Article

Hammill Institute on Disabilities

Assessment for Effective Intervention
2023, Vol. 48(4) 201–210

DOI: 10.1177/15345084231182718
aei.sagepub.com
Sage

Monitoring Elementary Students' Progress Using Word Dictation: Technical Features of Slope and Growth Analysis

Seohyeon Choi, MA[1], Emma Shanahan, PhD[1], Jechun An, MA[1], and Kristen McMaster, PhD[1]

Abstract

The purpose of this study was to examine the technical features of slopes produced from the curriculum-based measurement in writing (CBM-W) word dictation task. Seventy-nine elementary students in the U.S. Midwest with intensive learning needs responded to weekly word dictation probes across 20 weeks; responses were scored for correct letter sequences (CLS). Scores showed evidence of high reliability and sensitivity to growth in a short period. Linear mixed modeling revealed that students gained an average of 0.91 CLS for each additional week of instruction. Initial writing levels and growth rates did not significantly differ depending on students' demographic characteristics. Based on these findings, we illustrate how word dictation slopes can be used as indicators of writing growth for students with intensive learning needs.

Keywords

word dictation, curriculum-based measurement, slope, early writing, struggling learners

Writing difficulties can begin as early as preschool (Puranik & Lonigan, 2012) yet are often not identified until intermediate grades when students are expected to complete more complex writing tasks (Berninger et al., 2002). Early identification of writing difficulties along with writing interventions can prevent long-lasting negative consequences for many students (McMaster et al., 2018). Curriculum-based measurement (CBM; Deno, 1985) can play a vital role in quantifying students' writing improvement and assessing their responsiveness to instruction (Fuchs, 2004), facilitating early identification of writing difficulties and progress monitoring during interventions. Among different types of CBM for writing (CBM-W) tasks, word dictation, the main focus of this study, serves as a progress monitoring tool for beginning writers who are developing writing skills at the word level (Hampton & Lembke, 2016).

Despite its promise as a progress monitoring tool for early writing proficiency, research on word dictation is limited. Thus far, researchers (Hampton & Lembke, 2016; Lembke et al., 2003; Poch et al., 2019) have investigated the technical adequacy of word dictation using scores collected at a single time point. Given that research suggests three stages of research necessary for validating CBM (Fuchs, 2004)–Stage 1, assessing the technical adequacy of static scores; Stage 2, investigating technical features of slopes; and Stage 3, determining instructional utility–word dictation still requires Stage 2 research. In this study, we aimed to examine the technical features of word dictation slopes by conducting a conceptual replication of studies by McMaster et al. (2011) and Romig and Olsen (2021) and to examine whether students' demographic characteristics were associated with their initial performance levels and progress.

CBM for Students Struggling With Writing

Although many students respond to standardized writing intervention protocols, others likely require more intensive interventions (Al Otaiba et al., 2018). Such interventions may be delivered at a higher dosage in smaller student groups (Wanzek & Vaughn, 2009). Teachers may also make data-based decisions to increase intervention intensity if students are not making adequate progress through a process termed data-based instruction (DBI; Deno & Mirkin, 1977). Research has shown promising results in improving the writing skills of struggling

[1]University of Minnesota-Twin Cities, Minneapolis, USA

Corresponding Author:
Seohyeon Choi, Department of Educational Psychology, University of Minnesota-Twin Cities, Education Sciences Building, 56 E River Road, Minneapolis, MN 55455, USA.
Email: choi0836@umn.edu

Associate Editor: Abigail Allen

learners through intensive writing interventions (Jung et al., 2017; McMaster et al., 2020).

To monitor student response to writing intervention, CBM (Deno, 1985) is one of the few validated assessment tools that teachers can adopt. CBM serves as a *general outcome measure*, meaning it provides an overall indicator of academic performance and progress in a given academic domain, enabling educators to make decisions about students' overall progress toward important learning goals (Fuchs & Deno, 1991). Research demonstrates that CBM yields valid writing proficiency data for elementary students (Deno et al., 1982; McMaster & Espin, 2007), and teachers' use of CBM for progress monitoring improves outcomes for struggling learners (Stecker et al., 2005). For written expression, researchers proposed simple scoring procedures, such as the number of words written (WW), words spelled correctly (WSC), and correct letter sequences (CLS; any two adjacent letters placed according to the correct spelling of the word) (Deno et al., 1980, 1982) prompting researchers to explore various CBM-W tasks.

CBM-W Word Dictation

Early investigations of CBM-W tasks paid attention to sentence- or passage-level prompts, although researchers also developed CBM tasks focused on spelling (Fuchs et al., 1991a, 1991b; Ritchey, 2006; Ritchey et al., 2010). While these measures had established reliability, validity, and sensitivity evidence for upper-elementary students, relatively less has been examined for beginning writers with intensive learning needs (McMaster et al., 2011; McMaster & Espin, 2007). In response to the need for an adequate progress monitoring tool for early elementary students and students experiencing foundational writing needs, researchers developed CBM-Ws at the word level, including word dictation (Lembke et al., 2003).

Word dictation was designed to measure students' transcription skills. The Simple View of Writing (Berninger & Amtmann, 2003) identifies three key components of developing early writing: transcription, text generation, and self-regulation. Transcription relates to skills to translate language representations in working memory into orthographic symbols using a writing implement (e.g., pen or pencil) and includes skills for handwriting (e.g., holding a pen or pencil comfortably, writing fluently without excessive erasing/scribbling) as well as spelling (e.g., using correct alphabetic principles in isolation). Difficulties in transcription can interfere with the development of text-generation skills (Berninger & Amtmann, 2003).

Previous research has shown evidence of the validity (i.e., degree to which specific interpretations of test scores for proposed uses of tests are supported by evidence and theory; American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA, APA, NCME], 2014) of word dictation. Specifically, researchers have shown that word dictation, particularly scored for CLS, has evidence of criterion validity (r = .52 to .92; Lembke et al., 2003) in relation to students' written responses to a picture story starter, assessed using CBM-W scoring indices, as well as predictive validity (r = .48 to .50; Hampton & Lembke, 2016) with regard to the Test of Early Written Language-2 (TEWL-2) for beginning writers. Poch and colleagues (2019) provided evidence of predictive and concurrent validity in relation to the spelling subtest of the Wechsler Individual Achievement Test-III (WIAT-III) for students in Grades 1 to 3. Moreover, recent studies (Keller-Margulis et al., 2019; Smith & Lembke, 2022) have examined the technical adequacy of word dictation with specific test-taking populations, such as English Language Learners (ELLs), using a criterion measure designed and normed for ELLs (e.g., Accessing Comprehensive and Communication in English State-to-State [ACCESS] English Language Proficiency test). Collectively, findings indicate that word dictation has produced similar levels of validity evidence for ELLs as for the general population.

Although research has provided validity evidence for word dictation, it is important to acknowledge that this evidence has primarily been based on static scores, with limited attention paid to *slopes* generated from repeated measurements. This is concerning considering that slopes serve an important role in guiding teachers' progress monitoring and instructional decision-making for students. The next step in CBM validation research, as outlined by Fuchs (2004), is to examine the technical features of slopes to determine the measure's utility in measuring student progress. The primary purpose of this study was to add empirical evidence of the technical features of word dictation slopes using data obtained from struggling beginning writers.

Our secondary purpose was to examine whether and to what extent students' performance and/or growth, indicated by word dictation slopes, vary depending on their demographic characteristics. It is important to ensure that word dictation measures the same construct across diverse test-taking populations, without under- or overestimating the performance of some groups. Previous research on CBM in reading reported differences in intercepts on race, ethnicity, and gender (Kranzler et al., 1999) and in slopes on special education status and gender (Yeo et al., 2011); however, how demographic characteristics relate to students' performance on CBM-W remains understudied. Thus, we aimed to investigate whether word dictation produces discrepant intercepts or slopes depending on demographic grouping.

Purpose and Research Questions

The purpose of this study was to examine technical features of slopes produced from word dictation and to determine whether word dictation scores and/or slopes significantly differed based on students' demographic characteristics. Specific research questions were:

RQ1. Does word dictation scored for CLS produce reliable and stable slopes of writing progress of elementary students with intensive learning needs?
RQ2. Are slopes from word dictation scored for CLS sensitive to growth in writing of elementary students with intensive learning needs?
RQ3. What is the growth pattern for elementary students with intensive learning needs measured by word dictation?
RQ4. Do students' performance levels and growth rates vary depending on their demographic characteristics (sex, grade level, race/ethnicity, and eligibility for free/reduced-price lunch [FRL], and ELL services)?

Method

Setting and Participants

We used data from a larger project that examined the effects of a professional development program on teacher and student outcomes. In this project, participating teachers were assigned randomly to treatment and control conditions, and their students who had intensive early writing needs were selected as "target students." Teachers first nominated students on their caseloads who needed writing intervention or had an Individualized Education Program (IEP) in writing, if applicable, and met the eligibility criteria (i.e., in early elementary grades, have access to the general education curriculum, have functional English skills, can write one or more alphabet letters). The nominated students were then screened using word dictation and picture word (CBM-W designed to capture sentence-level writing skills through prompts composed of words with accompanying pictures). Selected target students (roughly 68% of the nominated students) were those who wrote the fewest CLS (word dictation) and correct word sequences (picture word) compared to other nominated students.

Treatment teachers participated in the professional development program and provided intensive, data-based writing instruction across 20 weeks (typically, at least three times per week for 20 to 30 minutes per session) while control teachers implemented business-as-usual instruction. Treatment teachers learned to align writing instruction to individual students' needs using research-based writing interventions (see McMaster et al., 2018), and typically emphasized transcription (handwriting and spelling) and for more advanced students, text generation (e.g., sentence construction). To monitor students' progress, treatment teachers selected either word dictation, picture word, or story prompts (CBM-W designed to capture discourse-level transcription and text generation skills) based on students' needs and administered the selected task every week. Teachers also intensified instruction as needed based on student progress.

We used data from the treatment students who participated in the project during the 2018–2019 (Cohort 1) and 2019–2020 (Cohort 2) school years. We did not include control group data because the teachers did not collect weekly data. We excluded Cohort 3 (2021–2022) data because the data collection was still ongoing at the time of the current analysis. Cohort 3 data also involved substantial missing data due to the ongoing impact of the COVID-19 pandemic (the pandemic did not affect Cohort 2 progress monitoring data collected for this study). Data from students for whom word dictation was used as a progress monitoring tool were included.

Participants included 79 students taught by 40 teachers (selected from 53 treatment teachers across Cohorts 1 and 2) from 13 public school districts in two U.S. Midwestern states. Table 1 presents student demographic data. Analysis of variance revealed no statistically significant difference in writing skills, measured by word dictation, picture word, and story prompt, among student groups from different sites and cohorts, allowing us to aggregate data. Demographic data were collected from teachers or administrators in each district. Most of the students (90.9%) were in Grades 1–3, while eight students in Grade 4 were included given their significant difficulties in early writing. Of the student participants, 64.6% were male. In terms of the race/ethnicity reported by teachers, more than half (57.0%) were White, 21.5% were African American, 13.9% were Hispanic, 5.1% were Native American/Alaskan Native, 1.3% were Asian, and 1.3% were multi-racial. Most students (96.2%) were eligible for special education services. 44.3% were eligible for free/reduced lunch. All students completed the project without attrition.

Measure

Word dictation assesses word-level transcription skills and is recommended for students just beginning to write words. The measure consists of 20 parallel forms, and each form consists of a list of 30 words. The word lists reflect spelling patterns specified in the Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Word dictation is administered individually for 3 minutes. The teacher dictates each word with one repeat, and students write the word. If the student writes all 30 words prior to the expiration of the allotted time, prorated scores can be calculated.

Table 1. Demographic Data for the 79 Elementary Grades Children in the Study.

Demographic	*n* (%)	Demographic	*n* (%)
Grade		Sex	
First	25 (31.7%)	Male	51 (64.6%)
Second	22 (27.9%)	Female	28 (35.4%)
Third	24 (30.4%)	Primary disability category	
Fourth	8 (10.1%)	Autism	13 (16.5%)
Race/ethnicity		Deaf/hard of hearing	2 (2.5%)
Native American/Alaskan Native	4 (5.1%)	Deaf-blind	1 (1.3%)
African American	17 (21.5%)	Intellectual disability	9 (11.4%)
Asian	1 (1.3%)	Other health disability	17 (21.5%)
Hispanic	11 (13.9%)	Specific learning disability	8 (22.8%)
White	45 (57.0%)	Emotional/behavioral disorder	6 (7.6%)
Multi-racial	1 (1.3%)	Need alternative programming	6 (7.6%)
Special education services eligibility		Speech/language impairment	3 (3.8%)
Eligible	76 (96.2%)	None	3 (3.8%)
English language learner services eligibility		Not applicable	1 (1.3%)
Eligible	12 (15.2%)		
Free/reduced-price lunch eligibility			
Eligible	35 (44.3%)		

Table 2. Descriptive Statistics for Word Dictation Correct Letter Sequences Scores.

Week	*M* (*SD*)	*n*	Week	*M* (*SD*)	*n*
1	29.84 (21.44)	79	11	40.26 (20.91)	42
2	32.39 (20.08)	66	12	42.53 (24.96)	47
3	33.12 (21.44)	65	13	43.83 (24.60)	40
4	35.26 (22.74)	73	14	45.14 (28.83)	37
5	33.81 (22.41)	57	15	46.69 (26.09)	45
6	38.22 (23.56)	73	16	43.55 (28.64)	56
7	38.54 (21.89)	59	17	47.22 (30.80)	54
8	40.82 (24.61)	66	18	49.20 (27.73)	54
9	43.34 (24.60)	50	19	47.37 (32.45)	51
10	41.86 (21.79)	35	20	51.26 (31.57)	43

Previous researchers have reported alternate form reliability coefficients as $r \geq .89$ to .95 and criterion validity as ranging from $r = .11$ to .77 depending on criterion measures (e.g., writing samples in response to a story prompt, teacher ratings, TEWL-2, WIAT-III) for Grades 1-3 (Hampton & Lembke, 2016; Lembke et al., 2003; Poch et al., 2019). For the current sample, correlation coefficients between adjacent weeks' scores ranged from $r = .79$ to .98.

Procedures

Fidelity of word dictation administration and scoring. Researchers taught teachers to administer and score word dictation and checked for fidelity. Trained doctoral graduate research assistants (GRAs) collected video/audio recordings of each teacher administering word dictation on two occasions and scored them using the CBM administration fidelity checklist (see Table S1 in the online supplemental materials). To measure scoring fidelity, GRAs collected one unscored prompt from each teacher each month, scored it independently, and compared it with the prompt scored by the teacher. Administration fidelity and scoring fidelity averaged 90.2% and 96.6%, respectively.

Data coding

Word dictation weekly scores. The dataset consisted of 79 students' weekly responses to word dictation collected across 20 weeks (see Table 2 for descriptive statistics of the weekly scores). The first author coded the data by extracting the students' weekly scores from individual progress monitoring graphs. To ensure the reliability of coding, the third author randomly selected 20% of the graphs and coded them independently. The reliability of coding, calculated as the percentage of agreement, was 99.69%. Additionally, it must be noted that our dataset involved a substantial amount of missing data (30.9% across time points). Results of a missingness test using finalfit.R showed that among a set of demographic variables (grade, sex, ELL, FRL, and special education eligibility), grade was significantly related to the missing scores. As grade was included in our observed data set, this result allowed us to conclude that the scores were missing at random (MAR; in other words, the distribution of missing data depends on unobserved data).

Demographic variables. Students' demographic information was collected using a demographic survey. To include the demographic variables in conditional models as Level

2 predictors (described below), we created five dichotomous variables based on the survey responses as follows: *sex* (male, female), *grade level* (Grades 1–2, Grades 3–4), *race/ethnicity* (White, students from racially/ethnically minoritized backgrounds), *FRL* (eligible for FRL, not eligible for FRL), and *ELL* (eligible for ELL services, not eligible for ELL services). We coded the race/ethnicity dichotomously for the statistical testing due to the limited sample sizes of students with reported race/ethnicity as Native American/Alaskan Native (4), African American (17), Asian (1), Hispanic (11), and Multi-racial (1), as compared to White (45). However, we recognize the limitation of collapsing demographic categories and discuss this in the study limitation section. FRL and ELL variables had missing values ("non-applicable") for 23 and 13 students, respectively; records in these classes were deleted in the conditional model analysis.

Data Analysis

Analysis of technical features of slopes. We conducted a conceptual replication of McMaster et al. (2011) and Romig and Olsen (2021). We evaluated the alternate-slopes reliability by calculating Pearson *r* correlation coefficients between slopes yielded from odd and even weeks. Further, we calculated Pearson *r* correlation coefficients between slopes obtained from incrementally increasing durations (Weeks 1–3, 1–4, etc.) and the slope produced across the full 20-week duration. Regarding sensitivity, we adopted the definition of *sensitivity to growth* used in McMaster et al. (2011) as slopes statistically greater than zero. However, a limitation of this criterion is that it does not indicate how well the measure captures "typical" growth (McMaster et al., 2011), meaning that if a zero slope is produced, it is unclear whether word dictation is not sensitive to writing growth or the students' writing skills actually did not improve over time. Identifying the time point at which the slope first became significantly greater than zero reveals the minimum number of data points needed to capture students' writing improvement. We were also interested in knowing when the slope was *reliable*, in that it had a correlation ($r \geq .50$) with the overall slope, which can be considered a moderately strong relation (Taylor, 1990). Regarding stability, we examined how the standard errors (*SE*) of the mean slopes changed. A decreasing trend in *SE*s indicates decreasing variability among the individual-level slopes, or the increasing stability of the slopes.

Although the current study replicated previous studies, there were some significant departures, including differences related to study participants (first graders in McMaster et al., 2011 and secondary students in Romig & Olsen, 2021), CBM-W task used (picture word, story prompt, and sentence copying prompts in McMaster et al., 2011; story prompt in Romig & Olsen, 2021), and data collection period (12 weeks in McMaster et al., 2011 and 11 in Romig & Olsen, 2021). In addition, in calculating the slopes, we used Linear Mixed Modeling (LMM), whereas the previous studies used Ordinary Least Squares (OLS) regression. As Parker et al. (2011) described in their re-analysis of McMaster et al. (2011), LMM is more appropriate for longitudinal analysis given the hierarchical nature of the data. Using OLS regression without recognizing the hierarchical structure of data can underestimate *SE*s of coefficients.

Analysis of writing growth patterns and impacts of demographic variables. Given the linear pattern identified by a visual inspection of weekly scores over time (see Figure S1 in the online supplemental materials) and the nested data structure, we used two-level LMM (repeated measurements at Level 1 nested within students at Level 2). Based on the comparison of baseline models, we selected the model with randomly varying intercepts and slopes based on relative fitness (smaller Akaike information criterion and Bayesian information criterion values).

We used the baseline model to calculate the intraclass correlation coefficient (ICC) and to determine the intercept and slope for the entire sample. An ICC value above .40 is considered sufficient to use multilevel modeling (Raudenbush & Bryk, 2002). Additionally, we constructed baseline models for each demographic subgroup to compare coefficients. Furthermore, to statistically understand the differences in estimates across subgroups, we included demographic variables one by one as Level 2 predictors and examined whether the variances in coefficients were significantly accounted for by each variable. The final baseline model is as follows:

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{ti} + e_{ti}$$

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

Y_{ti} and a_{ti} are the CLS scores and the linear time variable for individual student i at time t, respectively; π_{0i}, π_{1i}, and e_{ti} represent the intercept (Week 1), the linear slope, and the error variance, respectively; β_{00} and β_{10} represent the mean intercept and the mean linear slope, respectively; and r_{0i} and r_{1i} represent error variance associated with the intercept and slope, respectively.

Results

Does Word Dictation Produce Reliable and Stable Slopes of Writing Growth?

To establish alternate-slope reliability, we calculated slopes using LMM models for odd weeks (Weeks 1, 3 . . . 19) and

Table 3. Correlation Between Slopes From Incremental Durations and the 20-Week Slope.

Slope	Estimate (*SE*)	*t* value	*r*	Slope	Estimate (*SE*)	*t* value	*r*
Week 1–3	2.21(0.72)	3.07*	−0.27	Week 1–12	1.16(0.15)	7.51**	0.70**
Week 1–4	2.25(0.43)	5.27**	−0.06	Week 1–13	1.00(0.14)	7.41**	0.75**
Week 1–5	1.43(0.33)	4.32**	0.10	Week 1–14	0.94(0.13)	7.25**	0.79**
Week 1–6	1.38(0.23)	5.93**	0.32*	Week 1–15	0.88(0.12)	7.20**	0.88**
Week 1–7	1.27(0.22)	5.90**	0.41**	Week 1–16	0.85(0.12)	7.06**	0.91**
Week 1–8	1.33(0.21)	6.38**	0.50**	Week 1–17	0.86(0.12)	7.45**	0.96**
Week 1–9	1.39(0.18)	7.69**	0.59**	Week 1–18	0.87(0.11)	7.79**	0.98**
Week 1–10	1.31(0.17)	7.93**	0.62**	Week 1–19	0.85(0.11)	7.49**	1.00**
Week 1–11	1.17(0.16)	7.40**	0.62**				

Note. r represents Pearson's correlation coefficient.
$^{*}p < .01$. $^{**}p < .001$.

Table 4. Summary Results of Unconditional Models in Each Subgroup.

Parameter	Subgroup	Estimate (*SE*)	*t* value	Subgroup	Estimate (*SE*)	*t* value
Intercept	Male	30.88 (2.92)	10.56**	Female	33.30 (4.91)	6.78**
Slope		0.90 (0.06)	13.91**		0.92 (0.08)	11.07**
Intercept	Grade 1–2	28.58 (2.76)	10.35**	Grade 3–4	36.37 (4.73)	7.69**
Slope		0.86 (0.06)	13.92**		0.96 (0.09)	11.19**
Intercept	White	35.40 (3.64)	9.73**	Students of SREMB	27.23 (3.30)	8.27**
Slope		1.02 (0.07)	14.73**		0.75 (0.07)	10.01**
Intercept	FRL not eligible	34.11 (5.41)	6.30**	FRL eligible	27.36 (3.65)	7.49**
Slope		1.09 (0.12)	9.35**		0.74 (0.07)	11.24**
Intercept	ELL not eligible	33.32 (3.15)	10.56**	ELL eligible	28.44 (4.52)	6.29**
Slope		0.90 (0.06)	14.39**		0.88 (0.12)	7.30**

Note. ELL = English language learners; FRL = free or reduced-price lunch; SREMB = students from racially/ethnically minoritized backgrounds.
$^{**}p < .001$.

even weeks (Weeks 2, 4 . . . 20), and analyzed their correlations. Results showed that the slopes for odd weeks and even weeks were .91 ($t = 7.71, p < .001$) and .83 ($t = 7.01, p < .001$), respectively. The correlation was $r = .79$ ($p < .001$). Additionally, as shown in Table 3, correlations between the slopes from incremental durations (Weeks 1–3, 1–4, etc.) and the full 20-week slope increased as each week was added ($r = −0.27$ to 1.00). Particularly, the correlation coefficients became statistically significant from Weeks 1 to 6 ($r = 0.32, p < .01$) and became equal to or greater than .50 from Weeks 1 to 8 ($r = 0.50, p < .001$). Furthermore, slope stability increased across weeks. *SE*s decreased over time: maximum value was .72 in Weeks 1–3, decreasing to 0.11 in Weeks 1–18 and Weeks 1–19 (see Table 3). In other words, as teachers collected more weekly data, the slopes became more stable.

Are Word Dictation Slopes Sensitive to Writing Growth?

To determine sensitivity to growth, we examined the minimum number of data points required to produce a slope significantly different from zero. Results (see Table 3) showed that only in a brief duration of three weeks, the slope was significantly larger than zero ($t = 3.07, p < .01$). However, the correlation coefficient between the slope from Weeks 1 to 3 and the slope from Weeks 1 to 20 was not statistically significant (the correlation coefficient was statistically significant and positive from Weeks 1 to 6 and became $r = .50$ from Weeks 1 to 8). Taken together, at least eight weeks of data collection are needed to produce a significant and sufficiently reliable slope.

Are the Relations Between Writing Growth and Demographic Characteristics Significant?

The baseline model confirmed considerable between-subject variability (ICC = .80), supporting multilevel modeling. Further, baseline model coefficients indicated that, on average, students scored 31.73 CLS at Week 1 and grew by 0.91 CLS each week. To investigate differences in intercepts and slopes depending on students' demographic characteristics, we identified estimates produced from each demographic group. Table 4 summarizes the predicted intercepts and slopes for each subgroup. Results indicated higher intercepts and slopes for female students, third/fourth

Table 5. Summary Results of Conditional Model.

Fixed effect	Estimate (*SE*)	*t* value	*p* value
For initial status			
Sex	2.45 (4.71)	0.52	0.61
Grade level	8.71 (4.49)	1.94	0.06
Race/ethnicity	−8.05 (4.45)	−1.81	0.07
FRL	−6.61 (5.50)	−1.20	0.23
ELL	−5.57 (5.83)	−0.96	0.34
For linear slope			
Sex	0.03 (0.23)	0.15	0.88
Grade level	−0.04 (0.23)	−0.16	0.87
Race/ethnicity	−0.28 (0.22)	−1.25	0.21
FRL	−0.42 (0.29)	−1.48	0.15
ELL	0.10 (0.28)	0.31	0.76

Note. ELL = English language learner; FRL = free/reduced-price lunch.

graders, White students, students not eligible for FRL, and students not eligible for ELL. These findings suggested a need to further examine whether the differences were statistically significant. Thus, we included each variable as a Level 2 covariate; results (see Table 5) indicate that none of the variables were significantly related to intercepts or slopes.

Discussion

We investigated technical features of slopes produced from word dictation and characteristics of beginning writers' growth. In the following sections, we interpret the major findings, describe study limitations, and discuss implications for research and practice.

Technical Features of Word Dictation Slopes

We found evidence of alternate-slope reliability, incremental slope reliability, increasing stability of slopes over time, and sensitivity to growth in a brief period. Additionally, based on the findings, we can confirm that at least eight weeks of data collection may be required before making an instructional decision using word dictation. These results are consistent with previous research reporting that the stability of CBM slopes increased with additional data points (Christ, 2006; Hintze & Christ, 2004; McMaster et al., 2011). With respect to the desired data collection period, our finding aligns with previous findings that at least eight data points were needed to determine reliable, stable weekly growth rates when using sentence-level CBM-W tasks (Filderman et al., 2019; McMaster et al., 2011) and more broadly, with previous CBM research recommending collecting data for anywhere from five to 14 weeks while administering assessments one to three times per week (Ardoin et al., 2013; Christ et al., 2012).

Our findings also underscore the promise of CLS scoring for elementary students with severe writing difficulties. The results related to sensitivity, where statistical significance was found, showed a different picture from Romig and Olsen's (2021) finding that CBM-W slopes did not significantly differ from zero. Despite various factors possibly contributing to these inconsistent findings (e.g., grade level, length of progress monitoring), one explanation may relate to the use of CLS, compared with WW, WSC, correct word sequences (CWS), and correct minus incorrect word sequences (CIWS) in Romig and Olsen (2021). CLS can capture very small changes in student progress, which likely allows teachers to detect growth in very short time periods. Also, our finding aligns with the literature suggesting that (a) CBM-Ws can capture writing growth more accurately as compared to traditional approaches to scoring writing (Allen et al., 2018), and (b) CLS can inform progress for these students (Hampton & Lembke, 2016).

Writing Growth Characteristics of Students With Intensive Learning Needs

Our finding that the writing skills of students improved over time is encouraging, but more research is needed. Students scored 31.73 CLS on average at the first week and gained 0.91 CLS per week. According to CBM-W benchmarks (McMaster et al., 2018), as of Fall, the initial score of 31.73 falls within the 10th to 25th percentile for first-grade students and below the 10th percentile for second- or third-grade students. We can estimate that the students would eventually score 49.02 CLS when 20 weeks of progress monitoring are completed, which still falls in the 10th to 25th percentile for first graders and below the 10th percentile for second or third graders. This finding aligns with research indicating that growth rates of students receiving special education services often do not approach growth of students in general education (Deno et al., 2001) and multiple years of intervention may be required for students with intensive learning needs (Al Otaiba & Fuchs, 2006). To optimize the impacts of support, educators could consider intensifying interventions based on word dictation progress data (Filderman & Toste, 2022).

Relation Between Demographic Characteristics and Word Dictation Intercepts and Slopes

Our results tentatively suggest that word dictation may have comparable accuracy across different student groups, which generally aligns with previous research indicating the psychometric adequacy of word dictation for particular student populations, including ELLs (Keller-Margulis et al., 2019; Smith & Lembke, 2022). Variance in intercept or slope might instead be expected to be associated with other factors such as teachers' DBI knowledge and skills (Bresina & McMaster, 2020), teachers' implementation fidelity (Kretlow & Bartholomew, 2010), students' initial level of academic skills (Parker et al., 2011), or students' attention and working memory capacity (Berninger & Amtmann, 2003). Still, given that our investigation is preliminary and based on a small sample, further investigation is needed. Regarding the effect of sex, our finding does not align with findings of McMaster et al. (2017) who reported a significant relation of students' gender with intercept and slope for CWS and CIWS; thus, the analysis should be replicated in other samples. Furthermore, patterns in the data identified in unconditional models did indicate potential subgroup differences, particularly suggesting gaps in performance favoring White, English-speaking students who are not eligible for FRL. Given that in CBM reading, studies reported differences in intercepts and slopes for demographic groups (Kranzler et al., 1999; Yeo et al., 2011), continued investigation is needed to ensure that word dictation or other types of CBM-Ws do not under/overestimate writing performance for some populations.

Study Limitations and Implications for Research and Practice

The following limitations should be considered when interpreting these findings. First, our student sample was composed of 79 elementary students. Future research should investigate the characteristics of word dictation slopes and students' growth in writing using larger sample sizes to increase generalizability. Another limitation relates to our analytic approach of examining the relation of students' race/ethnicity with slopes and intercepts. Given the limited sample size, we combined students with race/ethnicity other than White into a single group ("students from racially/ethnically minoritized backgrounds") for the statistical testing, meaning we failed to consider the heterogeneity within groups as well as intersecting identities. Thus, our results regarding the relation of student race/ethnicity to word dictation intercepts and slopes have little to do with any one of the specific race/ethnicity categories, and any further interpretation could be misleading. With larger samples of students, research is needed to pay attention to the writing growth of students with each different race/ethnicity reported. Future researchers must also avoid making a decontextualized comparison between racially minoritized students and White students as default. Additionally, by collapsing ELL status to a binary classification, we were unable to account for the diverse range of students' linguistic backgrounds, English language proficiency, home language(s), and the corresponding educational services they were receiving. Future research should better account for the diversity of ELLs' languages and backgrounds rather than treating them as a homogeneous group. Last, although students generally exhibited low writing skills, the similarity of initial writing skills among different groups during screening remains unclear. It is uncertain whether the groups showing higher slopes and intercepts in the unconditional models (non-FRL eligible students, non-ELL eligible students, female students, White students, third and fourth graders) had initially higher scores at screening. Future research should examine the writing proficiency of different groups during screening and consider including the initial proficiency as a covariate in statistical models to control for its potential influence.

Our findings have implications for the practical use of word dictation. First, the finding that word dictation slopes produce reliable, stable, and sensitive indicators of writing growth supports practitioners' use of this tool. Particularly, we recommend that teachers of elementary students with significant writing needs, with or without various disabilities, could monitor student progress using CLS. Also, given that collection of at least eight weekly data points was supported, our investigation adds evidence to practical guidelines for decision-making using word dictation data. In future studies, researchers may examine potential links between teacher- or system-level factors and students' writing progress, beyond student characteristics.

Conclusion

This study examined the technical features of word dictation, scored for CLS, for early elementary students receiving intensive writing intervention. Results indicated that word dictation is a useful indicator of student progress in intensive writing intervention, producing reliable slopes within eight weeks of weekly data collection. Future research is needed, however, to more conclusively determine whether variations in slope by students' demographic characteristics exist. Ultimately, the findings of this study indicate that word dictation can be used for accurate progress monitoring and effective instructional decision-making.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A170101 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.

ORCID iD

Seohyeon Choi https://orcid.org/0000-0003-1721-4956

Supplemental Material

Supplemental material for this article is available on the *Assessment for Effective Intervention* website with the online version of this article.

References

Allen, A. A., Poch, A. L., & Lembke, E. S. (2018). An exploration of alternative scoring methods using curriculum-based measurement in early writing. *Learning Disability Quarterly*, *41*(2), 85–99. https://doi.org/10.1177/07319487177254

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities*, *39*(5), 414–431. https://doi.org/10.1177/00222194060390050401

Al Otaiba, S., Rouse, A. G., & Baker, K. (2018). Elementary grade intervention approaches to treat specific learning disabilities, including dyslexia. *Language, Speech, and Hearing Services in Schools*, *49*(4), 829–842. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0022

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, *51*(1), 1–18. https://doi.org/10.1016/j.jsp.2012.09.004

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities*, *35*(1), 39–56. https://doi.org/10.1177/002221940203500104

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. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). Guilford.

Bresina, B. C., & McMaster, K. L. (2020). Exploring the relation between teacher factors and student growth in early writing. *Journal of Learning Disabilities*, *53*(4), 311–324. https://doi.org/10.1177/0022219420913543

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. https://doi.org/10.1080/02796015.2006.12088006

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children*, *78*(3), 356–373. https://doi.org/10.1177/001440291207800306

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219–232. https://doi.org/10.1177/001440298505200303

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, *30*(4), 507–524. https://doi.org/10.1080/02796015.2001.12086131

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

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Council for Exceptional Children.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. (1980). *Relationships among simple measures of written expression and performance on standardized achievement tests* (No. 20; p. 95). University of Minnesota Institute for Research on Learning Disabilities.

Filderman, M. J., Austin, C. R., & Toste, J. R. (2019). Data-based decision making for struggling readers in the secondary grades. *Intervention in School and Clinic*, *55*(1), 3–12. https://doi.org/10.1177/1053451219832991

Filderman, M. J., & Toste, J. R. (2022). Effects of varying levels of data use to intensify a multisyllabic word reading intervention for upper elementary students with or at risk for reading disabilities. *Journal of Learning Disabilities*, *55*(5), 393–407. https://doi.org/10.1177/00222194211048405

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*(2), 188–192. https://doi.org/10.1080/02796015.2004.12086241

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*(6), 488–500. https://doi.org/10.1177/001440299105700603

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991a). The contribution of skills analysis to curriculum-based measurement in spelling. *Exceptional Children*, *57*(5), 443–452. https://doi.org/10.1177/001440299105700507

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991b). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review*, *20*(1), 49–66. https://doi.org/10.1080/02796015.1991.12085532

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. https://doi.org/10.1080/10573569.2014.973984

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. https://doi.org/10.1080/02796015.2004.12086243

Jung, P.-G., McMaster, K. L., & delMas, R. C. (2017). Effects of early writing intervention delivered within a data-based instruction framework. *Exceptional Children*, *83*(3), 281–297. https://doi.org/10.1177/0014402916667586

Keller-Margulis, M. A., Ochs, S., Reid, E. K., Faith, E. L., & Schanding, G. T., Jr. (2019). Validity and diagnostic accuracy of early written expression screeners in kindergarten. *Journal of Psychoeducational Assessment*, *37*(5), 539–552. https://doi.org/10.1177/0734282918769978

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly*, *14*(3), 327–342. https://doi.org/10.1037/h0089012

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, *33*(4), 279–299. https://doi.org/10.1177/0888406410371643

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. https://doi.org/10.1177/073724770302800304

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. https://doi.org/10.1177/001440291107700203

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

McMaster, K. L., Kunkel, A., Shin, J., Jung, P.-G., & Lembke, E. (2018). Early writing intervention: A best evidence synthesis. *Journal of Learning Disabilities*, *51*(4), 363–380. https://doi.org/10.1177/0022219417708169

McMaster, K. L., Lembke, E. S., Shin, J., Poch, A., Smith, R. A., Jung, P., Allen, A., & Wagner, K. (2020). Supporting teachers' use of data-based instruction to improve students' early writing skills. *Journal of Educational Psychology*, *112*(1), 1–21. https://doi.org/10.1037/edu0000358

McMaster, K. L., Shin, J., Espin, C. A., Jung, P. G., Wayman, M. M., & Deno, S. L. (2017). Monitoring elementary students' writing progress using curriculum-based measures: Grade and gender differences. *Reading and Writing*, *30*(9), 2069–2091. https://doi.org/10.1007/s11145-017-9766-9

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards* [English language arts].

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. https://doi.org/10.1037/a0026833

Poch, A. L., Allen, A. A., & Lembke, E. S. (2019). Scoring measures of word dictation curriculum-based measurement in writing: Effects of incremental administration. *Psychology in the Schools*, *56*(5), 702–723. https://doi.org/10.1002/pits.22220

Puranik, C. S., & Lonigan, C. J. (2012). Early writing deficits in preschoolers with oral language difficulties. *Journal of Learning Disabilities*, *45*(2), 179–190. https://doi.org/10.1177/0022219411423423

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear model: Applications and data analysis methods* (2nd ed.). Sage.

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

Ritchey, K. D., Coker, D. L. Jr., & McCraw, S. B. (2010). A comparison of metrics for scoring beginning spelling. *Assessment for Effective Intervention*, *35*(2), 78–88. https://doi.org/10.1177/1534508409336087

Romig, J. E., & Olsen, A. A. (2021). Technical features of slopes for curriculum-based measures of secondary writing. *Reading & Writing Quarterly*, *37*(6), 535–551. https://doi.org/10.1080/10573569.2020.1860841

Smith, R. A., & Lembke, E. S. (2022). Technical adequacy of a spelling curriculum-based measure for English Language Learners in the first through third grades. *Learning Disability Quarterly*, *45*(2), 144–155. https://doi.org/10.1177/0731948720930423

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*(8), 795–819. https://doi.org/10.1002/pits.20113

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35–39. https://doi.org/10.1177/875647939000600106

Wanzek, J., & Vaughn, S. (2009). Students demonstrating persistent low response to reading intervention: Three case studies. *Learning Disabilities Research & Practice*, *24*(3), 151–163. https://doi.org/10.1111/j.1540-5826.2009.00289.x

Yeo, S., Fearrington, J., & Christ, T. J. (2011). An investigation of gender, income, and special education status bias on curriculum-based measurement slope in reading. *School Psychology Quarterly*, *26*(2), 119–130. https://doi.org/10.1037/a0023021