When the type of practice matters:

The relations between writing instruction, student practice, and writing achievement in first grade

David L. Coker, Jr.

Austin S. Jennings

Elizabeth Farley-Ripple

Charles A. MacArthur

University of Delaware

Published *Contemporary Educational Psychology*, 2018

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110484 to the University of Delaware. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

A substantial body of research has demonstrated the efficacy of early writing interventions that target skills and composing processes; however, much less is known about the effectiveness of classroom writing instruction outside of the context of an intervention. The goal of this study was to investigate if writing instruction and student practice predict first-grade writing achievement and if the relations between writing instruction, practice and achievement depend on student factors. Assessments of students' spelling, handwriting, vocabulary, and reading were collected in the fall of first grade, and norm-referenced and researcher-designed writing tasks were administered in the spring (N=391). During the school year, four full-day observations of classroom instruction and student writing practice were conducted in 50 classrooms. The effects of writing instruction and student writing practice on spring writing achievement were analyzed using two-level, fixed-effects hierarchical linear models. Composing instruction was negatively related to contextualized spelling, but no other main effects of instruction were found. One type of writing practice, generative writing, was positively related to all three measures of writing achievement. Interactions were also found between student gender, minority status and multiple types of writing instruction and practice. These results point to the potential benefit of generative writing practice and indicate that efforts to differentiate instruction and practice may be beneficial for students. Additionally, the findings raise doubts about the effectiveness of current writing instruction.

Keywords: writing instruction, writing, first grade

With the widespread implementation of the Common Core State Standards (CCSS, National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010), many schools are faced with the prospect of evaluating their approach to writing instruction. Before the implementation of the CCSS, there was fairly wide variation in state writing standards, and in many states components of those standards did not align with the CCSS (Troia, et al., 2016). As a result of the CCSS writing standards, many school districts faced much more challenging expectations in terms of both student writing outcomes and processes (Shanahan, 2015).

In light of changing expectations, schools may question whether their current writing instruction is preparing students to meet the writing CCSS (Graham & Harris, 2015). Surprisingly, there is relatively little direct data on elementary school teachers' instructional approaches to writing (for exceptions see: Coker et al., 2016; Kim, Al Otaiba, Sidler, & Gruelich, 2013; Puranik, Al Otaiba, Sidler, & Greulich, 2014). There is even less current evidence about the efficacy of instructional approaches in the primary grades. Some of the best data comes from large-scale national assessments, such as the National Assessment of Educational Progress (NAEP). These results suggest that many elementary school students struggle to write well; for example, only 28% of fourth graders scored at or above the proficient level in 2002 (Persky, Daane, & Jin, 2003). Moreover, the most recent NAEP writing data on elementary-aged students is over ten years old and may not reflect schools' efforts to align their writing instruction to the CCSS. Despite limited information on the impact of writing instruction that teachers provide in the absence of a specific intervention, there is a growing body of empirical evidence on effective writing interventions for young students.

Effective Writing Instruction

Although there is far less evidence for writing than reading interventions, the recent publication of consensus reports and meta-analyses on writing signals that this body of work is large enough to be summarized (Graham, Bollinger, et al., 2012; Graham, Harris, & Santangelo, 2015; Graham & Hebert, 2011; Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007; Graham & Santangelo, 2014; Santangelo & Graham, 2016). These findings point to three areas that benefit from teacher attention: writing component skills, composing processes, and opportunities for writing practice.

Writing component skills. There is solid theoretical and empirical support for the role of writing component skills in early development. In the Not-So-Simple-View of Writing (NSSVW; Berninger & Winn, 2006), both text generation skills and transcription skills figure prominently in the model. Text generation involves generating ideas and translating them into language, which depends on oral language knowledge such as vocabulary and syntax. Transcription skills are needed to put words on paper and include handwriting, typing, and spelling. These skills are thought to develop early in development (Berninger & Swanson, 1994). Fluency with transcription skills is critical to early writing development because when students struggle to inscribe letters (handwriting or typing) or to spell words, there is less cognitive capacity available for higher order tasks such as planning, evaluating, and revising (McCutchen, 2000). In contrast, when students can write and spell fluently more cognitive resources are available to generate text and to engage in writing processes.

The positive impact of writing component skills has been demonstrated in both correlational and intervention studies. Researchers investigating early writing have found that component skills predict important writing outcomes in multi-grade samples (Berninger, Abbott, Abbott, Graham, & Richards, 2002; Coker, 2006; Graham, Berninger, Abbott, Abbott, &

Whitaker, 1997) and those limited to first grade (Kim, Al Otaiba, Folson, Greulich, & Puranik, 2014; Kim & Schatschneider, 2017). Instructional studies also support the importance of writing skills. In a meta-analysis of studies with students in grades K-12, Santangelo and Graham (2016) reported that handwriting instruction had a positive effect on how much students wrote (ES = 1.33), how fluently they wrote (ES = .48), and the quality of their writing (ES = .84). Instruction in spelling, another important transcription skill, has been found to improve how well students spell when composing (ES = .94; Graham & Santangelo, 2014).

Composing. The processes involved in composing have also been important components in theoretical and empirical investigations of writing. The NSSVW identifies the importance of composing processes such as planning and revising as parts of an individual's executive functions (Berninger & Winn, 2006). In addition, self-regulation processes, such as goal-setting, monitoring, and evaluating, are also included with these composing processes. These executive functions contribute to a writer's ability to coordinate the writing process during composing.

Interventions designed to teach students components of the composing process have been found to be effective. A meta-analysis focusing on elementary school found that the self-regulated strategy development model, which includes direct instruction in the writing process and self-regulation strategies targeting goal setting, monitoring, and evaluating, had a large effect size (ES 1.17; Graham et al., 2012). Other approaches designed to teach the composing process also demonstrated a positive effect on writing quality (ES = .59). These results aligned with those from a meta-analysis with studies of older students (Graham & Perin, 2007). Neither of these meta-analyses included studies on composing instruction in first grade due to the lack of research, but there is limited evidence that composing instruction is effective at this grade range (Zumbrunn & Bruning, 2013).

Opportunities for writing practice. In addition to instruction in skills and composing, students also benefit from frequent opportunities to write. Practice writing has been identified as a way to improve writing quality, even though the evidence is less robust (Graham, Bollinger, et al., 2012; Graham & Perin, 2007). In both skills and composing interventions, students engage in writing practice as a component of the instruction. Frequent writing practice has also been identified is a characteristic of effective literacy teachers (Graham & Perin, 2007). In addition, the Institute of Education Sciences (IES) Practice Guide recommends at least 30 minutes a day should be devoted to writing practice beginning in first grade, although the authors note that there is relatively little evidence to support this recommendation (Graham, Bollinger, et al., 2012).

This selective review highlights several broad practices that may support writing development. For these practices to be effective, they must be widely applied in classrooms. However, data on teachers' instructional activities and student writing practice is sparse.

Current State of Early Writing Instruction

Surveys with teachers (Cutler & Graham, 2008; Richards, Sturm, & Cali, 2012) and observations in kindergarten and first grade (Coker et al., 2016; Kim et al., 2013; Puranik et al., 2014) have revealed, on average, modest amounts of writing instruction occur. Furthermore, large classroom variation is common. Time allocated to writing instruction varied from 1 minute a day in kindergarten, (range: 0-8.86 min; Puranik et al., 2014) to 26.4 minutes a day in first grade (range: 5.50- 74.25 min; Coker et al., 2016). Across the primary grades, teachers reported an average of 21 minutes a day for instruction (Cutler & Graham, 2008). Most kindergarten instruction was used for handwriting (Puranik et al., 2014), but more diversity was found in first grade with 32.55% for skills instruction, and 54.4% for composing instruction (Coker et al.,

2016). A wide mix of composing and skills activities were also reported by teachers across the elementary grades (Cutler & Graham, 2008; Richards et al., 2012).

In general, more time was allocated for practice than for writing instruction in kindergarten and first grade. During the daily 90-minute block for kindergarten literacy instruction, approximately 8 minutes was for writing practice (range: 0-20.58 min; Puranik et al., 2014). Across the entire day, Coker et al. (2016) found students engaging in some form of writing practice for 125 minutes out of approximately 405 minutes in the school day.

Teachers also provided time for varying types of writing practice. In first grade 40% of practice time involved either copying words or filling in an a one-word response (Coker et al., 2016). Another 25% of practice time was used for generative writing, which required students to create the content and produce connected text. Similarly, first-grade teachers reported having students complete worksheets, copy individual words, or engage in handwriting practice for at least 100 out of 180 school days (Richards et al., 2012). Although these were the most frequent practice activities, Richards et al. reported that students wrote a wide range of texts; however, many received very little attention on average.

Efficacy of current instructional practices. Available data reveal wide variation around the amount and type of writing instruction and practice in the early grades. In light of this extensive variation, it seems likely that many teachers are not following current recommendations for writing instruction and practice (e.g., Graham, Bollinger, et al., 2012; Graham et al., 2015). This could hinder students' ability to meet the writing CCSS, but it is difficult to predict because there is little data on whether the frequency and type of early writing instruction and practice contribute to student writing achievement.

Student characteristics. There is also evidence that student demographics, such as gender and ethnicity, are related to writing achievement and might also impact the efficacy of writing instruction and practice. In the 2002 NAEP writing assessment, girls scored higher than boys in grades 4, 8, and 12 (Persky et al., 2003). Other empirical investigations of writing achievement in elementary school have also found that girls outperformed boys across a range of writing outcomes (Berninger & Fuller, 1992; Kim, Al Otaiba, Wanzek, & Gatlin, 2015). In a study of first-grade writing, gender was found to interact with the first-grade classroom teacher, and in most but not all classrooms, girls were found to write more and generate higher quality descriptions at the end of first grade and through third grade (Coker, 2006).

Student ethnicity has also been linked to writing performance. On the 2002 NAEP writing assessment, Asian/Pacific Islanders posted the highest scores, and White students outperformed Black and Hispanic students at all three grade levels (Persky et al., 2003). Coker (2006) found no main effects of ethnicity, but there was an interaction between student ethnicity and classroom teacher, such that the quality and length of texts written by Black students varied depending on the specific first-grade teacher.

The Present Study

The goal of this study was to address questions about the effectiveness of the amount and type of writing instruction and practice on writing achievement in first grade. In order to assess the impact of writing instruction and practice, we controlled student achievement in the fall of the year. These control measures included writing skills such as spelling and handwriting because of their demonstrated relations to writing achievement (Kent & Wanzek, 2016). In addition, receptive vocabulary and reading achievement were also included as controls because both have been shown to be related to early writing (Coker, 2006; Berninger et al., 2002; Kim,

Otaiba, Puranik, Folsom, & Gruelich, 2014; Lerkkanen, Rasku-Puttonen, Aunola, & Nurmi, 2004). Finally, we controlled for the total amount of reading instruction and reading practice that students experienced during the school year. Control variables for reading were included because many of the recommended instructional targets, such as the alphabetic principle and morphological awareness (Foorman et al., 2016), support students' efforts to spell and write (Graham, Bollinger, et al., 2012).

Two types of writing instruction (skills and composing) and three types of student writing practice (correct/copy, generative writing, and writing about text) were investigated. Specifically, we asked, RQ1) if writing instruction and practice predict first-grade writing achievement? And RQ2) if the relations between writing instruction, practice and achievement depend on student factors (i.e., demographics)?

Method

This analysis is a component of a larger project on first-grade writing instruction and student literacy achievement. In previous publications, the observational measures (Coker et al., 2016; Coker et al. 2018), student fall measures (Coker et al. 2018b), and researcher-designed writing outcomes (Coker et al., 2018) have been reported with this sample.

Participants

Students. All student participants (N = 391) were first-graders in a Mid-Atlantic state. The students were in 50 classrooms in 13 schools in three school districts across two school years. The size of the participating schools varied with a range of two to six first-grade classrooms per school. Between 10,000-17,400 students from urban and suburban neighborhoods were enrolled in the three school districts.

The demographic background of participating students was half female (51.9%) and White (50.6%) and included students of other ethnicities (African American 28.6%, Hispanic 12.3%, Asian 4.9%, Other 3.3%). The sample also included a small percentage of English Language Learners (8.7%) and students with disabilities (11.7%). Schools reported slightly over half of their participating students (54.9%) qualified for free or reduced-price meals (range: 15.9%-84.8%). The school districts only provided school-level socio-economic status (SES) data. Furthermore, during data collection the state department of education revised how it calculated student SES. As a result, school-level SES appeared to change even though there was little variation in the participating schools' populations. We report SES information from year one of data collection to enable comparisons among schools.

Teachers. First-grade teachers in the participating schools were asked to participate in the project. In total, fifty-seven teachers from 50 first-grade classrooms volunteered. In three classrooms a co-teaching model was used, and in four other classrooms, the teacher was replaced with a long-term substitute. Teachers were predominately female (96.3%) and White (90.7%), but the sample included two African-American teachers and one Hispanic teacher. Teachers averaged 14.94 (*SD* = 7.98) years of teaching, and they reported 8.69 (*SD* = 6.97) years of experience teaching first grade. Close to half of the teachers held a master's degree (48.1%); others reported a bachelor's degree (14.8%), a bachelor's degree with additional coursework (18.5%), or a master's degree with additional coursework (18.5%). A $200 honorarium was given to teachers for each semester they participated.

School Context

In each participating classroom, there were fewer than 22 students. The writing curriculum varied across the participating classrooms. In approximately half, there was no

standard writing curriculum (*n* = 23), and in 22 classrooms, it was part of the reading curriculum. A curriculum resource, *Explorations in Nonfiction Writing* (Stead & Hoyt, 2011), was also used in five classrooms. Similarly, the reading curricula varied across the classrooms. The most common curricula included Houghton Mifflin Harcourt's *Journey's* (*n* = 32; Baumann et al., 2011) or no published reading curriculum (*n* = 10). Other classrooms used Pearson Scott Foresman's *Reading Street* (*n* = 5; Afflerbach et al., 2011), and *Discover Intensive Phonics for Yourself* (*n* = 3; Lockhard & Eversole, 2006).

Observational Data

The classroom observational data were collected over two years in 50 classrooms. In the first year, 21 classrooms in five schools were observed; 29 classrooms from nine schools were observed in the second year. Teachers and students only participated in data collection during a single year.

Participating first-grade classrooms were observed four times during the year. The decision to conduct four observations was based on other studies of elementary literacy instruction have relied on three or fewer observations (Connor, Morrison, & Petrella, 2004; Foorman et al., 2006; Hoffman, Sailors, Duffy, & Beretvas 2004; Kim et al., 2013; Silverman & Crandell, 2010; Taylor, Pearson, Peterson, & Rodriguez, 2003). In addition, our observations extended across the entire school day, which allowed us to capture reading and writing throughout the day. This is a departure from previous work that has observed the literacy instructional block (Foorman et al., 2006; Puranik et al., 2014; Silverman & Crandell, 2010; Taylor et al., 2003).

Observation schedules. The four, day-long observations began in late October and were completed by the end of May. We notified teachers of the observations in advance. Observers

coded whenever the class was engaged in academic content, but no coding occurred during special classes (e.g., art, music, library, etc.), recess, and lunch. On average, the number of days between observations was 54.89. Every effort was made to maintain consistent time intervals between observations, but interruptions due to school holidays, state testing, closures due to inclement weather, and special events complicated this effort.

Observational codes. Coding relied on a time-sampling procedure to record the presence of instructional and practice activities in each 5-minute block. All academic instruction was observed for each classroom, and codes were recorded every 5 minutes for each group. After watching classroom instruction for three minutes, coders had two minutes to enter codes. An iPad application called iSeeNCode was used for the project (Hofstetter, 2016). The application reminded coders when to observe and when to code. Entering the codes was accomplished using the touch screen, and the application stored the data in a spreadsheet. In total, iSeeNCode was loaded with 111 individual dichotomous (present or absent) codes organized into multiple dimensions, four of which were used in this analysis: 1) Writing instructional focus, 2) Reading instructional focus, 3) Student writing practice, and 4) Student reading practice. The codes were chosen for the coding system based on relevant theory and empirical research on writing instruction (Coker et al., 2016). Codes for grouping, management of instruction, oral instructional focus, transitions, other academic instruction, not academic, and materials were not included in this analysis.

Writing instructional focus. To identify variability in teachers' approaches to instruction, twelve writing instructional codes were used. These codes were grouped in terms of two domains: skills writing instruction (spelling, grammar, handwriting, keyboarding, and punctuation/capitalization) and composing writing instruction (process writing, revising, editing,

narrative composing, informative composing, sharing student writing, and sharing teacher writing). We combined the codes to reflect the types of first-grade instruction, which includes both skills and composing (Coker et al., 2016; Cutler & Graham, 2008), and to align with developmental theory and empirical evidence. Transcription skills and composing processes are central in the NSSVW (Berninger & Winn, 2006). There is also substantial empirical evidence that both skills instruction (Graham et al., 1997; Santangelo & Graham, 2016) and composing instruction, particularly strategy instruction (Harris, Graham, & Mason, 2006; Zumbrunn & Bruning, 2013), in the primary grades can strengthen writing achievement.

Reading instructional focus. Our codes for reading instruction included both code-based and meaning-focused instructional activities. Examples include instruction in phonological/ phonemic awareness, word recognition/decoding, vocabulary, reading (fluency), and three types of comprehension codes (lower level, higher level, and strategy). In the analysis, individual codes were aggregated to control for the total amount of classroom reading instruction.

Student writing practice codes. The type and amount of student writing practice that was occurred in each group was also coded by observers. Three types of writing activities were coded: correct/copied writing, writing about text, and generative writing practice. A correct/copied response was noted when students either wrote a response that had a single correct answer, such as a worksheet, or when they were copying text. The writing about text practice code was applied when students were writing in response to something they had read or the teacher had read. The generative writing practice code was applied in situations when students were expected to write a text at least a sentence long, and there was not an expected response. As a result, students had some autonomy to determine the content of the text. Generative writing

practice encompassed narrative and informative texts, open-topic journal entries, and any other open-ended writing tasks.

Student reading practice codes. The amount and type of student reading practice was also coded. The reading activities were created to capture all instances of reading practice, including reading, reaching chorally, and reading taking turns. The reading code was used when students were reading individually. Reading chorally was applied when more than one student was reading the same passage out loud, and the taking turns reading code was used when students were in a group reading a text out loud by taking turns. As with reading instruction, individual codes were aggregated to control for the total amount of classroom reading practice.

Observer training and reliability. The observers were former teachers who had extensive classroom experience. Before data collection began, observers received extensive training in the use of iSeeNCode (Hofstetter, 2016). The observers practiced coding with videos of classroom instruction, and coding disagreements were resolved by the master coder (first author). After observers could code a 30-minute classroom video with at least 80% agreement with the master coder, coding practice was moved to a cooperating first-grade classroom. Practice sessions were repeated until observers demonstrated a minimum threshold of .80 of both kappa and inter-scorer agreement. For the four observers, the average agreement across coding dimensions (e.g., broad instructional focus, specific writing focus, etc.) averaged .92 with a range of .87 to .96.

Over the course of the study, observers' reliability was assessed in two ways. First, the master coder double coded 90 minutes of a school day with each observer for every data collection point (four times each year). Observers needed a minimum agreement of .8 to proceed

with data collection. Then, before the second year of data collection, a reliability check-up was conducted that required observers to reach the agreement threshold of .80 with the master coder.

Observations. Each classroom in the sample was observed for four full instructional days during an academic year. With 50 participating classrooms, there was a total of 200 observation days and 11034 five-minute observation blocks (M = 220.68 5-minute blocks per classroom). The coding protocol described above was applied to all five-minute blocks, creating a dataset of dichotomous variables characterizing the presence of instruction and practice during each observation block. To convert these data into instructional measures for each classroom, we aggregated observational block data to the observation day, then across days to the classroom, using several steps. First, we determined the number of observational blocks that an instructional code was observed each day. This metric provides an index of how much time was devoted to types of instruction and student practice in a given day. Second, we averaged the number of blocks in which a code was observed across observations. This provides an estimate of the average time in each classroom devoted to types of instruction and student practice across the year. This average measure of classroom instruction and practice was used for the current analysis. For example, if observers recorded five blocks of skills-focused writing instruction during the first observation, seven in the second, three in the third, and two in the last, the average would be 4.25 blocks of skills-focused writing instruction.

The average number of blocks was used because our objective was to characterize classroom instruction and student practice *over the entire school year*. As a result, we had a more representative assessment of classroom experiences in first grade. These data allowed us to determine the amount of time (i.e., average number of blocks) allocated to writing and reading instruction and practice during the year.

Student Measures

Decoding and word reading. Two subtests from the Woodcock-Johnson III Tests of Achievement (WJ-III; McGrew, Schrank, & Woodcock, 2007) were administered in the fall and spring to assess students' decoding skills: Letter Word Identification (LWID) and Word Attack (WA). For the LWID subtest, examiners showed students a list of letters and words to read aloud. In the WA subtest, examiners asked students to read a list of pseudowords aloud. Both assessments were discontinued when students missed six consecutive items. In the analyses, LWID and WA were combined to form the Basic Reading cluster.

Handwriting fluency. A researcher-developed assessment of letter-writing fluency was administered to students in the fall (Abbott & Berninger, 1993). Examiners gave students a piece of lined paper and a pencil with no eraser. Students were asked to write 26 lower-case alphabetic letters in one minute as fast as possible. Every 15 seconds the examiner marked the paper to indicate students' progress. Each correctly formed letter in alphabetical order received one point. Letters were scored as incorrect if they were illegible, incorrectly formed, or out of alphabetical order. Other researchers have used this assessment with both a 15-second (Berninger & Rutberg, 1992; Graham et al., 1997; Jones & Christensen, 1999) and a 60-second time limit (Kim, Al Otaiba, Folsom, et al., 2014; Wagner et al., 2011). In our sample, some students stopped writing before 60 seconds had elapsed, but none completed the task in under 45 seconds. To account for these differences, we used the total number of correctly written letters in the first 45 seconds for analysis.

Vocabulary. To assess the breadth of students' vocabulary knowledge, we administered the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007) in the fall.

For each item, the examiner read a word and students indicated which one of four pictures to represent the word. Testing was discontinued when students missed eight or more items in a set.

Spelling. The WJ-III Spelling subtest was administered in the fall and the spring (Woodcock, McGrew, & Mather, 2001; 2007). Students were asked to write letters or individual words that get increasingly difficult. Testing continued until students missed six consecutive items. Inter-scorer agreement was 100%.

Writing Outcome Measures

For these analyses, we used multiple writing assessments for several reasons. First, our theoretical framework, the NSSVW, depicts writing as a complex interaction among multiple skills and knowledge sources (Berninger & Winn, 2006). It would be challenging to index these multiple sources in a single assessment. Additionally, there is good evidence that writing is not a unidimensional construct, even in first grade (Kim, Al Otabia, Folsom, et al., 2014). We chose multiple assessments that are both norm-referenced and authentic tasks in an attempt to reliability assess students' early writing.

Broad Written Language Cluster. For the analyses the Broad Written Language cluster (BWL) was used as an index of global writing achievement. It is formed from WJ-III Spelling, Writing Fluency, and Writing Samples subtests following the assessment manual (Woodcock et al., 2001, 2007). The WJ-III Spelling subtest, as described previously, was included in the BWL cluster. Sentence writing fluency was assessed with the WJ-III Writing Fluency subtest. Students were given 7 mins to write as many simple sentences as possible three related words and a picture. Global writing proficiency in the spring was assessed using the WJ-III Writing Samples subtest. Students responded to a series of prompts that increased in difficulty in terms of the length, vocabulary, grammar and conceptual knowledge.

Extended writing prompts. In the spring, students were given researcher-developed writing prompts in two different genres, narrative descriptive. The narrative prompt read, "Think about one of your favorite activities. Write a story about a time that you had fun doing this activity." The descriptive prompt was, "Think about a person you know well. It could be someone in your family or a friend. Describe that person and tell what he or she is like to someone who doesn't know him or her." The examiner provided students with lined paper including the prompt at the and a pencil. Next the examiner read the prompt aloud. Then, students were given 20 minutes to write. Once students had finished, they were instructed to reread and check their work.

The narrative and descriptive tasks were scored for length, quality, contextualized spelling, syntactic complexity and mechanics to capture the multi-dimensional nature of first-grade writing (Kim, Al Otaiba, Folsom, et al., 2014). The narrative and descriptive texts were transcribed to reduce bias for poor handwriting before they were scored. Furthermore, students' spelling mistakes were corrected before the texts were scored for quality.

Length. Text length was computed as the total number of words using the word count formula in Microsoft Excel. However, letter or word sequences that were not legal words (e.g., *qlArqrsuus* or *MeaMyBIDBeISesMocaCat*) were excluded from the total (14 from narrative texts and 45 from descriptive texts were excluded, less than .5% of the sample).

Contextual Spelling. Contextualized spelling was the percentage of correctly spelled words. Inter-scorer agreement was calculated using 20% of the texts for each genre. Inter-scorer agreement was 99.3% for narrative contextual spelling and 98.9% for descriptive contextual spelling.

Quality. A 6-point holistic rubric was applied to assess the global quality of the texts. There were three dimensions of the rubric: (1) topic and detail, (2) organization and supporting details, and (3) word choice. We used the same quality rubric for both narrative and descriptive texts. Inter-scored agreement (± 1 point) was calculated for 100% of the texts. Inter-scorer agreement was 96.2% for narrative quality and 96.8% for descriptive quality. The Spearman rho correlation between scorers was .88 for narrative quality and .87 for descriptive quality.

Factor scores. In addition, the texts were also scored for syntactic complexity and mechanics; however, these scores are not used in this analysis. To simplify the data, a confirmatory factor analysis was conducted and the best fitting model contained four factors: Quality/Length, Spelling, Mechanics and Syntactic Quality (Coker et al., 2018b; χ^2 = 155.0, RMSEA = .06, CFI = .91, TLI = .94). For example, the length and quality measures for both narrative and descriptive texts loaded onto the Quality/Length factor, and the contextualized spelling measures for both texts loaded onto the Spelling factor. These two factor scores, Quality/Length and Spelling, were used as assessments of students' compositional skill.

Administration and scoring. Trained research assistants (RA) administered student assessments in a quiet location outside of the classroom during the school day. For all assessments, the RAs followed the administration manuals. The standardized assessments included discontinuation rules or the use of basal and ceiling sets to minimize student fatigue (i.e., LWID, WA, PC, WJ-III Spelling, WJ-III Writing Samples, and PPVT-4).

Data Analysis Strategy

In the current study, students (N = 391) were nested in classrooms (N = 50), nested in schools (N = 13). The study included an average of 7.82 students per classroom (range: 4-9 students) and an average of 3.85 classrooms per school (range: 2-6 classrooms).

In multilevel analyses, maximum likelihood estimation methods require approximately 30 to 50 cluster units for accurate parameter and standard error estimation (Maas & Hox, 2005) while restricted maximum likelihood (REML) estimation may approximate unbiased estimates in samples with fewer than 12 cluster units (Browne & Draper, 2006). In the current study, classrooms are nested in 13 schools, which approaches the minimum level-specific sample size recommendation for unbiased parameter and standard error estimation. An alternate method of accounting for the non-independence of classrooms within schools is to include school fixed effects. In the current study, we specified two-level fixed effects hierarchical linear models (HLMs) in which students (Level-1) are nested within classrooms (Level-2) with school fixed effects to account for school-level variability in student achievement.

Data missing at random was imputed via multiple imputation (m = 25) with the inclusion of auxiliary variables to increase the accuracy of imputed values (Baraldi & Enders, 2010; Rubin, 1987; Schafer, 2003). The inclusive multiple imputation procedure employed both student demographics (e.g., age in months, gender) and fall achievement variables (e.g., PPVT-4, WJ-III LWID). At the student-level, we imputed 18 values across 4 measures (range: 3-7 missing data points per measure; maximum of one data point per student). No classroom-level data was missing.

We developed HLM specifications to explore the relations between student-level demographics and fall achievement, classroom-level instruction and practice, and three measures of student-level writing achievement: WJ-III BWL, Quality/Length, and Spelling. Intraclass correlation coefficients (ICCs) indicate the proportion of variance between, as opposed to within, groups (Heck & Thomas, 2015). ICCs between .05 and .15 traditionally indicate adequate variance for multilevel modeling in education contexts (Hedges & Hedberg, 2007). ICCs for the

three measures of writing achievement reveal between 6 and 13 percent variance at the classroom level ($\rho_{BroadWrittenLanguage}$ = .13, $\rho_{Quality/Length}$ = .07, $\rho_{ContextualSpelling}$ = .06).

We developed unique model specifications for each measure of writing achievement through a four-step modeling building process. The general empirical model was:

$$Y_{ij} = \gamma_{00} + \boldsymbol{\gamma_{i0}X_{ij}} + \boldsymbol{\gamma_{0j}X_{j}} + \boldsymbol{\gamma_{ij}X_{ij}X_{j}} + r_{ij} + u_{0j}$$

where:

Y_{ij} = outcome variable

$\boldsymbol{\gamma_{i0}X_{ij}}$ = vector of main effects for student *i* in classroom *j*

$\boldsymbol{\gamma_{0j}X_{j}}$ = vector of main effects for classroom *j* (including fixed effects)

$\boldsymbol{\gamma_{ij}X_{ij}X_{j}}$ = vector of cross-level interaction effects for student *i* in classroom *j*

r_{ij} = random effect for student *i* in classroom *j*

u_{0j} = random effect for classroom *j*

First (Model 1), we estimated a baseline model to calculate the classroom-level variance component, or proportion of variation in spring writing achievement between classrooms. Second (Model 2), we estimated a student-level structural model to explore the relations between demographic variables, fall achievement, and measures of spring writing achievement. Third (Model 3), we estimated a classroom-level structural model to explore the relations between measures of writing and reading instruction, practice, and spring writing achievement. Finally (Model 4), we estimated a cross-level interaction model to explore interaction effects between student demographics, classroom writing instruction and practice, and spring writing achievement. Models 3 and 4 include four sets of fixed effects to control for classroom-level conditions (presence/absence of a co-teacher, long-term substitute, and formal writing

curriculum) and school-level variation. All models were estimated using the HLM 7 REML estimator (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011).

Student-level variables were group-mean centered while classroom-level variables were grand-mean centered in all model specifications. As a result, the student-level intercept may be interpreted as the outcome score of the average student in the average classroom. Student-level coefficients may be interpreted as the difference in outcome score between student categorizations (*gender* represents the differential performance of female students relative to male students; *minority status* represents the differential performance of non-white students relative to white students)while controlling for fall achievement. Classroom-level coefficients directly address RQ1 and may be interpreted as the difference in outcome score per one-unit increase in a given measure of instruction or practice. Cross-level interaction terms directly address RQ2 and may be interpreted as the differential relations between student-level variables and the outcome per one-unit increase in a given classroom-level measure of instruction or practice.

Results

Descriptive statistics, bivariate correlations, and reliability estimates for the student measures are presented in Table 1. Results from norm-referenced assessments of vocabulary and spelling were close to the national average. Students' Basic Reading and BWL performance was higher than average, which might signal stronger code-based kindergarten instruction. For the classroom measures, descriptive statistics and bivariate correlation may be found in Table 2. Classrooms averaged 1.84 5-minute blocks of skills writing instruction and 3.11 blocks of composing writing instruction per observation day. In terms of writing practice, the most common type was correct/copy practice (19 blocks) followed by generative writing practice

(6.78 blocks), and writing about text (3.2 blocks). However, it should be noted that there was extensive classroom variation in both writing instruction and practice.

Broad Written Language Model Specifications

Table 3 presents parameter, standard error, and variance component estimates for four spring BWL model specifications. The baseline model (Model 1) indicates 13 percent of variation in BWL occurs between classrooms ($\rho_{BroadWrittenLanguage}$ = .13). Variance components presented in Model 1 were used to determine the percent variation explained in subsequent models.

The student-level structural model (Model 2) explained approximately 70.7 percent of student-level variation in spring BWL. Model 2 indicates a significant positive association between gender and BWL while controlling for fall achievement, with female students performing better than male students (γ = 1.82, p = .009). However, the relations between minority status and BWL was not significant (γ = -0.15, p = .867).

The classroom-level structural model (Model 3) explained approximately 17.3 percent of classroom-level variation in spring BWL. Model 3 indicates a significant positive relation between generative writing practice and BWL (γ = 0.75, p = .045). However, the relations between other forms of writing practice, correct/copy and writing about text, and BWL were not significant (respectively, γ = -0.37, p = .138; γ = 0.00, p = .993). Furthermore, the associations between both types of writing instruction, skills and composition, and BWL was not significant (respectively, γ = 0.02, p = .962; γ = -0.55, p = .112).

The cross-level interaction model (Model 4) explored the relations between student demographics, writing instruction, generative writing practice, and spring BWL while controlling for fall achievement. At the student-level, gender maintained a significant positive

main effect (γ = 1.72, p = .016) while the main effect of minority status was not significant (γ = -0.35, p = .661). Although cross-level interactions including gender were not significant, there were two statistically significant cross-level interaction effects including minority status. The *minority status-generative writing practice* interaction term was significant and negative (γ = -0.42, p = .037) while the *minority status-skills writing instruction* interaction term was significant and positive (γ = 1.45, p = .003). These findings indicate, controlling for fall achievement, the spring BWL scores of the average white and non-white student in the average classroom are not significantly different. However, the spring BWL scores of the average white and non-white student are significantly different in classrooms that deviate from the average quantity of generative writing practice and/or skills writing instruction. Specifically, controlling for fall achievement, the average non-white student in a class that engaged in more than the average generative writing practice scored 0.42 points less than the average white student in the same class, per additional block of practice. Conversely, the average non-white student in a class that engaged in more than the average skills writing instruction scored 1.45 points higher than the average white student in the same class, per additional block of instruction.

Quality/Length Model Specifications

Table 4 presents parameter, standard error, and variance component estimates for four spring Quality/Length model specifications. The baseline model (Model 1) indicates 7 percent of variation in Quality/Length occurs between classrooms ($\rho_{Quality/Length}$ = .07). Variance components presented in Model 1 were used to determine the percent variation explained in subsequent models.

The student-level structural model (Model 2) explained approximately 39.9 percent of student-level variation in spring Quality/Length. Model 2 indicates a significant positive relation

between gender and Quality/Length while controlling for fall achievement, with female students performing better than male students (γ = 0.35, p = .001). However, the relation between minority status and Quality/Length was not significant (γ = 0.09, p = .420).

The classroom-level structural model (Model 3) explained approximately 61.5 percent of classroom-level variation in spring Quality/Length. Model 3 indicates a significant positive relation between generative writing practice and Quality/Length (γ = 0.05, p = .008). However, the relations between other forms of writing practice, correct/copy and writing about text, and Quality/Length were not significant (respectively, γ = 0.01, p = .527; γ = 0.00, p = .985). Furthermore, there was a significant negative relation between skills writing instruction and Quality/Length (γ = -0.09, p = .026) while the relation between composing writing instruction and Quality/Length was not significant (γ = -0.04, p = .066).

The cross-level interaction model (Model 4) explored the relations between student demographics, writing instruction, generative writing practice, and spring Quality/Length while controlling for fall achievement. At the student-level, gender maintained a significant positive main effect (γ = 0.34, p = .001) while the main effect of minority status was not significant (γ = 0.06, p = .555). Furthermore, there were two significant interaction effects. The *gender-generative writing practice* interaction term was significant and negative (γ = -0.06, p = .001). This finding indicates, controlling for fall achievement, the spring Quality/Length score of the average female student in the average class is 0.35 standard deviations higher than that of the average male student in the same class. However, in classes that engaged in more than the average generative writing practice, this achievement gap decreases by 0.06 standard deviations per additional block of generative writing practice. Conversely, the *minority status-skills writing instruction* interaction term was significant and positive (γ = 0.12, p = .050). This finding

indicates, controlling for fall achievement, the spring Quality/Length scores of the average white and non-white student in the average classroom are not significantly different. However, in classes that engaged in more than the average skills writing instruction, the average non-white student scores 0.12 standard deviations higher than the average white student in the same class, per block of additional skills writing instruction.

Narrative/Descriptive Contextual Spelling Model Specifications

Table 5 presents parameter, standard error, and variance component estimates for three spring Contextual Spelling model specifications. The baseline model (Model 1) indicates 6 percent of variation in Contextual Spelling occurs between classrooms ($\rho_{ContextualSpelling}$ = .06). Variance components presented in Model 1 were used to determine the percent variation explained in subsequent models.

The student-level structural model (Model 2) explained approximately 36.6 percent of student-level variation in spring Contextual Spelling. Model 2 indicates a significant positive association between gender and spring Contextual Spelling, with female students performing better than male students (γ = 0.13, p = .045). However, the relation between minority status and Contextual Spelling was not significant (γ = 0.15, p = .136).

The classroom-level structural model (Model 3) explained approximately 21.1 percent of classroom-level variation in spring Contextual Spelling. Model 3 indicates a significant positive association between generative writing practice and Contextual Spelling (γ = 0.06, p = .015). However, the relations between other forms of writing practice, correct/copy and writing about text, and Contextual Spelling were not significant (respectively, γ = -0.02, p = .115; γ = 0.01, p = .738). This finding indicates, controlling for fall achievement, the average student scores 0.06 standard deviations higher per additional block of generative writing practice. Furthermore, there

was a significant negative association between composition writing instruction and Contextual Spelling (γ = -0.05, p = .037), while the association between skills writing instruction and Contextual Spelling was not significant (γ = -0.06, p = .146). This finding indicates, controlling for fall achievement, the average student scores 0.05 standard deviations lower per additional block of composition writing instruction.

Cross-level interaction models (not reported) did not indicate any significant cross-level relations between student demographics, measures of classroom instruction or practice, and spring Contextual Spelling.

Discussion

Rising expectations for writing, as detailed in the CCSS, have raised awareness of the importance of effective writing instruction (Shanahan, 2015). Effective instruction is particularly important in the early grades to help students develop a strong foundation and to avoid future difficulties (Slavin, Madden, & Karweit, 1989). Observational studies of early writing instruction have revealed that many teachers typically engage in little instruction (Coker et al., 2016; Puranik et al., 2014) and what they do may not align with effective approaches (Graham, Bollinger, et al., 2012). The goal of this study was to determine if the amount and type of writing instruction and practice are associated with writing achievement in first grade, and if the relations between instruction and achievement and practice and achievement might depend on student factors. These two questions are discussed in turn.

Does Writing Instruction Predict Achievement?

Our analysis revealed that first-grade writing instruction had no direct, positive relation to students' writing achievement. This finding was somewhat surprising given the evidence that both skills and composing instruction strengthen writing skills, increase text length, and improve

writing quality (Graham, Bollinger, et al., 2012; Graham, et al., 2015; Graham & Perin, 2007; Graham & Santangelo, 2014; Santangelo & Graham, 2016) and that combining both may be effective, especially for students with writing difficulties (Limpo & Alves, 2017). There are several explanations for this unexpected finding.

The lack of impact could be attributed to the amount of writing instruction that was observed. Our observations revealed that on average teachers provided about 5 blocks of instruction each day (about 25 minutes). However, there was substantial variation around the average as the range extended from 1.1-14.85 blocks (or about 5.50 to 74.25 minutes). In fact, there was no writing instruction at all in 11% of our day-long observations.

Another explanation is that the type of writing instruction that teachers delivered was not effective. This could be explained by the lack of curricular guidance. For example, close to half of the teachers had no writing curriculum (46%), and a similar percentage indicated that their writing curriculum was part of the reading curriculum (44%). A small group (10%) had access to a supplementary curriculum that targeted nonfiction writing, but it is not clear whether those lessons would benefit a broader range of writing tasks (Coker et al., 2016).

In addition to having access to limited curricular resources, many teachers may not have received sufficient preparation to teach writing in their education courses, which could have also limited the effectiveness of their instruction. Cutler and Graham (2008) found that among primary-grade teachers, 28% reported that their training in writing was inadequate or poor. Nearly two thirds of teachers in grades 4-8 teachers responded that they received little or no training in writing instruction in their coursework (Gilbert & Graham, 2010).

Surprisingly, the analysis did reveal two negative effects of writing instruction. Instruction in writing skills was negatively related to Quality/Length, and composition

instruction was negatively related to students' Contextual Spelling. The magnitude of these effects was fairly small. Several potential reasons exist for why these relations were found. It may be that in classrooms where skills instruction was more common, teachers did not prioritize composing, which might have impacted students' Quality/Length scores. Similarly, in classrooms that prioritized composing, there might have been less attention to correct spelling, which might have resulted in lower Spelling scores. Additionally, the instruction that was provided may not have been sufficiently strong to move student scores. All of these potential explanations deserve future investigation.

Although no positive, direct relations were detected between writing instruction and achievement, our analysis did reveal that writing instruction moderated the association between student demographic variables and writing achievement. These relations are discussed in a subsequent section.

Does Writing Practice Predict Achievement?

We also investigated the association between the amount and type of writing practice and student achievement. Writing practice had a positive, significant relation to achievement. However, only one type of writing practice, generative writing, was a significant predictor; other forms of writing practice were not significantly related to writing achievement. Generative writing practice was related to higher BWL scores as well as higher Quality/Length and Contextual Spelling factor scores based on narrative and descriptive writing tasks. The consistent effect of generative writing practice across the three writing outcomes suggests that its impact may be broad even if it is modest.

When students engaged in generative writing practice, they produced connected text of at least a sentence in length, and they were expected to create the content of the text. Generative

writing required students to form ideas, to transcribe those sentences onto paper, and to manage the processes of planning what to write, and potentially revising and editing the text. As a result, generative writing engaged students in several cognitive processes highlighted in the NSSVW (Berninger & Winn, 2006). In contrast, engaging in tasks such as correct/copying writing relies almost exclusively on transcription skills of handwriting and spelling because this type of writing practice does not require writers to create content or to manage the challenges associated with producing longer text. It may be that having opportunities to integrate multiple cognitive processes when writing contributes to writing growth. Engaging in tasks that integrate these processes may be more beneficial than forms of writing practice that are less cognitively demanding.

This finding contributes to the body of evidence pointing to the importance of writing practice, and it suggests that some forms of practice may be more beneficial than others. The authors of the IES Practice Guide recommend providing at least 30 minutes of daily writing time for students beginning in first grade, even though there is relatively little direct evidence for this recommendation (Graham, Bollinger, et al., 2012). Indirect evidence for the importance of writing practice can be found in the design of nearly every effective writing intervention because writing practice is a central instructional component.

However, there is sparse research on the potential benefits of different types of student writing practice. A recent study indicated that generative writing practice was related to student reading achievement (Coker et al., 2018). However, in that analysis, generative writing practice mediated the association between composing instruction and first-grade reading achievement. Although the relation was small, this mediated association was found even when reading instruction was controlled. Taken together, the findings of these studies on generative writing

practice suggest that it may be a useful practice for literacy development and certainly deserves more investigation.

Does the Impact of Writing Instruction or Practice Depend on Student Factors?

Our results further revealed that the impact of writing instruction and practice on achievement also depended on two specific student demographic factors, ethnicity and gender. Minority students were predicted to have higher scores on both BWL and Quality/Length in classrooms that had more than the average amount of skills instruction than average. Minority status was also found to interact with generative writing practice; however, in this association minority students had lower BWL scores when they were in classrooms with more than the average amount of generative writing practice. Gender also interacted with generative writing practice such that boys had higher Quality/Length scores in classrooms with more than the average amount generative writing. It is important to note that these relations with student characteristics remained significant even after controlling for students' fall transcription skill (spelling and handwriting fluency), vocabulary knowledge, reading skill, and writing and reading instruction and practice.

Based on our data, it is unclear what might explain these interactions, but it seems likely that the relations can be explained by cognitive or environmental factors that were not assessed in this study. For example, other research with young students has revealed that factors such as attention (Kent, Wanzek, Petscher, Al Otaiba, & Kim, 2014; Kim et al., 2013), additional components of oral language, including grammatical knowledge (Kent & Wanzek, 2016; Kim et al., 2013), and familial socio-economic status (Coker, 2006; Kim et al., 2015) are all related to writing achievement. Even though other researchers have identified writing differences related to

gender (Berninger & Fuller, 1992; Kim et al., 2015) and ethnic status (Coker, 2006; Persky et al., 2003), it may be that these demographic variables are simply proxies for other factors.

These findings may provide further support for the potential benefits of targeted writing instruction, and they suggest the potential advantages of differentiating instruction and practice. Previous research has identified subtypes of writers based on their performance on a range of cognitive and linguistic tasks (Coker et al., 2018b; Roid, 1994; Hooper, Wakely, de Kruif, & Swartz, 2006; Wakely, Hooper, de Kruif, & Swartz, 2006). Furthermore, some of these subtypes have shown a differential response to intervention (Hooper et al., 2006). While the results of this study offer no evidence that students might respond differentially to instruction, this remains an important area for future research.

Overall, the findings of this analysis provide further support for the role of both cognitive and sociocultural forces in early writing development. In the level-one models, many of the cognitive predictors support the NSSVW. In addition, the results at level two provide insight into potential ways that classroom factors—both instruction and practice—may be related to writing development. Perhaps generative writing practice offers young writers a contextualized task (or activity system) that could strengthen writers' knowledge and skills (Russell, 1997). While this possibility is intriguing, much more work is needed to understand the ways that cognitive skills and the classroom contexts might interact to facilitate development.

Instructional Implications

The results of this study have several instructional implications worth considering. Our models indicated that writing instruction did not have a direct, positive effect on achievement. Clearly, more attention should be devoted to providing effective writing instruction for students. In schools, teachers should draw from instructional approaches that have research support (e.g.,

Graham, Bollinger, et al., 2012). Furthermore, administrators should prioritize scheduling daily writing instruction.

In addition, schools should consider providing more opportunities for student writing practice. Previous summary reports and meta-analyses have recommended time for student writing (Graham, Bollinger, et al., 2012; Graham & Perin, 2007). However, these results put a finer point on those suggestions and provide intriguing data that the type of practice may matter. In general, generative writing practice may be more supportive of students' writing progress. However, the interaction results suggest that for some students, minority students or boys, specific forms of instruction and practice may be more or less helpful.

Limitations

In a complex observational study such as this one, there are potential sources of error (Raudenbush & Sadoff, 2008). These sources include non-random assignment of observers to classrooms, and the time during the year when observations were scheduled. Our attempts to minimize these limitations included the use of four, day-long observations, rigorous training of observers, and the use of a small number of observers. In the future, these limitations might be mitigated by increasing the number of observations to strengthen the reliability of the estimates. Additionally, the use of classroom video could be employed to increase the precision of observations. Finally, randomly assigning observers to observations and randomly scheduling observation sessions could also be used to reduce bias.

Another potential limitation was our use of an average measure of instruction and student practice. This decision was made because our goal was to describe typical classroom instruction in first grade, but we recognize that variability was lost with the measure. Similarly, our observers recorded the type and length of writing instruction and practice. Other measures of

instructional quality could be used in the future to capture elements that may be related to student achievement such as teacher responsiveness (Kim et al., 2013).

In these models we attempted to include control variables for important literacy skills. A substantial amount of the variation at the student-level was explained in the BWL models; however, considerably less was accounted for in the other two models. Additional measures of linguistic knowledge and cognitive skills such as syntactic knowledge, attention and self-regulation might explain a greater proportion of the student-level variance in those models (Kim & Schatschneider, 2017; Hooper et al., 2011).

Conclusion

The results of this study offered reasons for both concern and optimism about the effectiveness of writing instruction and student practice in first grade. Our findings signal that current instructional efforts may not be effective in supporting students' writing growth. Given the importance of writing in the CCSS, teachers may want to reevaluate their own approach to early writing instruction to take advantage of evidence-based practices. On a more positive note, this study identified a specific type of writing practice, generative writing, that was associated with stronger student writing across multiple assessments. There may be value in engaging students in this type of writing practice. Finally, there was evidence that some forms of instruction and practice may have benefits for specific groups of students, suggesting that efforts to differentiate instruction and practice may be beneficial for students.

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of Educational Psychology*, *85*(3), 478–508.

Afflerbach, P., Blachowicz, C., Boyd, C. D., Cheyney, W., Juel, C., Kame'enui, E., …Wixon, K. K. (2011). *Reading Street*. Glenview, IL: Pearson Scott Foresman.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5-37.

Baumann, J. F., Chard, D. J., Cooks, J., Cooper, J. D., Gersten, R., Lipson, M., … Vogt, M. (2011). *Journeys*. Orlando, FL: Houghton Mifflin Harcourt.

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*, 39–56.

Berninger V. W., Fuller F. (1992). Gender differences in orthographic, verbal, and compositional fluency: Implications for diagnosis of writing disabilities in primary grade children. *Journal of School Psychology*, *30*, 363–382.

Berninger, V. W., & Rutberg, J. (1992). Relationship of finger function to beginning writing: Application to diagnosis of writing disabilities. *Developmental Medicine & Child Neurology*, *34*(3), 198–215.

Berninger, V. W., & Swanson, H.L. (1994). Modifying Hayes & Flower's model of skilled writing to explain beginning and developing writing. In E. Butterfield (Ed.), *Children's writing: Toward a process theory of development of skilled writing* (pp. 57-81). Greenwich, CT: JAI Press.

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

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis, 1*(3), 473-514. doi: 10.1214/06-BA117

Coker, D. (2006). The impact of first-grade factors on the growth and outcomes of urban schoolchildren's primary-grade writing. *Journal of Educational Psychology, 98*, 471-488.

Coker, D. L., Farley-Ripple, E., Jackson, A. F., Wen, H., MacArthur, C. A., & Jennings, A. S. (2016). Writing instruction in first grade: An observational study. *Reading and Writing, 29*(5), 793–832. http://doi.org/10.1007/s11145-015-9596-6

Coker, D. L., Jennings, A. S., Farley-Ripple, E., & MacArthur, C.A. (2018). The type of writing instruction and practice matters: The direct and indirect effects of writing instruction and student practice on reading achievement. *Journal of Educational Psychology, 110*(4), 502-517. http://dx.doi.org/10.1037/edu0000232

Coker, D. L., Ritchey, K. D., Uribe-Zarain, X., & Jennings, A. S. (2018b). An analysis of first-grade writing profiles and their relationship to compositional quality. *Journal of Learning Disabilities, 51*(4), 336–350. https://doi.org/10.1177/0022219417708171

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child x instruction interactions. *Journal of Educational Psychology, 96*(4), 682–698. https://doi.org/10.1037/0022-0663.96.4.682

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology, 100*(4), 907–919.

Dunn., L., M., & Dunn, D. M., (2007). *The Peabody Picture Vocabulary Test, Fourth Edition*. Bloomington, MN: NCS Pearson, Inc.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., … others. (2016). Foundational skills to support reading for understanding in kindergarten through 3rd grade. Educator's practice guide. NCEE 2016-4008. *National Center for Education Evaluation and Regional Assistance*. Retrieved from http://eric.ed.gov/?id=ED566956

Foorman, B., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*, 1-29.

Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4–6: A national survey. *The Elementary School Journal, 110*(4), 494–518.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*, 170-182. http://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* (No. 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_reviews.aspx#pubsearch

Graham, S., & Harris, K. R. (2015). Common Core State Standards and Writing: Introduction to the Special Issue. *The Elementary School Journal*, *115*(4), 457–463. https://doi.org/10.1086/681963

Graham, S., Harris, & Santangelo, T. (2015). Research-based writing practices and the Common Core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, *115*(4), 498–522. https://doi.org/10.1086/681964

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, *81*(4), 710–744.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*(4), 879–896. http://doi.org/10.1037/a0029185

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445-476. doi: 10.1037/0022-0663.99.3.445.

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, *27*(9), 1703–1743. http://doi.org/10.1007/s11145-014-9517-0

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal*, *43*(2), 295–340. https://doi.org/10.3102/00028312043002295

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York, NY: Routledge.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Education Evaluation and Policy Analysis*, *29*(1), 60-87. doi: 10.3102/0162373707299706

Hoffman, J. V., Sailors, M., Duffy, G. R., & Beretvas, S. N. (2004). The Effective Elementary Classroom Literacy Environment: Examining the Validity of the TEX-IN3 Observation System. *Journal of Literacy Research*, *36*(3), 303-334. DOI:10.1207/s15548430jlr3603_3

Hofstetter, F. (2016). ISeeNcode (Version 1.2) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/iseencode/id1084763871?mt=8

Hooper, S., Costa, L.-J., McBee, M., Anderson, K., Yerby, D., Knuth, S., & Childress, A. (2011). Concurrent and longitudinal neuropsychological contributors to written language expression in first and second grade students. *Reading and Writing*, *24*(2), 221–252. https://doi.org/10.1007/s11145-010-9263-x

Hooper, S. R., Wakely, M. B., de Kruif, R. E. L., & Swartz, C. W. (2006). Aptitude-treatment interactions revisited: Effect of metacognitive intervention on subtypes of written expression in elementary school students. *Developmental Neuropsychology*, *29*(1), 217–241. http://doi.org/10.1207/s15326942dn2901_11

Jones, D., & Christensen, C. A. (1999). Relationship between automaticity in handwriting and students' ability to generate written text. *Journal of Educational Psychology*, *91*(1), 44–49.

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.

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S., & Kim, Y.-S. (2014). Writing fluency and quality in kindergarten and first grade: the role of attention, reading, transcription, and oral language. *Reading and Writing*, *27*, 1163–1188. https://doi.org/10.1007/s11145-013-9480-1

Kim, Y.-S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014). Evaluating the dimensionality of first-grade written composition. *Journal of Speech, Language, and Hearing Research*, *57*(1), 199–211. https://doi.org/10.1044/1092-4388(2013/12-0152

Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., & Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergartners. *Reading and Writing*, *27*(2), 237–253. https://doi.org/10.1007/s11145-013-9440-9

Kim, Y.-S., Al Otaiba, S., Sidler, J. F., & Gruelich, L. (2013). Language, literacy, attentional behaviors, and instructional quality predictors of written composition for first graders. *Early Childhood Research Quarterly*, *28*(3), 461–469. http://doi.org/10.1016/j.ecresq.2013.01.001

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology*, *107*(1), 79–95. https://doi.org/10.1037/a0037210

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

Lerkkanen, M., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. (2004). The developmental dynamics of literacy skills during the first grade. *Educational Psychology*, *24*(6), 793–810. http://doi.org/10.1080/0144341042000271782

Lockhard, C., & Eversole, L. (2006). *Discover intensive phonics for yourself* (3rd ed.). North Salt Lake, UT: Reading Horizons.

Limpo, T., & Alves, R. A. (2017) Tailoring multicomponent writing interventions: Effects of self-regulation and transcription training. *Journal of Learning Disabilities*. Advance online publication. DOI: 10.1177/0022219417708170

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*(1), 13–23. https://doi.org/10.1207/S15326985EP3501_3

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual*. *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86-92. doi: 10.1027/1614-1881.1.3.86

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects K–5. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C. Retrieved April 23 2017, from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Persky, H. R., Daane, M. C., & Jin, Y. (2003). The nation's report card: Writing 2002. U. S. Department of Education, Institute of Education Sciences. Washington, DC: National Center for Education Statistics.

Puranik, C. S., Al Otaiba, S., Sidler, J. F., & Greulich, L. (2014). Exploring the amount and type of writing instruction during language arts instruction in kindergarten classrooms. *Reading and Writing, 27*(2), 213-236. http://doi.org/10.1007/s11145-013-9441-8

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., Jr., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S. W. & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness, 1*, 138-154. DOI: 10.1080/19345740801982104

Richards, S., Sturm, J., & Cali, K. (2012). Writing instruction in elementary classrooms: Making the connection to Common Core State Standards. *Seminars in Speech and Language*, *33*(02), 130–145. https://doi.org/10.1055/s-0032-1310313

Roid, G. H. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education, 7*(2), 159-170.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons, Inc.

Russell, D. R. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication*, *14*(4), 504–554.

Santangelo, T., & Graham, S. (2016). A comprehensive meta-analysis of handwriting instruction. *Educational Psychology Review*, *28*(2), 225–265. https://doi.org/10.1007/s10648-015-9335-1

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*(1), 19-25.

Shanahan, T. (2015). Common Core State Standards: A new role for writing. *The Elementary School Journal*, *115*(4), 464–479. https://doi.org/10.1086/681130

Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly, 29*, 104-122.

Slavin, R., Madden, N., & Karweit, N. (1989). *Effective programs for students at-risk: Conclusions for practice and policy*. In R. Slavin, N. Karweit, & N. Madden (Eds.), Effective programs for students at-risk (pp. 21–54). Boston: Allyn & Bacon.

Stead, T. & Hoyt, L. (2011). *Explorations in Nonfiction Writing: Grade 1*. Portsmouth, NH: Firsthand Heinemann.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal*, 3–28.

Troia, G.A., Olinghouse, N.G., Wilson, J., Stewart, K. A., Mo, Y., Hawkins, L., & Kopke, R. A. (2016). The Common Core Writing Standards: A descriptive study of content and alignment with a sample of former state standards. *Reading Horizons*, *55*(3), 99-141.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2), 203–220. https://doi.org/10.1007/s11145-010-9266-7

Wakely, M. B., Hooper, S. R., de Kruif, R. E. L., & Swartz, C. (2006). Subtypes of written expression in elementary school children: A linguistic-based model. *Developmental Neuropsychology, 29* (1), 125–159. http://doi.org/10.1207/s15326942dn2901_7

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.

Zumbrunn, S., & Bruning, R. (2013). Improving the writing and knowledge of emergent writers: The effects of self-regulated strategy development. *Reading and Writing: An Interdisciplinary Journal, 16*, 91-110. doi: 10.1007/s11145-012-93

Table 1: *Descriptive Statistics and Bivariate Correlations for Within-Group (Student) Measures*

Variable	*Reliability*	M	SD	1	2	3	4	5	6	7	8	9
1. Female	N/A	0.52	0.50	1								
2. Minority Status	N/A	0.49	0.50	.06	1							
3. Vocabulary*	.96	101.89	13.91	-.06	-.40	1						
4. Handwriting Fluency	1.00†	13.32	6.22	.19	-.10	.32	1					
5. Spelling*	.85	106.31	13.77	.02	-.03	.33	.50	1				
6. Basic Reading*	.89††	110.97	11.81	-.07	-.05	.38	.45	.83	1			
7. WJ Broad Written Language*	.84††	111.70	15.13	.08	-.09	.41	.53	.76	.78	1		
8. Writing Quality/Length	.96‡	0.00	1.00	.19	-.04	.32	.45	.52	.53	.62	1	
9. Contextualized Spelling	.99‡	0.00	1.00	.09	.05	.18	.36	.57	.52	.58	.63	1

Notes: sample-based reliability reported as Cronbach's alpha unless otherwise indicated; mean and standard deviation presented for non-centered variables; bivariate correlations calculated based on the maximum-likelihood estimated within covariance matrix for variables as centered in subsequent analyses. *Standard scores reported for interpretability; W and raw PPVT-4 scores used in HLM analyses. †Inter-scorer agreement. ††Composite measure; value represents the lowest reliability of component measures. ‡ Standardized factor score; value represents the lowest inter-scorer agreement of manifest variables.

Table 2: *Descriptive Statistics and Bivariate Correlations for Between-Group (Classroom) Measures*

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
1. Skills Writing Instruction	1.84	1.44	1									
2. Composing Writing Instruction	3.11	2.04	-.10	1								
3. Total Reading Instruction	17.54	4.14	-.25	.13	1							
4. Correct/Copy Practice	19.00	5.61	.21	-.11	.09	1						
5. Writing About Text Practice	3.20	2.69	-.41	.11	.14	-.15	1					
6. Generative Writing Practice	6.78	4.04	-.19	.31	.15	-.17	.31	1				
7. Total Reading Practice	13.96	5.14	-.30	.05	.51	.26	.37	.40	1			
8. WJ Broad Written Language	*	*	-.17	.04	.18	-.18	.18	.40	.35	1		
9. Writing Quality/Length	*	*	-.25	-.05	.37	-.07	.06	.36	.42	.85	1	
10. Contextualized Spelling	*	*	-.28	-.13	-.01	-.28	.22	.29	.24	.94	.72	1

Notes: $N = 50$; mean and standard deviation measured as number of 5-minute instructional blocks and presented for non-centered variables; bivariate correlations calculated based on the maximum-likelihood estimated between covariance matrix for variables as centered in subsequent analyses. All classroom instruction and practice variables represent 5-minute observation blocks per classroom.
*Values presented in Table 1 as these variables are measured at the within-group (student) level.

Table 3: *Broad Written Language Model Specifications*

Variable	Model 1	Model 2	Model 3	Model 4
Level 1: Student				
Gender (female)		**1.82(0.70)****	**1.82(0.70)****	**1.72(0.71)***
Minority Status (non-white)		-0.15(0.92)	-0.15(0.92)	-0.35(0.79)
Vocabulary		**0.07(0.02)****	**0.07(0.02)****	**0.06(0.02)****
Handwriting Fluency		**0.26(0.06)*****	**0.26(0.06)*****	**0.26(0.06)*****
Spelling		**0.22(0.04)*****	**0.22(0.04)*****	**0.21(0.04)*****
Basic Reading		**0.25(0.03)*****	**0.25(0.03)*****	**0.27(0.03)*****
Level-2: Classroom				
Skills Writing Instruction			0.02(0.50)	0.02(0.50)
Composing Writing Instruction			-0.55(0.34)	-0.55(0.34)
Total Reading Instruction			0.37(0.34)	0.37(0.33)
Correct/Copy Writing Practice			-0.37(0.24)	-0.37(0.24)
Writing About Text Practice			0.00(0.31)	0.00(0.31)
Generative Writing Practice			**0.75(0.36)***	**0.75(0.36)***
Total Reading Practice			0.45(0.24)	0.45(0.24)
Cross-level Interactions				
*Female*Skills Instruction*				0.75(0.51)
*Female*Composing Instruction*				-0.15(0.30)
*Female*Generative Writing*				0.04(0.20)
*Minority*Skills Instruction*				**1.45(0.49)****
*Minority *Composing Instruction*				0.16(0.44)
*Minority *Generative Writing*				**-0.42(0.20)***
Fixed Effects				
Classroom conditions	No	No	Yes	Yes
School	No	No	Yes	Yes
Variance Components				
Student-Level	**125.66**	**36.79**	**36.81**	**35.68**
Classroom-Level	**18.15*****	**28.86*****	**23.86*****	**24.00*****

Notes: parameter estimates, robust standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 4: *Narrative/Descriptive Writing Quality/Length Model Specifications*

	Model 1	Model 2	Model 3	Model 4
Level 1: Student				
Gender (female)		**0.35(0.08)*** **	**0.35(0.08)*** **	**0.34(0.08)*** **
Minority Status (non-white)		0.09(0.11)	0.09(0.11)	0.06(0.10)
Vocabulary		**0.01(0.00)***	**0.01(0.00)***	**0.01(0.00)***
Handwriting Fluency		**0.03(0.01)*** **	**0.03(0.01)*** **	**0.03(0.01)*** **
Spelling		0.01(0.00)	0.01(0.00)	0.01(0.00)
Basic Reading		**0.02(0.00)*** **	**0.02(0.00)*** **	**0.02(0.00)*** **
Level-2: Classroom				
Skills Writing Instruction			**-0.09(0.04)***	**-0.09(0.04)***
Composing Writing Instruction			-0.04(0.02)	-0.04(0.02)
Total Reading Instruction			0.01(0.02)	0.01(0.02)
Correct/Copy Writing Practice			0.01(0.01)	0.01(0.01)
Writing About Text Practice			0.00(0.02)	0.00(0.02)
Generative Writing Practice			**0.05(0.02)** **	**0.05(0.02)** **
Total Reading Practice			0.01(0.01)	0.01(0.01)
Cross-level Interactions				
*Female*Skills Instruction*				-0.06(0.05)
*Female*Composing Instruction*				0.03(0.03)
*Female*Generative Writing*				**-0.06(0.02)*** **
*Minority*Skills Instruction*				**0.12(0.06)***
*Minority *Composing Instruction*				0.05(0.04)
*Minority *Generative Writing*				-0.04(0.03)
Fixed Effects				
Classroom conditions	No	No	Yes	Yes
School	No	No	Yes	Yes
Variance Components				
Student-Level	**0.93**	**0.56**	**0.56**	**0.55**
Classroom-Level	**0.07** **	**0.11*** **	**0.04***	**0.05***

Notes: parameter estimates, robust standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 5: *Narrative/Descriptive Spelling Model Specifications*

	Model 1	Model 2	Model 3
Level 1: Student			
Gender (female)		**0.13(0.07)***	**0.13(0.07)***
Minority Status (non-white)		0.15(0.10)	0.15(0.10)
Vocabulary		0.00(0.00)	0.00(0.00)
Handwriting Fluency		**0.01(0.01)***	**0.01(0.01)***
Spelling		**0.03(0.00)*****	**0.03(0.00)*****
Basic Reading		**0.01(0.00)***	**0.01(0.00)***
Level-2: Classroom			
Skills Writing Instruction			-0.06(0.04)
Composing Writing Instruction			**-0.05(0.02)***
Total Reading Instruction			0.03(0.02)
Correct/Copy Writing Practice			-0.02(0.01)
Writing About Text Practice			0.01(0.02)
Generative Writing Practice			**0.06(0.02)***
Total Reading Practice			0.02(0.02)
Fixed Effects			
Classroom conditions	No	No	Yes
School	No	No	Yes
Variance Components			
Student-Level	**0.94**	**0.60**	**0.60**
Classroom-Level	**0.06***	**0.10*****	**0.08*****

Notes: parameter estimates, robust standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$