What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change?

Katherine A. Valentine

Adrea J. Truckenmiller

Gary A. Troia

Sydney Aldridge

Michigan State University

September 2021

Author Note: This research was supported in part by a grant from the Institute for Research on Teaching and Learning at Michigan State University and grant #R305A210061 from the U.S. Department of Education, Institute of Education Sciences to the second author and grant #R305A160049 from the U.S. Department of Education, Institute of Education Sciences, to the third author. The authors would like to acknowledge Greenflux, LLC for programming the Writing Architect.

Publication History: Received 23 April 2021; Received in revised form 25 August 2021; Accepted 17 September 2021; Available online 28 September 2021

Citation: Valentine, K. A., Truckenmiller, A. J., Troia, G. A., & Aldridge, S. (2021). What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change? *Assessing Writing*, *50*, 100567. https://doi.org/10.1016/j.asw.2021.100567

Abstract

Understanding students' progress towards meeting grade-level expectations within an academic year is the goal of many education stakeholders so that they can make decisions to adjust instruction and improve students' learning trajectories. The purpose of this study was to explore 4th and 5th grade students' progress towards meeting expectations in writing, using a curriculum- based measurement metric. Typed writing samples in narrative, persuasive, and passaged- based informational genres were collected in a one-year longitudinal design with four time points. Multilevel change models revealed evidence of quadratic change across the school year with significant change in the narrative genre but no statistically significant change in the other genres. Type of change (fan-spread growth or mastery learning) and growth based on initial transcription, spelling, and reading abilities were also explored.

Keywords: curriculum based measurement, written expression, annual change, genre

Understanding students' progress towards meeting grade-level expectations is the goal of many education stakeholders so that they can make decisions to adjust instruction and improve students' learning trajectories. The assessment framework curriculum-based measurement (CBM) has been used successfully to monitor change in reading and mathematics (Stecker et al., 2005), as well as early writing skills (McMaster et al., 2011). As an assessment technology, CBM has been favored to measure progress because assessment stimuli and scoring metrics can be adjusted to match community cultural expectations, are agnostic to specific curriculum content, are relatively efficient (i.e., quick, low-cost, easy-to-administer), demonstrate an acceptable level of technical adequacy for low to moderate-stakes decisions, and can be sensitive to change (Deno, 2003). However, detecting incremental change in writing and understanding what those changes imply has been challenging, particularly in grades when students are expected to compose extended responses in specific genres (e.g., narrative, persuasive, and informational papers with or without source texts as a prompt component) (Institute for Education Sciences, 2017). The purpose of this study is to more comprehensively explore change in writing in the grade levels that students are first administered high-stakes assessments in writing (grades 4 and 5) and in activities that more closely match many schools' expectations. Features of those activities include typed responses to narrative, persuasive, and passage-based informational prompts. It is important for increasing face validity that tasks more closely match intended expectations (American Educational Research Association, 2014). Furthermore, student performance is different depending on genre (Graham, et al., 2016; Romig et al., 2020) and typewritten transcription is different from handwritten transcription (Connelly et al., 2007).

We also highlight pressing issues of trajectory of change, the important role of where students begin the school year, and factors that influence students' initial writing performance in

grades 4 and 5. We begin by exploring the potential of CBM metrics as a tool for helping educators understand academic year change in writing.

Curriculum-Based Measurement

"CBM provides the tool by which teachers and schools can model students' trajectories of learning in a time frame that permits practitioners to use the data to tailor instructional programs, with the goal of optimizing student outcomes" (Fuchs, 2017, p. 6). To reach this goal of having change data that can inform instructional decisions that then leads to improved student outcomes, Fuchs (2004) suggests three stages of research. Stage 1 research examines various features of task administration (e.g., narrative, informational, or persuasive prompt) and different scoring metrics that best represent desired writing achievement, which may vary by grade level. In CBM-WE, limited work has been done in Stages 2 and 3 (Romig & Olsen, 2021). Stage 2 research comprises evaluation of the amount and type of change students make in the context of CBM-WE. Stage 3 research is the ultimate goal, which is evaluating the utility of CBM-WE to help educators adjust instruction when students are not making adequate progress.

Curriculum-Based Measurement in Written Expression

In Stage 1 research of CBM-WE, most often, the prompts are sentence starters or pictures for the narrative or expository genre and students have one minute to plan and three minutes to handwrite a response to the prompt (Romig et al., 2020). Adjustments to this administration format is needed to meet current writing expectations for grades 4 and beyond and to incorporate more recent research findings. Although some sentence starters are designed to elicit persuasive or informational responses, to our knowledge only one set of passage-based informational CBM-WE prompts have been explored and published that require a typed response (Truckenmiller et al., 2020). In addition to ensuring that narrative, persuasive, and passage-based informational

prompts are available in a digital (typing) environment, additional administration time (e.g., 7–15 min) is needed because of its association with increased criterion validity (McMaster & Campbell, 2008; Romig et al., 2020; Truckenmiller et al., 2020).

Research consensus on scoring metrics also has changed since the inception of CBM-WE. Originally, student responses were tallied for total words written (TWW), words spelled correctly, and number of correct writing sequences (CWS). Following three decades of development and evaluation, a meta-analysis shows that the metric Correct Minus Incorrect Writing Sequences (CIWS) has the highest criterion validity in grades 1 through 12 (Romig et al., 2017). The mean correlation coefficient of CIWS with criterion measures in grades three through five is.65, with a confidence interval of.54 to.73 and the correlation coefficients for TWW, WSC, and CWS ranging from.26 to.54 (Romig et al., 2017).

CIWS was originally introduced by Espin et al. (2000) as a metric to better evaluate the complexity of student writing in higher grade levels, and demonstrated utility for predicting performance on high-stakes state achievement tests (Espin et al., 2010). A CIWS score calculates the difference between the number of correct writing sequences in a writing sample and the number of incorrect writing sequences. A writing sequence is defined as the connection between two adjacent units of writing, such as words and punctuation. A correct writing sequence is one that adheres to correct spelling, punctuation, capitalization, syntax, and semantics in both of the adjacent units of writing. CIWS uniquely places more weight on key mechanisms of writing development: conventions of writing and sentence-level translation of ideas. Kim, Gatlin, Al Otaiba, and Wanzek (2018) recently explored the theoretical constructs of writing that are represented by CIWS. They empirically demonstrated that CIWS represent the consolidation of some transcription skills (i.e., handwriting and spelling) and some text generation (i.e., oral

language) skills that are necessary for writing quality. Their model of transcription and text generation explained 60–71% of the variance in CIWS, suggesting that transcription and text generation are the primary components of CIWS but there are also other unmeasured factors contributing to CIWS performance.

Measuring Change

Within a Multi-Tiered System of Support framework, understanding change is important for (a) monitoring student response to Tier 1 instruction at three to four time points per year and (b) more frequent monitoring of progress to detect students' response to interventions happening at Tiers 2, 3, and special education (Shapiro & Guard, 2014). Beyond use within MTSS and in the United States, Dockrell et al. (2018) highlight the power of CBM-WE as a formative assessment to tailor instruction over time and to compare progress in different genres of writing. In this study, we begin the exploration of change in CIWS by looking at typical change patterns across a school year in response to Tier 1 instruction (McMaster et al., 2011).

Some Stage 2 research studies have been conducted on CIWS. McMaster and Campbell (2008) found change on a narrative prompt of 0.70 CIWS per week and 0.89 CIWS per week on an expository prompt in grade 4. A major limitation of this study is that the statistical model constrains individual student data to assume the same shape and amount of intra-individual change over time, which is an untenable assumption (Singer & Willett, 2003). When allowing each student to have their own individual change trajectories in a multi-level model (MLM), Keller-Margulis , Mercer, Payan, and McGee (2015) found much lower estimates of change at 0.32 CIWS per week on a narrative prompt. They also found that change in writing is not necessarily linear, but rather varies by season. They found linear change in CIWS for younger writers (i.e., grades 2 through 4) which shifted to quadratic change in grade 5. The quadratic

pattern is a higher rate of improvement between the fall and winter and a slower rate from winter to spring.

MLMs also allow students to have their own intercept values (initial score in the fall), rather than constraining the intercept to the mean of the sample. This flexibility is necessary because of how widely students vary in their writing abilities at the beginning of the school year (Singer & Willett, 2003). The wide variability in initial performance on CBM-WE is important to explore because it has implications for the amount of change to expect (Codding et al., 2015) and represents an opportunity to understand what contributes to the variability of where students begin the school year. Change estimates are smaller for higher grade levels versus lower grade levels and for higher- versus lower-performing students (AIMSweb ROI Change Norms, 2012; McMaster & Campbell, 2008). This suggests an important general developmental trend of lower change rates for higher-performing and older students. This conclusion also is well supported in other studies that demonstrate decreasing amounts of change for higher-performing students and older students for more complex reading skills (Bloom et al., 2008; Scammacca, Fall, Capin, Roberts, & Swanson, 2020).

Grades 4 and 5 represent a developmental time period between the large growth in younger students and slower growth for older students and also a time when there is likely to be high inter-individual variability. Therefore, we are interested in determining if lower-performing students have a higher growth trajectory or if higher-performing students do. Meta-analyses in reading often show that a student's initial level of reading is related to how much that student grows (Huang et al., 2014; Pfost et al., 2014). Correlations between initial status (intercept) and change rates (slope) can be used to describe 3 types of change patterns. A strong positive correlation, known as fan-spread change, indicates students with higher initial performance (i.e.,

intercept) have a faster rate of change and that students with lower initial performance have a slower rate of change. On the other hand, strong negative correlations represent a slower rate of change for students with higher intercepts and a faster rate of change for those with lower initial intercepts, referred to as compensatory change. A correlation near zero indicates initial status has little bearing on rate of change. A meta-analysis of reading measures indicates that reading development can demonstrate both types of patterns depending on the skill measured, the age of the participants, and the time frame (e.g., one versus several school years; Pfost et al., 2014). CBM research also is mixed and may depend on age, skill, and time frame. In early elementary school, there is evidence of fan spread growth within a school year with 10-min written compositions (Wood, Schatschneider, & Wanzek, 2020) and there is evidence of compensatory growth within a school year in 3-min written compositions for middle schoolers (Codding et al., 2015).

Variables Related to Fall Performance and Change

We have established that initial performance on CBM-WE in the fall and change in CBM-WE varies. In the current study, we also seek to explore this variability in relation to different writing expectations or genres (narrative, persuasive, and passage-based informational writing) and student performance on components skills of writing (typing fluency, spelling, and word reading) that may impact their initial written composition performance (fall CIWS).

Change expectations and instruction should be different for different genres of writing. For example, change estimates are much smaller for expository (persuasive and informational) versus narrative genres (McMaster & Campbell, 2008). These differences translate into different instructional implications. For example, planning instruction for each genre requires a different strategy (De La Paz & Graham, 2002; Englert et al., 1988). Student performance in one genre is

only moderately related to performance in another genre, with correlations between informational, persuasive, and narrative genres ranging from 0.22 to 0.60 (Graham et al., 2016). Furthermore, instruction in the informational genre may lag behind that in other genres, as 83% of third and fourth grade teachers in a national survey reported being unprepared or minimally prepared to teach informational writing (Brindle et al., 2016).

In this study, we explore the role of foundational writing skills that are likely to influence students' initial CIWS performance in the fall of grades 4 and 5. In a study that sought to evaluate the developmental theoretical components of CIWS, Kim et al. (2018), found that CIWS captured components of the Not-So-Simple View of Writing (transcription, text generation, and executive functions within working memory constraints) relatively well. In the Not-So-Simple View of Writing (Berninger & Winn, 2006), children's writing develops from novice to expert as they become more fluent with spelling and handwriting/typing (transcription) and thus free cognitive resources (primarily working memory) to express their development of increasingly sophisticated language (text generation).

A meta-analysis of the component skills of written composition performance found that transcription (handwriting/typing fluency and spelling), oral language, and reading abilities influenced writing outcomes, with transcription and reading each accounting for 25% of the variance in writing outcomes, and oral language accounting for 10% of the variance (Kent & Wanzek, 2016). In the current study, we did not collect oral language measures, but we further evaluate the influence of transcription (typing fluency and spelling achievement) and printed word recognition on initial status and change during an academic year. Word-level reading was selected because it has been identified as the most consistent predictor of writing (Berninger et

al., 2010). Understanding the relation of component skill performance to change is important for determining some of the content components of intervention decisions (Berninger & May, 2011).

Purpose of This Study

Understanding the nature of change in writing can guide goal-setting and help to match students with the best content and instruction needed to meet those goals (Shapiro & Guard, 2014). In the current study, we sought to describe change in CIWS toward relevant grade level expectations in grades 4 and 5 (e.g., typing compositions using narrative, persuasive, and passage-based informational genres) to help researchers and educators understand the amount and type of change to expect. We explored the role of key characteristics of CBM-WE change through four research hypotheses. Based on previous research in CBM (Christ et al., 2010; Keller-Margulis et al., 2015) we hypothesized that (1) CIWS would demonstrate significant change across all genres and that (2) change would be better represented by a curvilinear rather than linear path. Similar to a previous study of narrative CBM (Codding et al., 2015), we expected that (3) students with lower initial CIWS would have a faster rate of change (a compensatory change pattern). Finally, given the influence of transcription (typing fluency and spelling) and reading on initial levels of writing, we hypothesized (4) these component skills would influence initial status of CIWS and we explored whether they moderated change trajectories within a school year.

We also controlled for two well-known factors influencing initial writing status, gender and grade level. Girls consistently outperform boys throughout schooling with an estimated average effect size of 0.43 on writing measures (Parr, 2010). However, differences in annual *change* due to gender has not been found in CBM-WE studies (Fearrington et al., 2014).

Method

Participants

Nine teachers from public schools in one school district in the Midwest volunteered their classrooms to participate in a larger study throughout the 2018–2019 school year. Data were collected from 92 students in fourth (n = 37) and fifth grade (n = 55) and demographics are summarized in Table 1.

Materials

Writing Architect 1.0

The Writing Architect is a web-based application that allows for group administration (on desktops, laptops, or netbooks with headphones) of writing prompts and a customized FileMaker scoring interface for human raters and computer tallying. In the web application, all instructions and passages (for informational prompts) were read aloud by a human voice recording. Students could follow along with the prompt or passage on their screen or on the provided paper copy. The web application gives students three minutes to plan their writing on blank paper and 15 min to write by typing into a textbox in the application. Students also had the option to end the planning period or writing period earlier if they were finished.

Four separate prompts in each of three genres (narrative, informational, and persuasive) were administered across the year. Each prompt was reviewed by a panel of experts for content validity. In a calibration study, the means for the four narrative prompts and the four persuasive prompts were all within the standard error of CIWS. The passage-based informational prompts were closely evaluated in a counter-balanced study design to determine the amount of variance explained by the different passages and the variance accounted for was less than 5%, indicating strong evidence for form equivalence (Truckenmiller et al., 2020). Narrative prompts took the

form of a story title and students were asked to "write a creative, fictional story—a make believe story—to match the title; write a story others will find interesting and enjoyable to read." Persuasive prompts posed a debate-style question and students were instructed to "write a persuasive essay that convinces readers to agree with your answer to the question." Informational prompts consisted of an informational passage (drawn from online informational text providers for children) and a passage-specific prompt. Passages were selected and modified based upon grade-level readability metrics (word count, Lexile, Flesch-Kincaid, and Coh-Metrix narrativity and syntactic simplicity score). For the prompts, students were instructed to "write an informative paper that will help others learn about the topic of the passage you read; be sure to use information from the article you just read to give reasons why it is important." Prompts for all genres also reminded students that papers should include a clear main idea, an introduction and conclusion, details to support their claims, and to follow the rules of writing.

Table 1

Demographic Characteristics of Participants by Grade

Characteristic		Grade 4		Grade 5	
		n	%	*n*	%
Race/Ethnicity	Asian	7	18.9	4	7.3
	Black	2	5.4	10	18.2
	Hispanic	0	0	3	5.5
	Indian American	2	5.4	3	5.5
	Multiracial/multi-ethnic	6	16.2	5	9.1
	Native American	0	0	0	0
	White	20	54.1	30	54.5
Sex	Male	16	43.2	32	58.2
	Female	21	56.8	23	41.8
Language Status	Multilingual	2	5.4	10	18.2
	Native English speaker	35	94.6	45	81.8
Disability Status	Disability	5	13.5	4	7.3
	No Disability	32	86.5	51	92.7

Note. $N = 92$ ($n = 37$ for Grade 4 and $n = 55$ for Grade 5).

Measures

CIWS

The CIWS metric was calculated for the final submission for each writing sample (at 15 min or after the student pressed submit) through the Writing Architect interface, which allows human scorers to identify correct and incorrect sequences in each sample. The program tallied the number of correct and incorrect writing sequences as rated by the human scorer, then calculated CIWS by subtracting the number of incorrect sequences from the number of correct sequences. Scoring was completed by undergraduate and graduate research assistants who were trained using a scoring manual (available at https://osf.io/tfvx2/). Research assistants demonstrated 95% agreement on three samples before scoring and calibrated their scoring every 100 samples to prevent scorer drift. Interscorer reliability was calculated using intraclass correlation for 10% of the samples at $r = .98$ for CIWS. All responses also were scored for TWW, which the computer automatically tallied as any set of letters separated by a space.

Component Skills of Writing

Typing fluency was assessed within the Writing Architect web application. Participants were instructed to type a paragraph as quickly and accurately as possible and the web application ended administration at 90 s. The paragraph was an extended version (147 words) of the Monroe Sherman (1966) handwritten paragraph copying task. The program captured the number of characters typed and a research assistant counted the number of errors and subtracted from the total (Graham et al., 1997). Interrater reliability for 10% of the sample was perfect, $r = 1.00$.

The Spelling subtest of the Test of Written Language-4 (Hammill & Larsen, 2009), which consists of a series of dictated sentences in which target words are embedded, was administered to evaluate spelling ability. Internal consistency reported in the technical manual is high (α = .90 to.93).

The Reading subtest of the Wide Range Achievement Test-3 (Wilkinson, 1993) was administered to measure students' word recognition ability on a list of words. Internal consistency reliabilities reported in the technical manual for grades 4 and 5 are high (α = .88 to .90).

Design and Analysis

A longitudinal design with four time points (October/November, January/February, March/April, and May/June) was used. At each timepoint, probes for each of the three genres were group-administered via the web during English/Language Arts class time. Prompts were randomly assigned to individual students by the researchers at each time point. Because the administration dates varied among classrooms and with unequal intervals between time points, multilevel modeling (MLM) was chosen instead of latent change curve models, which require the time points to be the same for all students (O'Connell et al., 2013). Time (level 1) is nested within students (level 2) and is represented by the number of calendar days since the first assessment date for each student. Students were nested within teachers (level 3) in an 'empty MLM' to determine if the teacher level should be included. Given the likelihood of missing data in a longitudinal school-based study, patterns of missingness analyses were conducted prior to estimating the MLM (McCoach, 2018). Statistics were obtained using R 4.0.0 and the R package nlme (v3.1–148; Pinheiro et al. , 2020). Fit statistics were used to compare subsequent models.

To examine the pattern of change, a correlation between intercept and slope was computed using Pearson's *r.* Large positive correlation coefficients are evidence of a fan-spread change pattern. Large negative correlation coefficients indicate an inverse relationship with a slower rate of change for students with a higher intercept (compensatory change). A correlation coefficient nearer to zero indicates students are growing at similar rates regardless of their initial performance (temporal stability).

Results

Due to absences, there were 88 missing data points out of 1092 possible data points. Little's MCAR Test indicated data were Missing Completely at Random for the narrative genre, $_{\chi}^{2}(17) = 10.32$, $p = .89$, for the informational genre, $_{\chi}^{2}(23) = 19.92$, $p = .65$, and for the persuasive genre, $\chi^2(14) = 12.34$, $p = .58$. Therefore, maximum likelihood estimation was deemed appropriate and used in the MLM. Assumptions of normality (skewness and kurtosis values were all under 1.96) and homogeneity of variance, $F(1, 339) = .08$, $p = .77$, were met. Descriptive statistics for each genre at each time point are in Table 2.

Table 2

Descriptive Statistics for CIWS Scores by Genre and Time Point

Measure	M (SD)	*n*
Narrative CIWS		
Time 1	96.72 (71.27)	88
Time 2	85.72 (70.52)	87
Time 3	83.74 (68.81)	84
Time 4	92.29 (68.81)	82
Informational CIWS		
Time 1	62.42 (59.67)	86
Time 2	63.87 (53.67)	84
Time 3	71.41 (64.86)	81
Time 4	76.95 (71.51)	78
Persuasive CIWS		

Time 1	63.21 (48.21)	90
Time 2	76.68 (58.94)	88
Time 3	66.22 (54.94)	80
Time 4	71.64 (64.34)	76
Typing Fluency	110.18 (42.14)	92
WRAT-3 Reading	109.55 (13.19)	92
TOWL-4 Spelling	10.53 (3.09)	92
*TOWL-4 Vocabulary	10.76 (3.13)	92

Note. TOWL-4 = Test of Written Language, Fourth Edition; WRAT-3 = Wide Range Achievement Test, Third Edition

* The Vocabulary subtest was not included in any models because it was not a hypothesized contributing variable. However, the descriptive statistics are presented here as a way to further describe the language achievement of the sample

Two additional sets of analyses were run to verify two other necessary assumptions in this study. The descriptive statistics suggested that intercepts and slopes would likely be different for each genre. A MLM of the entire data set with a dummy code for each genre confirmed statistically significant differences between genre intercepts and between genre slopes (see fixed effects of genre in Supplemental Materials, Table S1). Given the unexpected lack of change in the narrative means (compared to other studies), we were concerned that this might be due to CIWS not detecting change rather than to no change occurring. Therefore, we also ran all models with TWW as a dependent variable because, although TWW has lower validity coefficients, it is often more sensitive to change (McMaster & Campbell, 2008). Correlations indicated that CIWS and TWW scores were highly related ($r = .90$, $p < .001$). MLM results were remarkably similar between the two metrics across all genres and research hypotheses. Tables with TWW results are presented in Supplemental Online Materials.

Unconditional Change Models

An unconditional means model with the CIWS score as the dependent variable was run to determine the amount of variance between students (Level 2) and between teachers (Level 3). The intraclass correlation (ICC) at Level 2 was calculated at.72,.71, and.60 for narrative, informational, and persuasive genres, respectively. This indicates a high level of variance to be explained between students. Surprisingly, the ICC for Level 3 revealed 1% of the variance to be explained at the teacher level for each genre; thus, a two-level model (time nested within students) was used instead of three levels (time nested within students nested within teachers). Subsequent models were run with control variables, fixed and random effects, and hypothesized variables. MLMs for narrative, informational, and persuasive genres are summarized in Tables 3–5. Fig. 1 displays a scatter plot with the best-fit change trends of each genre.

Narrative Change Models

For the narrative genre, allowing the linear slope to vary randomly explained little additional variance, therefore a model with a random intercept and fixed slope was selected (see model parameters and fit indices in Table 3). Including a quadratic slope did yield the best fit model (Model 4), though allowing for a random quadratic slope (Model 5) did not lead to a significant change in fit. Decreasing AIC and BIC criteria and a likelihood ratio test between Model 3 and Model 4, $\chi^2(1) = 4.26$, $p = 0.039$, confirmed that Model 4 was the best fitting model. Predictors of initial level of CIWS were added to Model 4 which resulted in improved fit for Model 6. Therefore, Model 6 is the final model. On average, students began at 66 CIWS with more fluent typists (performing one standard deviation higher than the mean on typing fluency) writing 39 more CIWS and strong spellers writing 20 more CIWS than average students. Girls did not write significantly more CIWS than boys and grade 5 students did not write significantly

more CIWS than grade 4 students. Average change for narrative CIWS showed significant regression to the mean (reduction of 3 CIWS per day) from fall to winter (linear slope) and then significant acceleration of CIWS from winter to the end of the year. Although grade 5 students had significantly less regression to the mean, no other variables significantly moderated the change trajectory.

Table 3
Multilevel Model Estimates for CIWS Change in Narrative Writing

Estimate	Parameter	Model 2	Model 4	Model 5	Final Model (Model 4 with predictors)
Initial status			Fixed effects		
Intercept	γ_{00}	-4.15 (64.25)	5.60 (64.42)	30.57 (63.72)	65.75 (52.57)
Grade	γ_{01}	18.77 (13.63)	17.86 (13.65)	12.84 (13.51)	3.62 (10.92)
Gender	γ_{02}	18.64 (13.41)	18.34 (13.43)	14.31 (13.29)	8.34 (10.46)
Typing fluency	γ_{03}				38.68** (6.39)
Spelling	γ_{04}				19.64* (9.07)
Word reading	γ_{05}				9.29 (9.37)
Linear Rate of Change	γ_{10}	-0.03 (0.03)	-0.31**(0.001)	-0.29**(0.10)	-2.85** (1.00)
Grade	γ_{11}				0.54**(0.21)
Gender	γ_{12}				0.14 (0.20)
Typing Fluency	γ_{13}				-0.14 (0.12)
Spelling	γ_{14}				-0.18 (0.17)
Word reading	γ_{15}				0.06 (0.18)
Quadratic Rate of Change	γ_{20}		0.002**(0.001)	0.001* (0.001)	0.01* (0.01)
Grade	γ_{21}				-0.002 (0.001)
Gender	γ_{22}				0.0001 (0.001)
Typing Fluency	γ_{23}				0.001 (0.001)
Spelling	γ_{24}				0.001 (0.001)
Word reading	γ_{25}				-0.001 (0.001)
Variance components			Random Effects		
Level 1	σ^2_{ε}	1398.85	1352.99	1280.38	1228.35
Level 2	σ^2_{00}	3559.78	3570.58	3223.56	1009.71
Linear	σ^2_{11}			0.11	
Covariance	σ^2_{01}			0.22	
			Goodness-of-fit		
Likelihood Ratio		4.45	5.77*	2.42	121.52
p-value		0.22	0.02	0.30	<.0001
Log likelihood		-1825.03	-1820.84	-1819.63	-1760.08
AIC		3662.06	3655.68	3657.25	3560.15
BIC		3685.05	3682.50	3691.74	3636.79

Note. Standard errors are in parentheses. * $p < .05$; ** $p < .01$

Informational Change Models

For the informational genre, the change model with a random intercept, random linear slope, and random quadratic slope (Model 6) provided the best fit model (see Table 4). Hypothesized moderation variables of initial level of CIWS were added in Model 7 and model fit further improved. On average, students began the year writing 66 CIWS on informational essays with fluent typists (performing one standard deviation higher than the mean on typing fluency) writing 22 more CIWS. Girls did not write significantly more CIWS than boys and grade 5 students did not write significantly more CIWS than grade 4 students. Average change for informational CIWS was not significantly different from 0, although girls experienced significantly less regression to the mean (linear slope) than boys. No other variables explained the random variation in linear or quadratic slopes.

Table 4

Multilevel Model Estimates for CIWS Change in Informational Writing

Estimate	Parameter	Model 2	Model 5	Model 6	Model 7
Initial status			Fixed effects		
Intercept	γ_{00}	23.33 (53.01)	34.89 (52.55)	31.54 (51.86)	66.33 (52.07)
Grade	γ_{01}	6.51 (11.26)	4.54 (11.13)	4.93 (10.99)	-2.38 (10.82)
Gender	γ_{02}	15.62 (11.09)	13.84 (10.97)	17.02 (10.83)	-5.27 (10.37)
Typing fluency	γ_{03}				21.53** (6.39)
Spelling	γ_{04}				10.89 (8.98)
Word reading	γ_{05}				10.71 (9.23)
Linear rate of change	γ_{10}	0.07* (0.03)	-0.02 (0.10)	-0.01 (0.10)	-0.54 (1.03)
Grade	γ_{11}				0.09 (0.21)
Gender	γ_{12}				0.47* (0.20)
Typing fluency	γ_{13}				-0.12 (0.12)
Spelling	γ_{14}				0.03 (0.17)
Word reading	γ_{15}				-0.16 (0.18)
Quadratic rate of change	γ_{20}		0.001 (0.001)	0.0004 (0.001)	-0.001 (0.01)
Grade	γ_{21}				0.0004 (0.001)
Gender	γ_{22}				-0.001 (0.001)
Typing fluency	γ_{23}				0.001 (0.001)
Spelling	γ_{24}				-0.0001 (0.001)
Word reading	γ_{25}				0.001 (0.001)
Variance components			Random Effects		
Level 1	σ^2_{ε}	1515.05	1173.58	834.48	793.49
Level 2	σ^2_{0}	2235.74	2556.77	2770.15	1363.14
Linear	σ^2_{11}		0.06	0.55	0.23
Quadratic	σ^2_{22}			0.004	0.00001
Covariance	σ^2_{01}		-0.35	-0.25	0.09
Covariance	σ^2_{02}			0.09	-0.21
			Goodness-of-fit		
Likelihood Ratio		6.99	11.13**	17.66**	73.24
p-value		0.07	0.004	0.001	<.0001
Log likelihood		-1754.02	-1747.81	-1738.98	-1702.36
AIC		3520.04	3513.62	3501.95	3454.72
BIC		3542.82	3547.78	3547.51	3549.62

Note. Standard errors are in parentheses. * $p < .05$; ** $p < .01$

Persuasive Change Models

In the persuasive genre, Model 6, with a random intercept, random linear slope, and random curvature, was determined to have the best fit. In this genre, the average intercept is uninterpretable because there are large statistically significant and practically large differences on four explanatory variables. Girls, on average, wrote 24 more CIWS and grade 5 students wrote 31 more CIWS in the fall. Fluent typists wrote 14 more CIWS and strong spellers wrote 21 more CIWS. Average change for persuasive CIWS was not significantly different from 0 and no other variables explained the random variation in linear or quadratic slopes.

Table 5
Multilevel Model Estimates for CIWS Change in Persuasive Writing

Estimate	Parameter	Model 2	Model 5	Model 6	Model 7
Initial status			Fixed Effects		
Intercept	γ_{00}	-47.86 (48.39)	-31.22 (47.16)	-37.81 (46.95)	-94.25* (39.24)
Grade	γ_{01}	22.49* (10.25)	18.25 (9.98)	19.78* (9.95)	30.91** (8.13)
Gender	γ_{02}	24.16* (10.07)	25.87** (9.80)	24.83* (9.76)	23.55** (7.70)
Typing fluency	γ_{03}				13.70** (4.78)
Spelling	γ_{04}				20.66** (6.54)
Word reading	γ_{05}				3.60 (6.78)
Linear rate of change	γ_{10}	0.01 (0.03)	0.12 (0.08)	0.12 (0.08)	1.21 (0.86)
Grade	γ_{11}				-0.27 (0.18)
Gender	γ_{12}				0.17 (0.17)
Typing fluency	γ_{13}				0.06 (0.10)
Spelling	γ_{14}				-0.27 (0.14)
Word reading	γ_{15}				0.23 (0.15)
Quadratic rate of change	γ_{20}		-0.001 (0.0004)	-0.001 (0.0005)	-0.01 (0.005)
Grade	γ_{21}				0.002 (0.001)
Gender	γ_{22}				-0.001 (0.001)
Typing fluency	γ_{23}				0.0002 (0.001)
Spelling	γ_{24}				0.001 (0.001)
Word reading	γ_{25}				-0.001 (0.001)
Variance components			Random Effects		
Level 1	σ^2_{ε}	921.14	851.83	640.85	619.73
Level 2	σ^2_0	1977.96	1591.29	1649.08	617.04
Linear	σ^2_{11}		0.10	0.38	0.12
Quadratic	σ^2_{22}			0.003	0.00001
Covariance	σ^2_{01}		0.47	0.16	0.37
Covariance	σ^2_{02}			-0.05	-0.30
			Goodness-of-fit		
Likelihood Ratio		8.94*	6.81*	13.31**	90.41
p-value		0.03	0.03	0.004	<.0001
Log Likelihood		-1712.78	-1708.62	-1701.96	-1656.76
AIC		3437.56	3435.23	3427.92	3363.51
BIC		3460.42	3469.53	3473.66	3458.79

Note. Standard errors are in parentheses. * $p < .05$. ** $p < .01$.

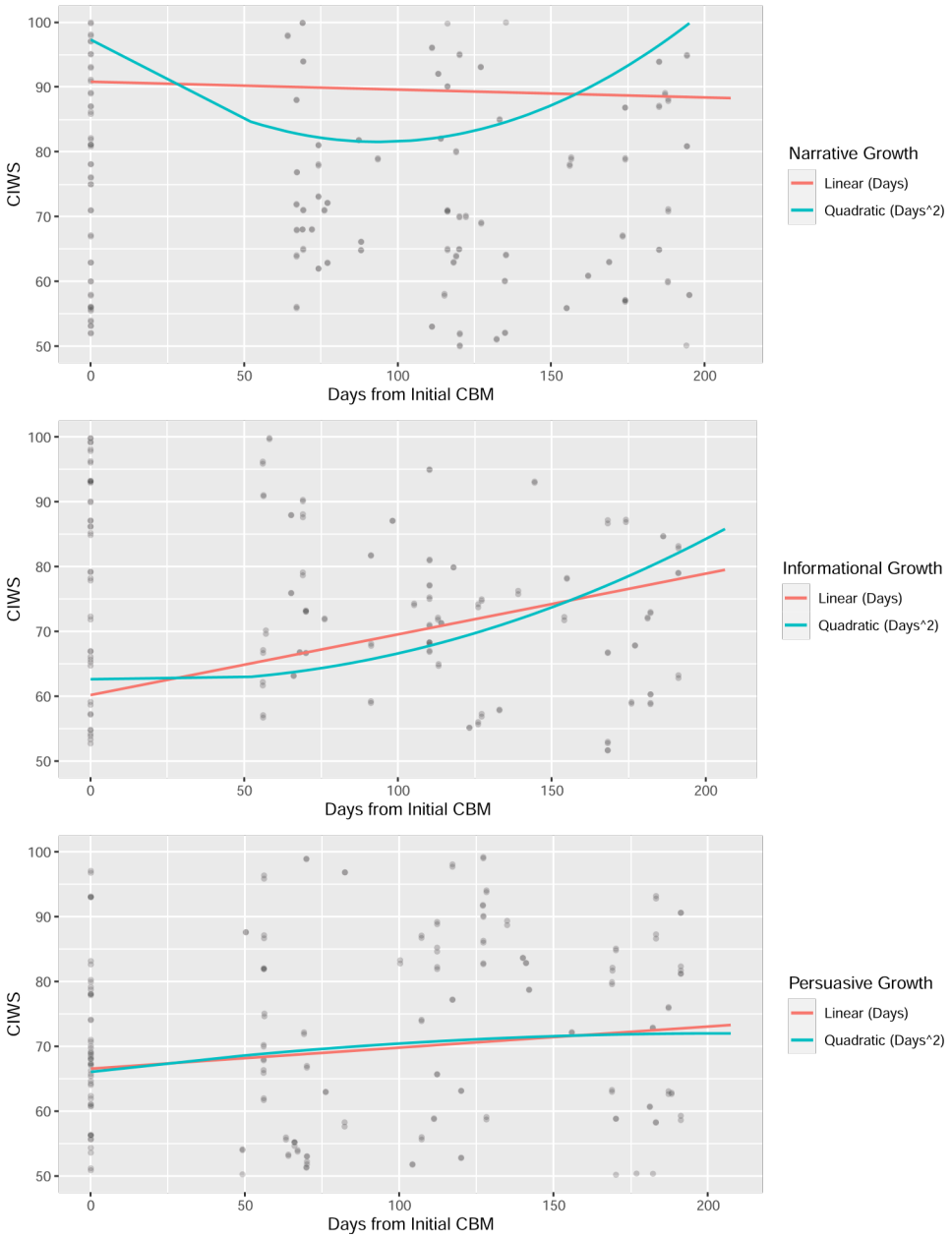

Fig. 1. Change in writing across a school year for narrative, informational, and persuasive genres.

Change Pattern

For research hypothesis 3, we examined the pattern of change between students through a correlation between the intercept and the slope. For the best-fitting models, the correlation between the intercept and quadratic curvature was found to be $r = .35$ for narrative, $r = .11$ for informational, and $r = .18$ for persuasive. In each genre, this correlation is interpreted as the

correlation between the average initial score of students and the amount of acceleration from winter to the end of the school year. The moderate positive correlation coefficient for narrative writing indicates that students with a higher initial score have slightly more acceleration in their growth. The small positive correlations in the other two genres suggest that initial status has little bearing on rate of change.

Discussion

Educators and researchers want to understand the nature of change in writing to guide goal-setting and matching students with the level and content of instruction needed to meet those goals (Shapiro & Guard, 2014). The research in writing lags behind other academic skills for understanding which tools have the highest likelihood of utility (Stage 1 research), understanding the amount of growth to expect (Stage 2 research), and ways to use CBM metrics to adjust instruction when students are not making adequate growth (Stage 3 research; see Fuchs, 2004).

The current study was conducted to understand CIWS change within a school year (Stage 2) for grades 4 and 5 and factors that may influence measurement of student change. This is the first study to explore CIWS sensitivity to growth with a task that more closely mirrors school expectations, which includes longer administration times, digital administration, and passage-based informational writing in addition to narrative and persuasive genres. A previous Stage 1 study of these features demonstrated incrementally higher validity coefficients with valued writing outcomes (i.e., writing quality ratings, state writing test; see Truckenmiller et al., 2020). We describe the complexities of CBM-WE for measuring change and implications for Tier 1 decisions.

Sensitivity to Change

Similar to other studies of narrative writing (e.g., McMaster & Campbell, 2008), CIWS in our narrative writing task was sensitive to change during the school year for grades 4 and 5, but the amount of growth was much smaller than expected. Slope estimates in the other two genres were not significant, which is contrary to other studies that have found significant growth from fall to spring in expository CIWS (Dockrell et al., 2018; McMaster & Campbell, 2008). We explore two potential conclusions from these findings: (a) there may be a trade-off between incrementally higher amounts of validity and less sensitivity to growth, and (b) general growth in written expression may begin to level off in grades 4 and 5.

To increase validity coefficients of CBM-WE in later elementary grades, we extended the administration time beyond most CBM-WE (3 min for planning instead of 1 min, and 15 min for writing instead of 3–10 min). This resulted in adequate levels of construct validity (coefficients greater than.60) for grades 5 through 8, but not grade 3 (Truckenmiller et al., 2020). Construct validity is higher because the construct measured is closer to a written composition achievement measure, rather than a written composition fluency measure, which is what CIWS was found to represent (Kim et al., 2018). Fluency is a useful construct to measure in elementary grade levels because fluency measures detect growth for skills that are rapidly developing and fluency measures have high concurrent validity with achievement in elementary grades (Lembke et al., 2013; Ritchey et al., 2016). We believe that the longer administration time more resembles achievement than fluency and thus gives up some of the sensitivity to growth that is associated with fluency measures. There is evidence to suggest that the shift from fluency measures is appropriate for late elementary grades because the growth occurring for students at Tier 1 is across grade levels instead of within grade levels.

More recent studies suggest that very little growth in writing occurs within an academic year beyond third grade. A recent study of students' writing change from fall to spring demonstrated significant growth in grades 1, 2, and 3 (Wood, Schatschneider, & Hart, 2020), but there was no change in grades 4 or 5 and some later grade levels seemed to show a decline from fall to spring. Wood and colleagues demonstrated this same pattern in a variety of metrics that are components of CIWS: total words written, number of spelling errors, and number of grammar errors. These authors suggest that productivity (TWW) and accuracy (spelling errors and grammar errors) may not be important indicators of growth as students become more proficient, but rather growth in writing *complexity* is more important. When compared to productivity and accuracy metrics, another study has shown that complexity metrics are more likely to detect differences between grades 4 and 6 (Troia et al., 2020).

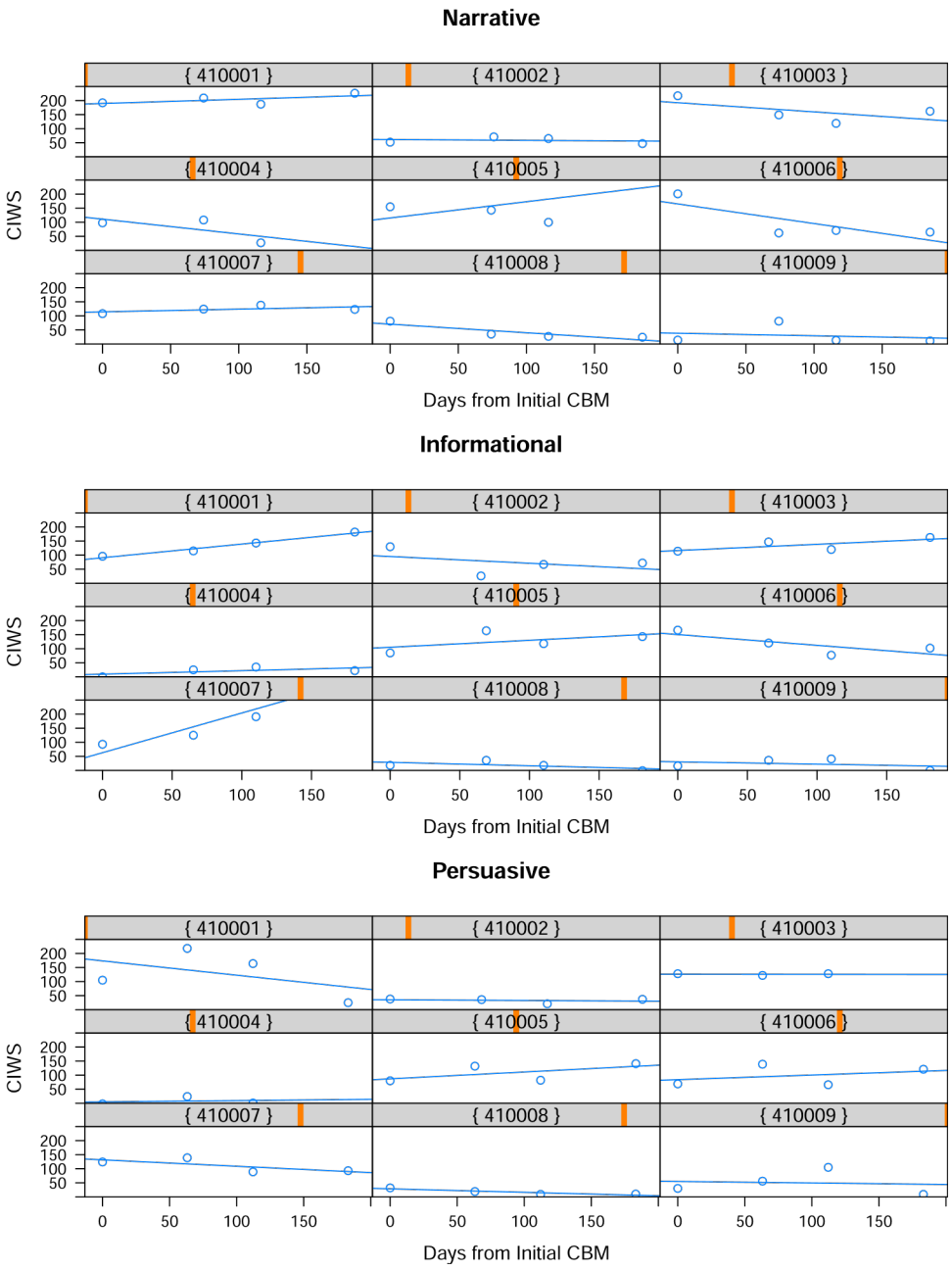

Fig. 2. A sampling of individual CIWS scores by genre. Note. The first nine students' CIWS scores across the school year are displayed as an example of the variety of scores and slopes among individuals.

Patterns of Writing Change

Normative data for CBM-WE (e.g., AIMSweb) include the unstated assumptions that all students have similar paths and that growth from fall to spring is linear. However, the large standard deviations in this sample (see Table 2) and the large variances around growth

trajectories (see Tables 3, 4, and 5) imply that there are major individual differences around mean values. We illustrate how misleading fixed effects means are in Fig. 2. In this figure, one can see students in our sample having every combination of widely varying initial status, and both large and small positive and negative change. By using MLM, the fixed and random effects capture the general characteristics of growth for both the group as a whole and for the individuals within the group (Curran et al., 2010; Singer & Willett, 2003). MLM also allows for change to be curvilinear as well as linear.

Similar to other studies using MLM (Keller-Margulis et al., 2015), we found that models allowing random variation and quadratic change fit best. Though best-fit models for each genre did include a quadratic curvature, only the narrative genre was sensitive to statistically significant amounts of quadratic growth. For narrative writing, Keller Margulis and colleagues found that students' CIWS increased from fall to winter and then slowly decreased from winter to spring. We found an opposite U-shaped pattern with initially declining scores from fall to winter and then acceleration towards the end of the school year. CBM researchers have described this growth as seasonal differences (Keller-Margulis et al., 2015). Seasonal differences may be due to systematic factors like spacing of instruction. For example, in the current sample, the classrooms mostly started with a narrative writing unit in the fall, followed by informative, then persuasive, and then response to literature, and finally perhaps another writing unit that may have included poetry or another type of informational writing experience. Further research is being conducted that includes observation of writing instruction to determine if these seasonal changes are due to instruction. This research will be helpful to illuminate the level of instruction needed to effect growth in writing. It is likely that typical general education writing instruction results in small

amounts of growth within a school year compared to interventions that have larger effect sizes on writing achievement (Graham & Perin, 2007; Graham et al., 2012).

Initial Writing Factors

Another developmental consideration is that students come to grades 4 and 5 with wide variability in their writing instruction, practice, and skill development, particularly with informational and persuasive genres (Brindle et al., 2016). The wide variability is evident in our sample and has implications for how much they change across a school year. We begin with narrative writing. The correlation between the intercept and the quadratic slope is positive with a small-to-moderate magnitude. This suggests that those students with higher starting points have steeper, more negative slopes to the 2nd data point, but then become more positive at a faster rate after the 2nd data point. Their quadratic curves are deeper U-shapes, and though slopes away from the minimum are steeper (more negative), they change at a faster rate and become positive more quickly. It also suggests that students with lower initial CIWS have shallower U-shapes with less decrease at the beginning and less increase later. This pattern suggests that higher performing students have greater degrees of change. This finding is similar to another study that measures growth from fall to spring in second grade (Wood, Schatschneider, & Wanzek, 2020), but differs from another study of grades 7 and 8 where investigators found that students who initially perform more poorly had steeper slopes than the students who were higher performers (a compensatory learning profile; Codding et al., 2015). Our study is the first to our knowledge to evaluate this type of relationship with a quadratic change curve. Because growth occurs so slowly, we believe that future research should examine this pattern across time spans greater than one academic year and in response to an intervention (Stage 3 research) to better understand if higher performing students tend to grow more or if lower performing students catch up.

The slow-moving nature of writing within a school year suggests that educators should attend more to where a student begins the year as a more informative piece of information to guide instructional decisions. This conclusion is supported by other CBM research demonstrating that intercept values have more utility for instructional decisions than slope values (Codding et al., 2015; Keller-Margulis et al., 2015). Remarkably, students within a grade level differed from each other as much as they differed from students in the other grade level. This suggests that educators in late elementary grades really need information to understand the factors that contribute to where students begin the year so that they can differentiate instruction. This study further elucidates some areas educators should attend to: genre, transcription fluency, spelling, and reading.

Initial status and change patterns were significantly different across the three genres, with narrative writing being the most privileged. This aligns with previous studies showing that students write more text and write more accurately in the narrative genre (Dockrell et al., 2018) and suggesting that students transfer some, but not much, of their writing skills between genres (Graham et al., 2016). Specifically, in order to generate ideas, students need to be taught the persuasive and informational text structures and planning strategies that are different from narrative (Troia & Graham, 2002).

In addition to idea generation, students need sufficient skills with transcription fluency, spelling, and reading to facilitate their writing development (Kent & Wanzek, 2016). Our results provide evidence that typing fluency and spelling were key determinants of writing performance regardless of genre and thus could be targets for instruction that would improve overall writing. The one exception is spelling in the informational genre. Spelling may not have been statistically significant because students could refer back to the passage to obtain the correct spelling of

words that were in the passage. Word reading did not significantly influence writing. This may be due to the specific sample having sufficiently mastered word reading (over 80% of the sample scored above the national standard score of 100).

Limitations and Direction for Future Research

As with any study involving the dynamics of students and schools, this study is not without limitations. Introducing passages as part of the informational prompt creates the opportunity for construct-irrelevant error to cause scores to be different between time points. We previously explored the amount of error between informational forms in a counterbalanced design (Santi et al., 2015) and found that differences in CIWS due to form was less than 5% (Truckenmiller et al., 2020). Reading abilities were mediated by allowing students to listen to the passage and all instructions. Further evidence in the current study suggests that decoding abilities did not play a significant role in students' initial status or change in CIWS. Therefore, the fluctuation in scores from time point to time point is likely due to unsystematic error.

Of particular interest with this sample are the large ICCs (>.7) calculated at the student level indicating that about 70% of variation is due to between-student differences, with the remaining 30% due to within-student differences. In an integrative data analysis of reading and mathematics, Hedges and Hedberg (2007) found the average ICC in grades 4 and 5 to be .25 in reading and .22 in mathematics. The ICCs found in this study are substantially larger, indicating that there is much more that we need to understand about between-student variation that will provide insight about writing development and instruction. Future research should include additional measures of student characteristics that have not been explored as thoroughly, such as knowledge about writing and motivation to write in varied genres. Explaining this additional

variance at the student level would simultaneously allow for better estimation of performance and provide more precise information for schools to understand their students' needs.

Finally, the estimation of within-year change without the context of a specific curriculum or intervention is probably ill advised. The ICC at the teacher level in this study was less than 1%, indicating that there was not much variation between the classrooms. This estimation was possibly limited due to the small number of teachers (n = 9). Either way, these results provide further evidence that future research needs to occur within the context of specific intervention and suggests that the progress monitoring power of CBM (and CIWS specifically) may be best suited to detecting changes due to a specified instructional program or intervention.

The external validity of this study is limited due to the high-performing nature of the students' reading abilities. Although, the sample's spelling and vocabulary abilities had means and standard deviations similar to national norms, in which the national scaled score mean = 10 and standard deviation = 3. Similar to most studies of writing, it is difficult to generalize our findings because the instructional environments are not described and the schools come from adjacent geographic locations.

Conclusions and Implications for Practice

In grades 4 and 5, students vary widely in their writing performance making it difficult to identify a good Tier 1 general outcome metric. There may be many students who are more similar to earlier elementary students where fluency metrics work well as a general outcome measure for writing. There also may be many students who begin the grade level with higher proficiency and therefore complexity metrics may be more appropriate for them. It also is likely that growth is slower or stretched over periods of time longer than an academic year, similar to more complex reading skill development observed in later grades (Bloom et al., 2008;

Scammacca et al., 2020). Regardless, it is evident that, if educators want to detect significant growth in writing, educators need to implement instruction that demonstrates significant impact on writing performance.

Based on the results of our study, we recommend implementing genre-specific text structure instruction (e.g., De La Paz & Graham, 2002; Englert et al., 1988; Troia & Graham, 2002) and genre-agnostic opportunities (typing fluency and spelling) for improving writing at Tier 1. For detecting student change, we recommend that schools use a reliable and valid pretest, instruction that has been shown to be effective (Graham & Perin, 2007; Graham et al., 2012), and a posttest. More frequent progress monitoring may not be needed to detect if there was or was not an impact of instruction on student performance. Typically, CBMs are used to monitor students' progress in response to a more intensive intervention (Tier 2, Tier 3, or special education) and progress is measured weekly or biweekly. This frequency of progress monitoring is likely not feasible with the length of the prompts described here because they are too time-intensive for administration and scoring to be administered on a weekly basis. Our study also suggests that there may not be enough change in student performance to detect on a weekly basis. More research with instruction included is needed to determine if more frequent progress monitoring will provide useful information. A 1997 paper by Graham and Harris sums up the results of this study well with the title: *It Can Be Taught, But It Does Not Develop Naturally* (Graham and Harris, 1997).

References

AIMSweb ROI Growth Norms. (2012). Pearson Education, Inc.

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

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In *Handbook of writing research*. (pp. 96–114). The Guilford Press.

Berninger, V. W., Abbott, R. D., Lee, S. H., Lovitt, D., Pam, T., Lin, S.J. (Cindy), Gould, L., Youngstrom, M., Shirley, S., & Amtmann, D. (2010). Relationship of word- and sentence-level working memory to reading and writing in second, fourth, and sixth grade. *Language, Speech, and Hearing Services in Schools, 41*(2), 179–193. https://doi.org/10.1044/0161-1461(2009/08-0002).

Berninger, V. W., & May, M. O. (2011). Evidence-Based Diagnosis and Treatment for Specific Learning Disabilities Involving Impairments in Written and/or Oral Language. *Journal of Learning Disabilities, 44*(2), 167–183. https://doi.org/10.1177/0022219410391189.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*, 289 –328. https://doi.org/10.1080/19345740802400072

Brindle, M., Graham, S., Harris, K. R., & Hebert, M. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing*, *29*(5), 929–954. https://doi.org/10.1007/s11145-015-9604-x

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39*(3), 447–462. https://doi.org/10.1080/02796015.2010.12087765

Codding, R. S., Petscher, Y., & Truckenmiller, A. J. (2015). Middle-school CBM predictors: Examining the contributions of writing in addition to reading and mathematics. Journal of Educational Psychology, 107(2), 437-450. http://dx.doi.org/10.1037/a0037520

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboarded and handwritten compositions and the relationship with transcription speed. *British Journal of Educational Psychology, 77*, 479–492. https://doi.org/10.1348/000709906X116768

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121–136. https://doi.org/10.1080/15248371003699969.

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, *94*(4), 687–698. https://doi.org/10.1037/0022-0663.94.4.687

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

Dockrell, J., Connelly, V., Walter, K., & Critten, S. (2018). The role of curriculum based measures in assessing writing products. In B. Miller, P. McCardle, & V. Connelly (Eds.), *Writing development in struggling learners* (pp. 182–197). Brill. https://doi.org/10.1163/9789004346369_011

Englert, C. S., Stewart, S. R., & Hiebert, E. H. (1988). Young writers' use of text structure in expository text generation. *Journal of Educational Psychology*, *80*(2), 143–151. https://doi.org/10.1037/0022-0663.80.2.143

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The Journal of Special Education*, *34*(3), 140–153. https://doi.org/10.1177/002246690003400303

Espin, C., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice*, *25*(2), 60–75. https://doi.org/10.1111/j.1540-5826.2010.00304.x

Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum- based measurement in third- through eighth-grade students. *Psychology in the Schools, 51*(1), 85–96. https://doi.org/10.1002/pits.21733.

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. (2017). Curriculum-based measurement as the emerging alternative: Three decades later. *Learning Disabilities Research & Practice, 32*(1), 5–7. https://doi.org/10.1111/ldrp.12127

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. https://doi.org/10.1037/0022-0663.89.1.170.

Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide (NCEE 2012-4058)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications

Graham, S., & Harris, K. R. (1997). It can be taught, but it does not develop naturally: Myths and realities in writing instruction. *School Psychology Review*, *26*(3), 414–424. https://doi.org/10.1080/02796015.1997.12085875

Graham, S., Hebert, M., Paige Sandbank, M., & Harris, K. R. (2016). Assessing the writing achievement of young struggling writers: Application of generalizability theory. *Learning Disability Quarterly*, *39*(2), 72–82. https://doi.org/10.1177/0731948714555019

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445. https://doi.org/10.1037/0022-0663.99.3.445

Hammill, D. D., Larsen, S. C. (2009). Test of written language (4th ed.). San Antonio, TX: Pearson.

Huang, F. L., Moon, T. R., & Boren, R. (2014). Are the reading rich getting richer? Testing for the presence of the Matthew effect. *Reading & Writing Quarterly*, *30*(2), 95-115. https://doi.org/10.1080/10573569.2013.789784

Institute for Education Sciences, Technical Working Group. (2017). *Future directions for writing research at the secondary level*. Washington, DC. Retrieved from https://ies.ed.gov/ncer/whatsnew/techworkinggroup/

Keller-Margulis, M. A., Mercer, S. H., Payan, A., & McGee, W. (2015). Measuring annual growth using written expression curriculum-based measurement: An examination of seasonal and gender differences. *School Psychology Quarterly*, *30*(2), 276–288. https://doi.org/10.1037/spq0000086

Kent, S. C., & Wanzek, J. (2016). The relationship between component skills and writing quality and production across developmental levels: A meta-analysis of the last 25 years. *Review of Educational Research, 86*(2), 570–601.

Kim, Y.-S. G., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2018). Theorization and an empirical investigation of the component-based and developmental text writing fluency construct. *Journal of Learning Disabilities, 51*(4), 320–335. https://doi.org/10.1177/0022219417712016

Lembke, E., Hampton, & Hendricker, (2013). Data-based decision making in academics using curriculum-based measurement. In B. Cook & M. Tankersley (Eds.) *Research-Based Practices in Special Education* (pp. 235-248). Pearson.

McCoach, D. B. (2018). Multilevel modeling. In Gregory R. Hancock (Ed.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (2nd ed., pp. 292–312). Routledge.

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. https://doi.org/10.1080/02796015.2008.12087867

McMaster, K. L., Ritchey, K. D., & Lembke, E. (2011). Curriculum-based measurement for beginning writers: Recent developments and future directions. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Assessment and Intervention* (Vol. 24, pp. 111–148). Emerald Group Publishing Limited. https://doi.org/10.1108/S0735-004X(2011)0000024008

Monroe, M., & Sherman, E. (1966). Group Diagnostic Reading Aptitude and Achievement Tests; C. H. Nevins Printing Co.

O'Connell, A. A., Logan, J. A. R., Pentimonti, J. M., & McCoach, D. B. (2013). Linear and quadratic growth models for continuous and dichotomous outcomes. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 125–168). Taylor & Francis Group.

Parr, J. M. (2010). A dual purpose data base for research and diagnostic assessment of student writing. *Journal of Writing Research, 2*(2), 129–150.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development. *Review of Educational Research, 84*(2), 203–244. https://doi.org/10.3102/0034654313509492

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2020). *nlme: Linear and Nonlinear Mixed Effects Models* (2.1-148) [R]. https://cran.r-project.org/package=nlme

Ritchey, K. D., McMaster, K. L., Al Otaiba, S., Puranik, C. S., Grace Kim, Y.-S., Parker, D. C., & Ortiz, M. (2016). Indicators of fluent writing in beginning writers. In K. D. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 21–66). Springer. https://doi.org/10.1007/978-1-4939-2803-3_2

Romig, J. E., Miller, A. A., Therrien, W. J., & Lloyd, J. W. (2020). Meta-analysis of prompt and duration for curriculum-based measurement of written language. *Exceptionality*. https://doi.org/10.1080/09362835.2020.1743706

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

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education*, *51*(2), 72–82. https://doi.org/10.1177/0022466916670637

Santi, K. L., Barr, C., Khalaf, S., Francis, D. J. (2015). Different approaches to equating oral reading fluency passages. In Cummings, K. D., Petscher, Y. (Eds.), *The fluency construct* (pp. 223–265). New York, NY: Springer. doi:10.1007/978-1-4939-2803-3_9.

Scammacca, N., Fall, A.-M., Capin, P., Roberts, G., & Swanson, E. (2020). Examining factors affecting reading and math growth and achievement gaps in grades 1–5: A cohort-sequential longitudinal approach. *Journal of Educational Psychology, 112*(4), 718–734. https://doi.org/10.1037/edu0000400

Shapiro, E. S., & Guard, K. B. (2014). Best practices in setting progress monitoring goals for academic skill improvement. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Student-level services* (pp. 51–66). National Association of School Psychologists.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press.

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

Troia, G. A., Brehmer, J. S., Glause, K., Reichmuth, H. L., & Lawrence, F. R. (2020). Direct and indirect effects of literacy skills and writing fluency on writing quality across three genres. *Education Sciences, 10*, 297. doi:10.3390/educsci10110297.

Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine: Changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities*, *35*(4), 290–305. https://doi.org/10.1177/00222194020350040101

Truckenmiller, A. J., McKindles, J. V., Petscher, Y., Eckert, T. L., & Tock, J. L. (2020). Expanding curriculum-based measurement in written expression for middle school. Journal of Special Education, 54, 133-145. doi:10.1177/0022466919887.

Wilkinson, G. S. (1993). Wide range achievement test: WRAT3. Western Psychological Services. https://search.library.wisc.edu/catalog/999714707202121.

Wood, C. L., Schatschneider, C., & Hart, S. (2020). Average one year change in lexical measures of written narratives for school age students. *Reading & Writing Quarterly*, *36*(3), 260–277. https://doi.org/10.1080/10573569.2019.1635544

Wood, C., Schatschneider, C., & Wanzek, J. (2020). Matthew effects in writing productivity during second grade. *Reading and Writing*, *33*(6), 1377–1398. https://doi.org/10.1007/s11145-019-10001-8