on word level probes. Unfortunately, WD CBM-W are not as well researched as other CBM-W probes, especially those evaluating text generation at the sentence or story/paragraph level. Future research must continue to examine the validity of word-level probes as well as the scoring methods that are commonly used. Research over about the last 15 years has begun to explore a number of alternative scoring methods for sentence and story/paragraph level CBM-W (e.g., Allen, Poch, & Lembke, 2018; Wagner, Smith, Allen, McMaster, Poch, & Lembke, 2018; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). It is possible that comparable scoring indices such as the mean length of correct letter sequences or the average number of correct letter sequences per response may provide additional information about students' spelling progress that also contains acceptable technical adequacy.

Implications for Practice

Spelling remains a critical aspect of writing ability, a skill that can significantly limit the cognitive reserves that a student has for producing longer more connected text (Berninger & Amtmann, 2003; Berninger et al., 2002; Berninger & Winn, 2006; Juel, et al., 1986; McCutchen, 1996). Educators should feel confident in teaching spelling (particularly using direct instructional techniques) and keeping spelling a part of their writing curricula, as research has continued to demonstrate strong connections between transcription and text generation (e.g., Berninger et al., 2002). Educators should also feel confident using WD CBMs to monitor students' progress when implemented with fidelity for a minimum of 2 to 3 minutes in the early grades. The data gleaned from these measures provides educators useful information for informing future intervention and instruction. The more information an educator can glean about a student's spelling abilities, the better informed he/she will be for supporting the often unique and individual spelling needs of struggling writers and writers with identified disabilities in orthography.

References

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281-298. doi:10.1037/a0019318

Allen, A. A., Jung, P–G., Poch, A. L., Brandes, D., Shin, J., Lembke, E. S., & McMaster, K. L. (2018). Technical adequacy of curriculum-based measures in writing in grades 1–3. Manuscript in preparation.

Allen, A., Poch, A. L., & Lembke, E. (2018). An exploration of alternative scoring methods in early writing using curriculum-based measurement of writing. *Learning Disability Quarterly, 41*(2), 85-99. doi:10.1177/0731948717725490

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345-363). New York, NY: Guilford Press.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the Simple View of Writing. *Journal of Educational Psychology, 94*(2), 291-304. doi:10.1037//0022-0663.94.2.291

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96-114). New York, NY: The Guilford Press.

Calfee, R. C., & Miller, R. G. (2013). Best practices in writing assessment for instruction.

In S. Graham, C. MacArthur, and J. Fitzgerald (Eds.), *Best practices in writing instruction (2nd ed.)* (pp. 351-380). New York: Guilford.

Carlisle, A., Poch, A., & Lembke, E. (2015). Technical report for the DBI-TLC project: Curriculum-based measurement screening study 1. Columbia, MO: DBI-TLC.

Clemens, N. H., Oslund, E. L., Simmons, L. E., & Simmons, D. (2014). Assessing spelling in kindergarten: Further comparison of scoring metrics and their relation to reading skills. *Journal of School Psychology, 52*(1), 49-61. https://doi.org/10.1016/j.jsp.2013.12.005

Coker, D. L., & Ritchey, K. D. (2013). Universal screening for writing risk in kindergarten. *Assessment for Effective Intervention, 39*(1), 1–12. doi:10.1177/1534508413502389

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232. doi:10.1177/001440298505200303

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184–192. doi:10.1177/00224669030370030801

Deno, S. L., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study. (Research Report No. 87). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P., Lowry, L., & Kuehnle, K. (1980a). Relationships among simple measures of spelling and performance on standardized achievement tests. (Research Report No. 21). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980b). Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Espin, C., Wallace, T., Campbell, H., Lembke, E., Long, J. D., & Ticha, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children, 74*(2), 174-193. doi:10.1177/001440290807400203

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 alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*(4), 477–497.

Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: A review. *Learning Disability Quarterly, 22*(2), 77-98. doi:10.2307/1511268

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*(1), 170-182. doi:10.1037/0022-0663.89.1.170

Hammill, D. D., & Larsen, S. C. (1978). The test of written language. Austin, TX: PRO-ED.

Hampton, D. D., & Lembke, E. S. (2016). Examining the technical adequacy of progress monitoring using early writing curriculum-based measures. *Reading & Writing Quarterly, 32*(4), 336-352. doi:10.1080/10573569.2014.973984

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. doi:10.1037/0022-0663.78.4.243

Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, *21*(5), 517–525. doi:10.1016/j.lindif.2011.06.004

Kim, Y-S., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35-50. doi:10.1037/edu0000129

Larsen, S. C. & Hammill, D. D. (1976). *Test of written spelling*. Austin, TX: Empiric Press.

Larsen, S. C., Hammill, D. D., & Moats, L. C. (1999). Test of written spelling (4th ed.). Austin, TX: Pro-Ed.

Lee, L., & Canter, S. M. (1971). Developmental sentence scoring. *Journal of Speech and Hearing Disorders, 36*, 335-340.

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*(3-4), 23–35. doi:10.1177/073724770302800304

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1978). *Stanford achievement test*. New York: Harcourt Brace Jovanovich.

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

Marston, D., & Deno, S. L. (1981). The reliability of simple, direct measures of written expression (Research Report No. 50). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Masterson, J. J., & Apel, K. (2010). The spelling sensitivity score: Noting developmental changes in spelling knowledge. *Assessment for Effective Intervention, 36*, 35–45. https://doi.org/10.1177%2F1534508410380039

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychologist, 8*(3), 299-325. doi:10.1007/BF01464076

McMaster, K. L., & Campbell, H. (2008). New and existing curriculum-based writing measures: Technical features within and across grades. *School Psychology Review*, *37*(4), 550–566.

McMaster, K. L., Du, X., & Petursdottir, A.-L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, *42*(1), 41–60. doi:10.1177/0022219408326212

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*(2), 185-206. doi:10.1177/001440291107700203

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in Writing: A literature review. *The Journal of Special Education*, *41*(2), 68-84. doi:10.1177/00224669070410020301

Parker, D. C., McMaster, K. L., Medhanie, A., & Silberglitt, B. (2011). Modeling early writing growth with curriculum-based measures. *School Psychology Quarterly*, *26*(4), 290–304. doi:10.1037/a0026833.

Parker, R., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality*, *2*, 1–17. doi:10.1080/09362839109524763

Pearson. (2009). Wechsler Individual Achievement Test-Third Edition.

Ritchey, K. D. (2006). Learning to write: Progress-monitoring tools for beginning and at-risk writers. *Teaching Exceptional Children*, *39*(2), 22–27. doi:10.1177/004005990603900204

Ritchey, K. D., & Coker, D. L. (2014). Identifying writing difficulties in first grade: An investigation of writing and reading measures. *Learning Disabilities Research & Practice*, *29*(2), 54–65. doi:10.1111/ldrp.12030

Ritchey, K. D., & Coker, D. L. (2013). An investigation of the validity and utility of two curriculum-based measurement writing tasks. *Reading & Writing Quarterly*, *29*(1), 89–119. doi:10.1080/10573569.2013.741957

Ritchey, K. D., Coker, D. L., Jr., & McCraw, S. B. (2010). A comparison of metrics for scoring beginning spelling. *Assessment for Effective Intervention, 35*, 78–88. https://doi.org/10.1177%2F1534508409336087

Taylor, R. L. (2003). *Assessment of exceptional students: Educational and psychological procedures (6th ed.)*. Boston: Allyn & Bacon.

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.

Wagner, K., Smith, A., Allen, A., McMaster, K., Poch, A., & Lembke, E. (2018). Exploration of new complexity metrics for curriculum-based measures of writing. *Assessment for Effective Intervention*. Advance online publication. https://doi.org/10.1177/1534508418773448

Weiser, B., & Mathes, P. (2011). Using encoding instruction to improve the reading and spelling performances of elementary students at risk for literacy difficulties: A best-evidence synthesis. *Review of Educational Research*, *81*(2), 170–200. https://doi.org/10.3102%2F0034654310396719

Table 1

Means and Standard Deviations (in parentheses) For First, Second, and Third Grade

	Fall			Winter			Spring		
	Minutes			Minutes			Minutes		
Scoring Method	1	2	3	1	2	3	1	2	3
First Grade									
WW	6.34	12.86	20.06	6.76	13.52	20.48	7.90	16.27	24.42
	(2.35)	(4.35)	(6.09)	(2.28)	(4.26)	(6.13)	(2.22)	(4.36)	(6.41)
WSC	3.13	6.37	9.34	3.48	7.10	10.40	4.92	10.33	14.74
	(2.30)	(4.52)	(6.16)	(2.20)	(4.22)	(5.92)	(2.71)	(5.29)	(7.44)
CLS	24.68	48.94	72.26	27.17	52.79	76.20	34.38	68.44	97.74
	(12.25)	(21.73)	(28.60)	(12.04)	(19.98)	(26.72)	(12.53)	(23.22)	(32.96)
C-ILS	17.04	34.84	49.02	20.09	39.12	54.34	28.15	56.35	76.96
	(14.35)	(24.81)	(32.75)	(14.30)	(23.49)	(30.64)	(16.21)	(28.30)	(38.47)
WIAT Spelling	103.20 (10.65)								
Second Grade									
WW	8.94	17.63	26.94	8.77	17.83	27.46	10.54	21.40	31.94
	(2.50)	(5.00)	(7.40)	(2.71)	(5.67)	(7.74)	(3.12)	(6.54)	(9.27)
WSC	6.10	11.88	17.60	6.23	12.69	18.12	8.12	15.77	22.28
	(3.77)	(6.95)	(9.89)	(3.87)	(7.24)	(9.99)	(4.45)	(8.58)	(12.02)
CLS	40.65	76.06	111.54	40.45	75.15	113.74	49.36	94.00	137.16
	(16.17)	(29.42)	(42.75)	(16.76)	(29.30)	(42.77)	(17.85)	(35.44)	(53.58)
C-ILS	35.29	64.23	91.08	35.55	64.19	92.70	44.66	81.65	115.62
	(19.57)	(34.28)	(50.14)	(19.47)	(32.76)	(47.93)	(21.07)	(41.25)	(62.06)
WIAT Spelling	94.24 (12.86)								
Third Grade									
WW	11.32	22.94	33.46	11.47	23.56	34.93	12.77	25.87	38.98
	(3.14)	(5.77)	(8.93)	(3.68)	(7.31)	(9.84)	(3.33)	(6.10)	(8.93)
WSC	9.57	18.85	26.16	10.23	20.02	27.32	11.19	22.04	32.59
	(3.65)	(6.38)	(9.81)	(3.98)	(7.78)	(9.63)	(3.71)	(6.49)	(9.84)
CLS	55.92	106.72	151.34	57.17	109.23	153.87	62.77	119.11	181.96
	(15.97)	(28.04)	(47.13)	(19.01)	(33.90)	(48.45)	(17.13)	(31.04)	(48.30)
C-ILS	52.51	97.92	135.30	54.15	100.31	137.02	59.21	109.73	166.59
	(16.76)	(29.34)	(50.06)	(20.39)	(36.15)	(52.36)	(18.36)	(32.87)	(51.58)
WIAT Spelling	94.48 (10.58)								

Table 2

Correlations of Matched Scoring Indices across Time and Grade

		Fall–Winter			Fall–Spring			Winter–Spring		
Grade	Scoring Index	1m	2m	3m	1m	2m	3m	1m	2m	3m
1	WW - 1	0.62	0.63	0.67	0.65	0.65	0.70	0.74	0.73	0.76
	WW - 2	0.76	0.77	0.80	0.72	0.75	0.80	0.77	0.78	0.82
	WW - 3	0.77	0.78	0.81	0.72	0.74	0.80	0.76	0.77	0.82
	WSC - 1	0.78	0.81	0.87	0.83	0.80	0.80	0.72	0.66	0.72
	WSC - 2	0.77	0.80	0.87	0.84	0.84	0.86	0.83	0.79	0.86
	WSC - 3	0.78	0.81	0.89	0.83	0.82	0.86	0.85	0.81	0.87
	CLS - 1	0.72	0.75	0.79	0.84	0.79	0.81	0.70	0.68	0.73
	CLS - 2	0.73	0.77	0.82	0.83	0.83	0.84	0.84	0.82	0.88
	CLS - 3	0.75	0.78	0.83	0.84	0.83	0.85	0.84	0.84	0.89
	C-ILS - 1	0.73	0.77	0.82	0.82	0.78	0.80	0.66	0.64	0.69
	C-ILS - 2	0.69	0.76	0.80	0.80	0.81	0.83	0.83	0.81	0.88
	C-ILS - 3	0.67	0.73	0.80	0.78	0.78	0.82	0.83	0.83	0.88
2	WW - 1	0.78	0.81	0.85	0.77	0.82	0.82	0.74	0.79	0.78
	WW - 2	0.83	0.87	0.89	0.73	0.80	0.81	0.76	0.80	0.80
	WW - 3	0.82	0.86	0.88	0.72	0.78	0.80	0.82	0.86	0.87
	WSC - 1	0.89	0.89	0.92	0.87	0.92	0.92	0.84	0.91	0.89
	WSC - 2	0.92	0.93	0.95	0.86	0.91	0.92	0.84	0.90	0.90
	WSC - 3	0.91	0.92	0.95	0.85	0.91	0.92	0.87	0.94	0.93
	CLS - 1	0.86	0.81	0.92	0.85	0.91	0.91	0.81	0.87	0.86
	CLS - 2	0.87	0.82	0.94	0.82	0.89	0.90	0.77	0.84	0.84
	CLS - 3	0.84	0.82	0.93	0.80	0.86	0.87	0.86	0.92	0.92
	C-ILS - 1	0.86	0.83	0.93	0.85	0.92	0.92	0.82	0.89	0.88
	C-ILS - 2	0.87	0.85	0.94	0.84	0.91	0.92	0.79	0.88	0.88
	C-ILS - 3	0.85	0.84	0.94	0.81	0.89	0.91	0.86	0.93	0.93
3	WW - 1	0.71	0.76	0.71	0.58	0.62	0.64	0.80	0.84	0.87
	WW - 2	0.83	0.84	0.82	0.69	0.70	0.74	0.79	0.82	0.85
	WW - 3	0.80	0.84	0.82	0.68	0.73	0.74	0.82	0.87	0.89
	WSC - 1	0.67	0.79	0.70	0.63	0.71	0.70	0.79	0.83	0.85
	WSC - 2	0.76	0.86	0.81	0.76	0.80	0.83	0.81	0.86	0.88
	WSC - 3	0.74	0.85	0.83	0.75	0.83	0.82	0.81	0.88	0.89
	CLS - 1	0.73	0.78	0.69	0.64	0.69	0.68	0.80	0.84	0.86
	CLS - 2	0.80	0.86	0.78	0.75	0.77	0.79	0.78	0.81	0.85
	CLS - 3	0.82	0.86	0.83	0.74	0.80	0.79	0.80	0.85	0.88
	C-ILS - 1	0.72	0.78	0.72	0.66	0.71	0.70	0.77	0.81	0.83
	C-ILS - 2	0.75	0.83	0.80	0.76	0.80	0.80	0.78	0.82	0.85
	C-ILS - 3	0.79	0.85	0.85	0.76	0.83	0.81	0.79	0.86	0.86

All correlations significant at $p \leq .01$ except where indicated. [a] = not significant, [b] = $p \leq .05$, m = minute; WW = Words Written, WSC = Words Spelled Correctly, CLS = Correct Letter Sequences, C-ILS = Correct Minus Incorrect Letter Sequences

Table 3

Predictive and Concurrent Validity (uses age norms)

WIAT Subtest	Scoring method	First			Second			Third		
		Minutes			Minutes			Minutes		
		1	2	3	1	2	3	1	2	3
		Predictive Validity								
Spelling	WW	.28	.09	.15	.44**	.44**	.48**	.49**	.51**	.58**
	WSC	**.50****	.41**	.49**	**.73****	**.73****	**.74****	**.68****	**.74****	**.76****
	CLS	.43**	.31*	.41**	**.66****	**.66****	**.67****	**.60****	**.66****	**.69****
	C-ILS	.47**	.42**	**.53****	**.72****	**.73****	**.73****	**.66****	**.73****	**.75****
		Concurrent Validity								
Spelling	WW	.13	.11	.11	.43**	.48**	**.51****	**.54****	**.53****	**.51****
	WSC	.48**	.48**	.49**	**.70****	**.74****	**.77****	**.62****	**.65****	**.66****
	CLS	.41**	.37**	.37**	**.62****	**.68****	**.70****	**.58****	**.60****	**.59****
	C-ILS	.47**	.45**	.48**	**.69****	**.76****	**.77****	**.59****	**.64****	**.63****

**p ≤ .05; **p ≤ .01*

Table 4

Word Dictation One-Way ANOVA Results

		One Minute					Two Minutes					Three Minutes				
		SS	df	MS	*F*	*p*	SS	df	MS	*F*	*p*	SS	df	MS	*F*	*p*
Fall																
WW	Between	582.89	2	291.45	40.58	.000	2437.91	2	1218.96	47.53	.000	4490.08	2	2245.04	39.23	.000
	Within	1005.58	140	7.18			3616.06	141	25.65			8412.06	147	57.22		
	Total	1588.48	142				6053.97	143				12902.14	149			
WSC	Between	978.69	2	489.34	44.57	.000	3750.29	2	1875.15	51.59	.000	7073.56	2	3536.78	45.77	.000
	Within	1537.21	140	10.98			5124.60	141	36.34			11357.94	147	77.26		
	Total	2515.90	142				8874.89	143				18431.50	149			
CWS	Between	22929.85	2	11464.93	51.48	.000	80146.29	2	40073.15	56.78	.000	156343.41	2	78171.71	48.19	.000
	Within	31176.97	140	222.69			99517.03	141	705.79			238461.26	147	1622.19		
	Total	54106.83	142				179663.33	143				394804.67	149			
CIWS	Between	29571.00	2	14785.50	50.77	.000	95522.06	2	47761.03	54.14	.000	186144.84	2	93072.42	45.83	.000
	Within	40769.66	140	291.21			124376.83	141	882.11			298549.16	147	2030.95		
	Total	70340.66	142				219898.89	143				484694.00	149			
Winter																
WW	Between	519.47	2	259.74	29.74	.000	2291.52	2	1145.76	32.43	.000	5221.84	2	2610.92	40.33	.000
	Within	1196.50	137	8.73			4768.96	135	35.33			9517.65	147	64.75		
	Total	1715.97	139				7060.47	137				14739.49	149			
WSC	Between	1074.56	2	537.28	44.93	.000	3795.41	2	1897.70	42.45	.000	7173.99	2	3587.00	47.26	.000
	Within	1638.33	137	11.96			6034.91	135	44.70			11157.80	147	75.90		
	Total	2712.89	139				9830.32	137				18331.79	149			
CWS	Between	21028.54	2	10514.27	39.95	.000	73240.19	2	36620.09	44.66	.000	150874.55	2	75437.28	46.28	.000
	Within	36056.86	137	263.19			110705.53	135	820.04			239615.39	147	1630.04		
	Total	57085.40	139				183945.72	137				390489.94	149			
CIWS	Between	27067.34	2	13533.67	40.53	.000	85672.88	2	42836.44	42.99	.000	171195.57	2	85597.79	42.96	.000

	Within	45743.23	137	333.89			134516.03	135	996.42			292876.70	147	1992.36		
	Total	72810.57	139				220188.91	137				464072.27	149			
Spring																
WW	Between	565.05	2	282.52	32.85	.000	2145.98	2	1072.99	32.61	.000	5249.28	2	2624.64	38.14	.000
	Within	1221.32	142	8.60			4540.16	138	32.90			10045.98	146	68.81		
	Total	1786.37	144				6686.14	140				15295.26	148			
WSC	Between	935.34	2	467.67	34.05	.000	3187.08	2	1593.54	33.19	.000	7942.50	2	3971.25	40.13	.000
	Within	1950.22	142	13.73			6625.06	138	48.01			14447.54	146	98.96		
	Total	2885.56	144				9812.14	140				22390.04	148			
CWS	Between	19170.36	2	9585.18	37.31	.000	59674.01	2	29837.01	32.48	.000	175700.96	2	87850.48	41.93	.000
	Within	36477.20	142	256.88			126786.26	138	918.74			305894.26	146	2095.17		
	Total	55647.56	144				186460.27	140				481595.22	148			
CIWS	Between	22962.76	2	11481.38	32.86	.000	66187.24	2	33093.62	27.65	.000	199898.49	2	99949.25	37.52	.000
	Within	49611.07	142	349.37			165190.76	138	1197.03			388899.54	146	2663.70		
	Total	72573.83	144				231378.00	140				588798.03	148			

Table 5

RM-ANOVA Results Within Minute of Administration by Fall, Winter, Spring

	1 minute			
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Time	167.06	35.95	2	<.001
Time * Grade	5.21	0.56	4	0.69
Within-Subjects Contrasts				
Fall to Winter	7.81	0.86	2	0.43
Winter to Spring	7.28	0.95	2	0.39
Fall to Spring	0.55	0.05	2	0.95
WSC				
Time	221.28	42.28	2	<.001
Time * Grade	11.10	1.06	4	0.38
Within-Subjects Contrasts				
Fall to Winter	11.57	1.19	2	0.31
Winter to Spring	20.35	2.00	2	0.14
Fall to Spring	1.38	0.12	2	0.89
CLS				
Time	4976.08	46.02	2	<.001
Time * Grade	196.77	0.91	4	0.46
Within-Subjects Contrasts				
Fall to Winter	193.49	0.96	2	0.39
Winter to Spring	275.75	1.28	2	0.28
Fall to Spring	121.07	0.52	2	0.59
CILS				
Time	5739.13	40.69	2	<.001
Time * Grade	443.78	1.57	4	0.18
Within-Subjects Contrasts				
Fall to Winter	226.38	0.91	2	0.41
Winter to Spring	607.36	1.95	2	0.15
Fall to Spring	497.60	1.75	2	0.18

	2 minute			
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Time	684.77	53.27	2	<.001
Time * Grade	19.98	0.78	4	0.54
Within-Subjects Contrasts				
Fall to Winter	28.63	1.39	2	0.25
Winter to Spring	28.89	1.09	2	0.34
Fall to Spring	2.42	0.08	2	0.92
WSC				
Time	778.15	67.53	2	<.001
Time * Grade	28.28	1.23	4	0.30
Within-Subjects Contrasts				
Fall to Winter	32.84	1.73	2	0.18
Winter to Spring	43.81	1.75	2	0.18
Fall to Spring	8.20	0.33	2	0.72

CLS				
Time	17715.79	63.07	2	<.001
Time * Grade	1578.04	2.81	4	0.03
Within-Subjects Contrasts				
Fall to Winter	1676.32	3.28	2	0.04
Winter to Spring	2057.24	3.35	2	0.04
Fall to Spring	1000.55	1.78	2	0.17
CILS				
Time	18228.71	53.99	2	<.001
Time * Grade	2012.74	2.98	4	0.02
Within-Subjects Contrasts				
Fall to Winter	1696.99	2.67	2	0.07
Winter to Spring	2422.92	3.37	2	0.04
Fall to Spring	1918.30	2.86	2	0.06

	3 minute			
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Time	1951.01	88.00	1.84	<.001
Time * Grade	23.60	0.53	3.67	0.70
Within-Subjects Contrasts				
Fall to Winter	33.75	0.84	2	0.43
Winter to Spring	23.85	0.67	2	0.51
Fall to Spring	13.19	0.23	2	0.80
WSC				
Time	2414.60	122.01	1.92	<.001
Time * Grade	19.15	0.48	3.83	0.74
Within-Subjects Contrasts				
Fall to Winter	12.34	0.36	2	0.70
Winter to Spring	7.95	0.22	2	0.80
Fall to Spring	37.17	0.78	2	0.46
CLS				
Time	61570.82	123.40	1.87	<.001
Time * Grade	171.25	0.17	3.73	0.95
Within-Subjects Contrasts				
Fall to Winter	83.51	0.10	2	0.91
Winter to Spring	295.39	0.35	2	0.71
Fall to Spring	134.87	0.11	2	0.90
CILS				
Time	64416.42	108.66	1.92	<.001
Time * Grade	385.18	0.33	3.83	0.85
Within-Subjects Contrasts				
Fall to Winter	417.31	0.41	2	0.67
Winter to Spring	349.70	0.32	2	0.73
Fall to Spring	388.52	0.27	2	0.76

Table 6

RM-ANOVA Results Within Time Point by Minute of Administration

Fall				
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Minute	23146.40	1713.64	1.13	<.001
Minute x Grade	1037.22	38.40	2.25	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	620.44	38.29	2	<.001
2 minute to 3 minute	423.10	30.13	2	<.001
1 minute to 3 minute	2068.11	40.72	2	<.001
WSC				
Minute	9582.39	565.05	1.12	<.001
Minute x Grade	1495.13	44.08	2.24	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	842.35	42.15	2	<.001
2 minute to 3 minute	656.63	37.00	2	<.001
1 minute to 3 minute	2986.40	46.65	2	<.001
CLS				
Minute	369099.18	1065.69	1.11	<.001
Minute x Grade	32958.87	47.58	2.22	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	16680.24	44.97	2	<.001
2 minute to 3 minute	16285.42	41.40	2	<.001
1 minute to 3 minute	65910.94	50.17	2	<.001
CILS				
Minute	236562.77	544.86	1.13	<.001
Minute x Grade	37134.40	42.76	2.25	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	18133.28	40.96	2	<.001
2 minute to 3 minute	19018.25	35.85	2	<.001
1 minute to 3 minute	74251.68	45.50	2	<.001

Winter				
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Minute	22804.34	1387.81	1.15	<.001
Minute x Grade	1068.86	32.52	2.29	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	646.64	29.85	2	<.001
2 minute to 3 minute	438.65	27.67	2	<.001
1 minute to 3 minute	2121.30	34.73	2	<.001
WSC				
Minute	9336.84	507.96	1.22	<.001
Minute x Grade	1188.77	32.34	2.44	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	828.92	32.16	2	<.001
2 minute to 3 minute	394.25	21.10	2	<.001
1 minute to 3 minute	2343.14	35.60	2	<.001

CLS				
Minute	351447.38	822.61	1.36	<.001
Minute x Grade	25292.59	29.60	2.72	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	15570.78	31.97	2	<.001
2 minute to 3 minute	10952.51	17.05	2	<.001
1 minute to 3 minute	49354.47	34.41	2	<.001
CILS				
Minute	223024.86	430.43	1.35	<.001
Minute x Grade	26471.90	25.55	2.69	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	16282.87	27.44	2	<.001
2 minute to 3 minute	11166.24	14.61	2	<.001
1 minute to 3 minute	51966.58	29.67	2	<.001

		Spring		
Source	**Sum of Squares**	***F***	***df***	***p***
WW				
Minute	32311.95	1988.13	1.06	<.001
Minute x Grade	1009.69	31.06	2.12	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	518.79	27.34	2	<.001
2 minute to 3 minute	491.83	31.84	2	<.001
1 minute to 3 minute	2018.46	31.99	2	<.001
WSC				
Minute	16115.51	700.38	1.09	<.001
Minute x Grade	1492.33	32.43	2.18	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	715.06	27.83	2	<.001
2 minute to 3 minute	779.48	32.21	2	<.001
1 minute to 3 minute	2982.46	33.83	2	<.001
CLS				
Minute	572367.51	1136.05	1.08	<.001
Minute x Grade	34464.27	34.20	2.16	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	11920.64	24.06	2	<.001
2 minute to 3 minute	23197.36	39.19	2	<.001
1 minute to 3 minute	68274.82	35.28	2	<.001
CILS				
Minute	403350.64	620.87	1.09	<.001
Minute x Grade	38545.21	29.67	2.19	<.001
Within-Subjects Contrasts				
1 minute to 2 minute	12349.39	19.27	2	<.001
2 minute to 3 minute	27162.67	34.99	2	<.001
1 minute to 3 minute	76123.56	30.69	2	<.001