The type of writing instruction and practice matters: The direct and indirect effects of writing instruction and student practice on reading achievement

David L. Coker, Jr.

Austin S. Jennings

Elizabeth Farley-Ripple

Charles A. MacArthur

University of Delaware

Published 2018

Journal of Educational Psychology

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

Previous research has demonstrated that writing instruction can support reading achievement (Graham & Hebert, 2011); however much of this work involved carefully designed interventions. In this study, we evaluated a conceptual framework of the direct and indirect effects of typical writing instruction and student writing practice on reading achievement in first grade. Fall reading, vocabulary, and writing data were collected from 391 students, and classroom writing instruction and student writing practice were observed in 50 classrooms. The effects of writing instruction on spring reading achievement were evaluated using a two-level, fixed effects structural equation model. In a multiple mediator model, the total indirect effect of composing writing instruction through student writing practice on spring reading achievement was positive and statistically significant ($\beta = .17$, $p = .029$), with a modest effect of composing writing instruction mediated by generative writing practice ($\beta = .15$, $p = .024$). The final model explained 86% and 59% of the variability in spring reading achievement at the student and classroom levels, respectively. These results suggest that generative writing practice mediates the relationship between composing instruction and spring reading achievement. The results also highlight some potentially positive effects of typical writing instruction and student writing practice after controlling for reading instruction and fall reading achievement.

Keywords: writing instruction, reading, first grade

Educational Impact and Implications Statement

This study suggests that typical, first-grade, composing instruction combined with opportunities for students to compose may contribute to reading achievement. The findings provide support for efforts to combine writing instruction and practice and to integrate both with reading instruction.

During first grade, students experience significant literacy development that is crucial for future success (Snow, Burns, & Griffin, 1998). The authors of the Common Core State Standards (CCSS) recognized the importance of first-grade reading growth and set expectations to ensure that students become independent readers by the end of first grade (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). For example, in the CCSS literature standards, first-grade students are expected to "read with sufficient accuracy and fluency to support comprehension" (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010, p. 16).

To help students meet these standards, researchers have devoted considerable attention to understanding the predictors of reading success and developing effective interventions (e.g., Adams, 1994; Foorman et al., 2016; National Institute of Child Health and Human Development [NICHD], 2000; Snow et al., 1998). Despite these efforts, many students continue to struggle with reading. Results from the 2015 National Assessment of Educational Progress reveal that only 36% of fourth-grade students achieve at or above the level of proficiency in reading, with even lower levels of achievement for students who are Black, Hispanic, or have disabilities (National Center for Educational Statistics [NCES], n.d.). Early reading difficulties are concerning because it is estimated that as many as three-fourths of primary-grade students at risk of reading difficulties will continue to have reading problems in the future (Juel, 1988; Scarborough, 1998).

One under-explored approach to developing early reading skills is writing instruction. Recent reviews of carefully designed writing interventions have demonstrated a positive impact on a range or reading outcomes across grade levels (Graham & Hebert, 2011; Graham &

Santangelo, 2014). Furthermore, there is emerging evidence that instruction in both writing skills, such as spelling and handwriting, and composing can strengthen young students' reading achievement (Adams-Boateng, 2001; Conrad, 2008; Denner, McGinley, & Brown, 1989; Ouellette & Sénéchal, 2008; Ouellette, Sénéchal, & Haley, 2013; Sénéchal, Ouellette, Pagan, & Lever, 2012; Uhry & Shepherd, 1993).

Given the importance of early reading success, we were interested in how typical writing instruction and practice could contribute to global reading achievement during first grade when students are in the early phases of learning to write and read. The relationship between writing instruction, practice, and reading has received little attention, and this relationship is complex for several reasons. First, the targets of writing instruction can vary considerably from focusing on skills like handwriting, spelling, and grammar to composing various types of text. Different types of writing instruction might be expected to have different effects of reading. Furthermore, writing instruction can be combined with practice in many ways; for example, a spelling lesson might be followed by a practice worksheet or opportunities to write a story. The interactions between the focus of writing instruction and the type practice might alter the effect on reading achievement. The overarching goal of this study is to explore this complex relationship by investigating the impact of different forms of writing instruction on reading and by testing whether the effect of writing instruction is mediated by the type of writing practice.

Conceptual Framework

Our understanding of how writing instruction can impact reading is based on the nature of global reading achievement in first grade and the types of writing instruction and practice that may impact reading. Figure 1 depicts the theorized relationships between reading achievement

and two types of writing instruction (skills and composing), two types of writing practice (correct/copy and generative writing), and reading instruction.

Reading achievement. In the current study, reading achievement is conceptualized as a global process that includes word reading, fluency and reading comprehension. Theoretical accounts have characterized reading as a multidimensional process that depends on letter- and word-reading skills related to fluent decoding and oral language skills (Gough & Tunmer, 1986). In first grade, the print skills related to decoding have been identified as powerful reading predictors. In some studies, print skills are the sole predictors of reading (Storch & Whitehurst, 2002; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Other researchers have found oral language skills may explain some variance in comprehension after accounting for print skills (Foorman, Herrera, Patscher, Mitchell, & Truckenmiller, 2015; Muter, Hulme, Snowling, & Stevenson, 2004). In all of these studies, decoding and print skills are strongly predictive of first-grade reading. As a result, it is difficult to separate reading comprehension from fluent decoding, except through targeted assessments of oral language skills (Foorman et al., 2015; Muter et al., 2004). To account for the inter-related nature of reading skills in first grade, we followed the example of other researchers who operationalized early reading as a global measure (e.g., Lerkkanen, Rasku-Puttonen, Aunola, & Nurmi, 2004; Oslund et al., 2015).

Theoretical foundation. Our conceptual framework of the impact of writing instruction and practice on reading is grounded in the simple view of writing. Juel and her colleagues (Juel, Griffith, & Gough, 1986; Juel, 1988) described writing quality as dependent on spelling and idea generation. Berninger and her colleagues extended the simple view based on empirical studies with typically achieving and learning disabled students (Berninger, 2000; Berninger, Fuller, & Whitaker, 1996; Berninger, & Swanson, 1994). Berninger and Swanson (1994) identified

transcription, which involves handwriting (or typing) and spelling, as central for young writers. Idea generation is understood as the process of "turning ideas into words, sentences, and larger units of discourse" (McCutchen, 2006, p. 123) and is distinct from transcription processes (Berninger & Swanson, 1994). Both of these processes contribute to writing quality.

Drawing on these models of early writing and empirical research on writing development, we have identified two types of writing instruction that are important to the development of early reading achievement—skills instruction targeting transcription skills, and composing instruction that addresses text generation and executive functions.

Writing Instruction. Skills instruction includes handwriting, spelling, and mechanics instruction. Due to the focus on transcription skills, skills instruction is likely to strengthen students' knowledge of print concepts, letter names, the alphabetic principle, and the orthographic patterns of phonetically irregular words, all of which are important for reading development (Adams, 1994; Snow et al., 1998). Skills instruction would then impact reading by improving students' decoding skills.

The second type of writing instruction that may be related to reading involves the creation of longer texts and is described as composing instruction. Composing instruction would address the challenges associated with generating ideas for writing (text generation) and the self-regulatory processes needed to plan, produce, and revise text (executive functions). Strengthening students' composing is likely to impact reading achievement through several paths. First, composing requires the application of both transcription skills and text generation. The transcription skills would contribute to decoding, as described previously. Instruction in text generation could strengthen a range of knowledge sources that could contribute to reading

achievement, including background knowledge, oral language skills, text structure knowledge, and executive functions (Foorman et al., 2016; NICHD, 2000; Snow et al., 1998).

Writing Practice. In addition to writing instruction, opportunities for students to practice writing may also play a role in their reading achievement by giving students opportunities to practice what they have learned and to build confidence in their skills (Graham et al., 2012). Specifically, two types of writing practice may facilitate reading achievement--correct/copy and generative writing. In correct/copy tasks students either copy words or write single-word responses with a definite correct answer. Examples of correct/copy tasks include practice spelling words, filling in blanks in worksheets, and copying words from the board. Engaging in correct/copy tasks provides practice with transcription skills such as handwriting and spelling (Berninger, & Swanson, 1994). This type of practice may bolster phonological and orthographic knowledge which in turn, may support decoding achievement.

In contrast, generative writing practice provides opportunities for students to write longer texts that require content generation and syntactical knowledge. Generative writing tasks may include writing a personal story, composing a paragraph about a science topic, or reflecting on a common classroom experience. Generative writing tasks would draw on a full range of writing skills and knowledge, such as transcription skills, vocabulary and syntactic knowledge to represent ideas in sentences, discourse knowledge to represent genres, and even process knowledge to plan and revise text (Berninger, & Swanson, 1994). Consequently, generative writing practice might strengthen both word reading and text comprehension. The transcription demands of generative writing practice might support word reading, much like correct/copy practice. The composing demands of generative writing practice might strengthen text reading and comprehension by supporting semantic, syntactical, discourse, and even process knowledge.

It is possible that the kind of writing practice that students experience may mediate the effect of writing instruction (Figure 1). Mediation may occur as specific types of practice align with the instructional focus in ways that build students' skills and confidence. For example, simply providing instruction in how to plan a narrative may provide some benefits for students. However, creating practice opportunities for students to plan their own narratives may extend and enhance the impact of writing instruction. Instruction and practice planning narratives would be likely to strengthen students' knowledge of narrative structure, which could support their reading.

Evidence of the Impact of Writing Instruction on Reading

There is growing evidence that various forms of writing skills instruction can contribute to student reading achievement. Limited work has been conducted on handwriting instruction, but there is some evidence that it can improve word reading (Berninger et al., 1997). More evidence for the impact of spelling instruction on word reading exists for students in grades k-2 (e.g., Berninger et al., 1998; Conrad, 2008; Ehri & Wilce, 1987; Fuchs et al., 2006; Graham, Harris, & Chorzempa, 2002; Sénéchal et al., 2012; Uhry & Shepherd, 1993). Research with struggling kindergarten students has signaled that various approaches to spelling instruction are associated with gains in word recognition (Berninger et al., 1998). Uhry and Shepherd (1993) investigated whether instruction in segmenting and spelling phonetically regular words was superior to a whole language approach to reading instruction in first grade. The intervention group demonstrated stronger word reading (of pseudo and real words) and reading fluency than the group that experienced typical reading instruction. Both spelling and decoding depend on shared knowledge of phonology and orthography, especially in the primary grades (Fitzgerald & Shanahan, 2000). Consequently, it is anticipated that instruction that strengthens spelling may

improve phonological, orthographic, and morphological knowledge and may result in stronger word reading and comprehension.

There is also evidence that various approaches to composing instruction are associated with reading improvements for young students (Adams-Boateng, 2001; Craig, 2006; Denner et al., 1989; Frey, 1993). Craig (2006) tested the impact of interactive writing instruction in kindergarten. This approach combined spelling instruction through word building activities with opportunities to learn about composing, text structure, and background information through cooperative writing and discussions of the text. The intervention was found to improve both word reading and reading comprehension. These effects may have extended beyond decoding because interactive writing addressed challenges associated with both transcribing words (e.g., spelling) and composing. Denner et al. (1989) tested a pre-reading activity in which first-grade students were given clues about a story and used those clues to write their own story. After reading the original story, their comprehension was assessed, and those in the story-writing group outperformed the control group. Another study that investigated the impact of journal writing found that having students write in journals after hearing a reading strengthened their reading comprehension more than extra reading instruction (Adams-Boateng, 2001). Taken together, there is emerging evidence that various forms of composing instruction may strengthen young students' reading.

Evidence of the Impact of Writing Practice on Reading

Although writing practice figures prominently in many instructional interventions, researchers have devoted less attention to it with young writers. One exception is spelling, where approaches to spelling practice have also been associated with improvements in reading outcomes. In a study with second-grade students, Conrad (2008) found that both practice spelling

and reading words transferred to the other modality. However, greater transfer was demonstrated from spelling to reading when transferring to new words with practiced rimes. Another line of recent work has found that teaching students to practice spelling using invented spelling (with and without feedback) has resulted in better gains in word reading than typical reading instruction (Ouellette & Sénéchal, 2008; Ouellette, et al., 2013; Sénéchal et al., 2012). We consider these types of spelling practice to be correct/copy tasks because students generate single words that have a single correct spelling.

Even less research has been conducted on other forms of writing practice. However, in a number of the intervention studies reviewed above, students engage in generative writing practice where they create new texts. For example, as part of an intervention, Craig (2006) had students engage in interactive writing activities with teachers, and Adams-Boateng (2001) asked students to compose extended texts. In these studies, it is not clear how specific types of student practice, such as copying specific words or having opportunities to generate connected text, may strengthen students' reading, but student practice is an integral part of those interventions and likely played a role in their efficacy.

Various forms of student writing practice are also common in typical primary-grade classrooms. Observations of kindergarten and first grade revealed more time for student writing practice than for writing instruction. Puranik, Al Otaiba, Sidler, & Greulich (2014) reported about 8 minutes of writing practice during the 90-minute literacy instructional block in the winter of kindergarten (range: 0-20.58 min). In first grade, Coker et al. (2016) observed some form of writing occurring for 125 minutes during the entire school day. Teachers in the primary grades provide time for students to write, but there is little evidence for how various types of writing practice—either

when associated with a writing intervention or simply provided by the teacher—would be beneficial for students (Graham, McKeown, Kiuhara, & Harris, 2012).

Typical Writing Instruction and Practice

The effects of writing instruction and practice on reading have been studied in researcher-designed interventions. However, typical classroom writing instruction may differ considerably from controlled interventions, which confound the potential effects of writing instruction and practice on reading. Converging data from survey research with teachers (Cutler & Graham, 2008) and observational research in kindergarten and first grade (Coker et al., 2016; Kim, Al Otaiba, Sidler, & Gruelich, 2013; Puranik et al., 2014) has suggested that only modest amounts of writing instruction occur and that large variation among classrooms exists. For example, estimates for the amount of writing instruction vary from 1 minute in the fall of kindergarten, (range: 0-8.86 min; Puranik et al., 2014), to about 26.4 minutes a day in first grade (range: 5.50-74.25 min; Coker et al., 2016), which was similar to the 21 minutes a day reported by teachers in the primary grades (Cutler & Graham, 2008). In terms of the nature of writing instruction, the majority of instructional time in kindergarten was devoted to handwriting (Puranik et al., 2014). In first grade, a more balanced instructional approach was reported—32.55% for skills instruction, 54.4% for composing instruction, which includes process writing, composition instruction, and sharing teacher and student writing, and small amounts for other approaches (Coker et al., 2016). A similar mix of composing and skills instruction was also reported by primary-grade teachers (Cutler & Graham, 2008).

In one study, Coker et al. (2016) coded the type of writing practice in first grade and found that 41% involved either copying or filling in an answer (correct/copy). These tasks did not require students to create new ideas. Alternately, 25% of the activities involved generative

writing, which required students to create the content of the text and generate sentences. Primary-grade teachers listed the types of writing their students completed by February of the school year, and a wide range of text types were produced—50% of the respondents listed 12 or more types of texts, suggesting a wide range in the types of tasks and texts produced by students (Cutler & Graham, 2008). Considering how much more varied typical classroom writing instruction and practice is compared to researcher-designed interventions, it is unclear whether typical writing instruction and practice would predict reading achievement.

The Current Study

The goal of this study was to investigate the direct and indirect effects of two types of typical writing instruction (skills and composing) and student writing practice (correct/copy and generative writing) on spring reading achievement in first grade. After controlling for reading achievement, vocabulary, and transcription skills in the fall, student demographics, and the total amount of reading instruction across the school year, we investigated the following research questions about writing instruction:

1. What are the direct effects of skills and composing writing instruction on global spring reading achievement in first grade?
2. Are the effects of writing instruction (skills and composing) on global reading achievement in first grade mediated by the type of student writing practice (correct/copy and generative writing)? Specifically,
    a. Is skills instruction mediated by correct/copy writing practice?
    b. Is composing instruction mediated by correct/copy and/or generative writing practice?

Two hypotheses guided this work. First, it seemed likely that skills and composing instruction would have direct effects on student reading achievement. This hypothesis was based on findings reviewed from intervention research indicating that both skills and composing instruction can improve reading achievement. A second hypothesis was that the effect of writing instruction would also be mediated by student writing practice. Furthermore, we hypothesized that the indirect effect of writing instruction through student writing practice would depend on the type of instruction and the type of practice. For example, it seemed likely that the effect of skills instruction might only be mediated by opportunities to practice writing or copying words (correct/copy) and not by other forms of writing practice. However, composing instruction might be mediated by both correct/copy and generative writing practice.

Considering that there is little relevant data pertinent to the second hypothesis and that fall achievement and reading instruction were included in the model, we predicted that any effects would be small but positive. However, given variation in the amount, type, and quality of writing instruction and practice in typical classrooms, it is possible that no effects would be found.

Method

The current study is part of a project designed to explore typical writing instruction and student literacy outcomes in first grade. As part of the larger project, a range of student and classroom assessments was collected. The observational data with this sample of teachers have been described in greater detail in a previous publication on the nature of first-grade writing instruction (Coker et al., 2016). Relationships among spring writing assessments with this sample of students have also been reported (Coker, Ritchey, Uribe-Zarain, & Jennings, 2017).

However, in this study we present a new analysis of fall student assessments and spring reading outcomes that have not been published elsewhere.

Participants

Students. Three hundred ninety-one first-grade students in a Mid-Atlantic state participated in the study across two school years. The students were drawn from 50 classrooms in 13 schools in three school districts that serve between 10,000-17,400 students in urban and suburban neighborhoods. The schools varied in size; the number of first-grade classrooms in each school ranged from two to six.

Students were invited to participate if their classroom teacher was part of the study. Participating students were evenly divided by gender (Female 51.9%), represented a range of ethnic backgrounds (White 50.6%, African American 28.6%, Hispanic 12.3%, Asian 4.9%, Other 3.3%), and included English Language Learners (8.7%) and students with disabilities (11.7%). Only school-level socio-economic status (SES) was provided by the school districts because the state department of education revised its method for calculating student SES before the second year of data collection. This policy change altered school-level SES statistics, even though there were no large demographic changes for the participating schools. To enable comparisons between schools from the two years of data collection, SES information from the first year of data collection was used. Just over half of the students (54.9%) in the participating schools qualified for free or reduced-price meals (range: 15.9%-84.8%).

Teachers. Within the schools, all first-grade teachers were invited to participate by the research team. Fifty-seven teachers volunteered from 50 first-grade classrooms. In four classrooms, the original classroom teachers were replaced with long-term substitutes, and three classrooms used a co-teaching model. Most teachers were female (96.3%) and White (90.7%);

however, there were two African-American teachers and one Hispanic teacher. Nearly half of the teachers held a master's degree (48.1%), but others held a bachelor's degree (14.8%), a bachelor's degree and additional coursework (18.5%), or a master's degree with additional coursework (18.5%). As a group, the teachers averaged 14.94 (*SD* = 7.98) years of teaching, with 8.69 (*SD* = 6.97) years of experience in first grade. Participating teachers were provided a $200 honorarium each semester they were observed.

Classroom Context

There were fewer than 22 students in each classroom. Although all schools were located in the same state, the adopted reading curricula varied across schools. Houghton Mifflin Harcourt's *Journey's* (*n* = 32; Baumann et al., 2011) was used in over half of the classrooms. Other classrooms used Pearson Scott Foresman's *Reading Street* (*n* = 5; Afflerbach et al., 2011), *Discover Intensive Phonics for Yourself* (*n* = 3; Lockhard & Eversole, 2006), or no published reading curriculum (*n* = 10). For writing instruction, the writing curriculum was integrated with the reading curriculum in 22 classrooms. Five teachers used an adaptable writing curriculum resource, *Explorations in Nonfiction Writing* (Stead & Hoyt, 2011). Approximately half of the teachers did not use a standard writing curriculum (*n* = 23).

Classroom Observations

Classroom observations were conducted in 50 classrooms over two years. Observations were conducted in 21 classrooms in five schools during the first year, and 29 classrooms from nine schools during the second year. There was no participant overlap; each teacher or student only participated in the data collection during a single year of data collection. Classrooms were selected for each year of the study based on their willingness to participate that year. Classrooms were observed four times during the first-grade year. Previous observational studies of

elementary literacy instruction have relied on three or fewer time points (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), and this number of observations enabled researchers to characterize typical instruction and to find meaningful correlations between literacy instruction and student achievement. Further, our observations covered the entire school day, allowing us to capture reading and writing instruction throughout the day rather than limiting our observations to literacy blocks, as is common in other observational studies (Foorman et al., 2006; Puranik et al., 2014; Silverman & Crandell, 2010; Taylor et al., 2003).

Scheduling observations. In each classroom, four day-long observations were conducted across the school year. They began in late October and were completed by the end of May with the average number of days between observations being 54.89. The research team made every effort to conduct the observations at equal intervals, but this effort was complicated by interruptions from school holidays, state testing, closures due to inclement weather, and special events at the schools. Teachers were informed of the observations in advance. The observations began with the start of the school day and ended when students were dismissed, and the observers continued coding whenever the class was engaged in academic content. The observers did not code during special classes (e.g., art, music, library, etc.), recess, and lunch.

Coding system. The coding system used a time-sampling procedure to capture the presence of instructional and practice activities at the individual, small group, or whole-class level. All academic instruction was observed for each classroom and codes were recorded every 5 minutes, so that the variables represent the approximate time that these activities were observed. Codes were not mutually exclusive such that more than one code could be entered in a

five-minute block. Observers were assigned to specific observations based on their availability and the logistics of travel; therefore, classrooms were not nested in observers. During the observations, observers watched classroom instruction and took field notes as needed for three minutes, and then they coded for two minutes. The coding was facilitated by an iPad application called iSeeNCode. The application reminded coders when to observe, when to code, simplified entering the codes, and stored the data in a spreadsheet. iSeeNCode included 111 individual dichotomous (present or absent) codes organized into seven dimensions, three of which are the subject of this analysis: 1) Broad instructional focus (reading and writing), 2) Specific writing instructional focus, and 3) Student writing practice.[1] Codes were selected for the protocol based on theory and the existing body of observational research on writing instruction (Coker et al., 2016).

Broad instructional focus. Observers identified when teachers were engaged in writing and/or reading instruction and selected specific codes to describe the instructional focus.

Specific writing instructional focus. In order to capture the potential variability in instruction, twelve different writing instructional codes were used. For the analysis, these codes were grouped in terms of two conceptual domains: skills-based instruction (including codes for instruction in spelling, grammar, handwriting, punctuation/capitalization, and keyboarding) and composition-based instruction (including codes for instruction in process writing, revising, editing, informative composing, narrative composing, sharing text by students, and sharing text by teachers). These groupings were made for both practical and theoretical reasons. The goal of

[1] The remaining coding categories included grouping, management of instruction, teacher instructional mode, and materials.

combining the individual codes was to represent accurately the types of instruction that occurred in first grade.

Student writing practice codes. Observers also coded the type and amount of student writing practice. Students' writing activities included correct/copied writing and generative writing. A correct/copied response occurred when students either were writing a response that had a single correct answer, such as a worksheet, or were copying text. Generative writing involved situations when students wrote but there was not a single expected answer. Examples include writing narratives and informative texts, open-topic journal entries, and any other open-ended writing task.

Reading instruction. Reading instruction was coded using nine specific codes designed to capture the full range of instructional targets and activities. The codes included assessment, lower-level comprehension, higher-level comprehension, strategies for comprehension, phonological and phonemic awareness, read aloud, vocabulary, word recognition, and other. The codes were selected based on existing curricular practices and then refined during the development of the coding system.

Observer training and reliability testing. Observations were conducted by four experienced classroom teachers. Training involved instruction in the use of iSeeNCode and practice coding videos of classroom instruction. Observers compared codes and discussed any coding questions with the research team. Any coding disagreements were resolved by the master coder (the first author). Once observers were able to code a 30-minute video segment with at least 80% agreement with the master coder, live coding was conducted in a classroom. A 90-minute literacy block was then coded with the master coder. Once kappa and percent agreement showed that all observers reached a minimum threshold of .80, coders could begin collecting

data. The average agreement across coding dimensions (e.g., broad instructional focus, specific writing focus, etc) ranged from .87 to .96 with an overall average of .92.

To guard against observer drift over time, observers' reliability was assessed over the course of the project in two ways. During each of the four waves of data collection, observers double coded 90 minutes of a school day with the master coder. A minimum agreement of .8 was required to proceed with data collection. Secondly, at the beginning of the second year, a reliability check-up was conducted that required observers to reach the agreement threshold of .80 with the master coder.

Observational data. The sample of 50 classrooms was observed for four full instructional days for a total of 200 observation days and 11034 5-minute observation blocks (*M* = 220.68 5-minute blocks per classroom). All blocks were coded using the procedures described above, resulting in a dataset of dichotomous variables characterizing the presence or absence of instruction and practice during observation blocks. In order to convert these data into measures of instruction in each classroom, we aggregated observational block data to the observation day, then across days to the classroom, using the following steps. First, we calculated the number of observational blocks each day that an instructional code was observed. This yielded a measure of the amount of time 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 provided a measure of the average time devoted to types of instruction and student practice for a classroom across the year. We used this as a measure of classroom instruction. For example, if observers coded 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 across observations would be 4.25 blocks, or 21.25 minutes of skills-focused writing instruction.

We elected to use the average number of blocks because our objective was to describe classroom instruction and student practice *over the school year*. This approach was designed to provide a more representative assessment of classroom experiences across the school year. The dataset ultimately included the average number of blocks dedicated to different writing and reading instructional actions and writing practice across the school year.

Student Measures

Decoding and word reading. To assess students' decoding skills, two subtests from the Woodcock Johnson III Tests of Achievement (WJIII; McGrew, Schrank, & Woodcock, 2007) were used the both the fall and the spring: Letter Word Identification (LWID) and Word Attack (WA). In the LWID subtest, students were shown a list of letters and words to be read aloud. Testing was discontinued when students missed six consecutive items. For the WA subtest, students were asked to read pesudowords aloud until six consecutive items were incorrect. LWID and WA were combined to form the Basic Reading Cluster.

Reading fluency. The DIBELS Next Oral Reading Fluency (ORF; Good et al., 2011) subtest was administered individually to students in the spring by the participating school districts. Students were asked to read as many words as possible in three, 60-second passages. The ORF score provided by the districts was the number of words read correctly for the median passage.

Reading comprehension. Students' reading comprehension was assessed with the WJIII Passage Comprehension subtest (PC; McGrew et al., 2007) in the spring. Students were asked to read a passage and then identify a single word to complete a sentence in a cloze task. The test was discontinued when students reached the ceiling of six, consecutive, incorrect items.

Handwriting fluency. A measure of letter-writing fluency was administered to students individually in the fall (Abbott & Berninger, 1993). Students were given a piece of lined paper and a pencil without an eraser. The examiner asked students to write 26 lower-case alphabetic letters as quickly as possible in one minute. Students were also told that every 15 seconds the examiner would mark the paper to indicate students' progress. One point was awarded for each correctly formed letter in the right order. Letters that were illegible, formed incorrectly or written in the wrong order were scored as incorrect. Trained research assistants (RAs) scored students' responses, and the inter-rater agreement was 100%. This assessment has been used by other researchers with both a 15-second (Berninger & Rutberg, 1992; Graham et al., 1997; Jones & Christensen, 1999) and a 60-second time limit (Kim, Al Otaiba, Puranik, Folson & Gruelich, 2014; Wagner et al., 2011). Some students completed the task and stopped writing in less than 60 seconds, but no student stopped writing in under 45 seconds. To facilitate comparisons among students, the total number of correctly written letters in the first 45 seconds was used for analysis.

Spelling. Spelling ability was measured using the WJIII Spelling subtest (McGrew et al., 2007) in the fall. Students were asked to write the letters and words dictated by the examiner. Testing continued until students missed six consecutive items.

Vocabulary. The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007) was administered in the fall. The examiner read a word, and students were asked to point to one of four pictures to represent the word. Testing was discontinued when students missed eight or more items in a set.

All assessments were administered by trained RAs, except for DIBELS Next ORF, which was administered by school personnel. Testing occurred during the school day in a quiet location

outside of the classroom. RAs followed the administration manual for all assessments. The standardized assessments included discontinuation rules or the use of basal and ceiling sets to minimize student fatigue (i.e., LWID, WA, PC, WJIII Spelling, and PPVT-4).

Data Analysis Strategy

In the current study, students (N = 391) were nested in classrooms (N = 50), nested in schools (N = 13). In multilevel analyses, sample size at the highest level of analysis is a primary restriction to accurate model estimation (Maas & Hox, 2005). Maximum likelihood estimation methods are asymptotic and require approximately 30 to 50 units at the highest level of analysis for accurate parameter and standard error estimation. Given the current study includes only 13 schools, we specified two-level fixed effects models in which students (Level-1) are nested within classrooms (Level-2) with school fixed effects to account for the non-independence of classrooms within schools. The average classroom-level cluster size was 7.82 students (range: 4-9 students).

We addressed our research questions through a four-step data analysis strategy. First, student-level 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). We imputed 18 values across 4 measures (range: 3-7 missing data points per measure). We did not impute more than one value per student. No classroom-level data was missing. All variables were centered to eliminate nonessential collinearity (Dalal & Zickar, 2012; Little, Card, Bovaird, Preacher, & Crandall, 2007); within-group-only variables were group-mean centered while remaining variables were grand-mean centered.

Second, we developed a latent measure of early reading to assess first-grade reading achievement comprised of four spring reading measures: LWID, WA, PC, and ORF. Use of a

latent measure reduces measurement error (Klein, 2005), allows the inclusion of relevant reading subskills, and aligns with previous research with young readers (Lerkkanen et al., 2004; Oslund et al., 2015). Intraclass correlation coefficients (ICCs) represent the proportion of total variance that exists between groups (Heck & Thomas, 2015); values between .05 and .15 traditionally indicate adequate variance for multilevel modeling in education contexts (Hedges & Hedberg, 2007). ICCs for the four indicators of latent spring reading achievement reveal between 12 and 18 percent variance at the classroom level (ρ_{LW} = .18, ρ_{WA} = .14, ρ_{PC} = .12, ρ_{ORF} = .17).

Third, we evaluated cross-level measurement invariance in latent spring reading achievement. This analysis provided insight into whether latent spring reading achievement is measured by the same metric at the student and classroom levels or if the construct varies by level. Multilevel confirmatory factor analyses (ML-CFAs) with free and constrained factor loadings at Level-2 are compared via the Santorra-Bentler scaled chi-square statistic to test for statistically significant differences in model fit.

Finally, we developed a multiple mediator model to explore the relationship between writing instruction, writing practice, and latent spring reading achievement. We estimated multilevel structural equation models (ML-SEMs) using the Mplus 7.4 robust maximum likelihood estimator (Muthén & Muthén, 1998-2015). ML-SEM provides a framework for simultaneously estimating measurement and structural models while reducing cluster bias in parameter estimates and correcting standard errors for the non-independence of structural data (Heck & Thomas, 2015; Hox, 2010; Mehta & Neale, 2005).

We used five indices to evaluate ML-SEM fit: chi-square, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and Tucker-Lewis index (TLI). RMSEA and SRMR are absolute fit indices; values

below .08 indicate reasonable fit and values below .06 indicate good fit (Hu & Bentler, 1999). CFI and TLI are relative fit indices that typically range between values of 0 and 1; values greater than .90 traditionally indicate good fit while more recent research suggests values greater than .95 are a preferable indicator of model fit (Bentler & Bonett, 1980; Hu & Bentler, 1999). RMSEA, CFI, and TLI are a function of the chi-square test statistic and, as a result, are more sensitive to within-group model misspecification (Hsu, Kwok, Lin, & Acosta, 2015; Ryu & West, 2009). However, ML-SEM covariance matrices for the within- and between-group models are computed separately which allows the SRMR to be computed separately for the within- (SRMR-W) and between-group (SRMR-B) models. To this end, SRMR-B is the most sensitive fit index to between-group model misspecification.

Results

Descriptive Statistics

Tables 1 and 2 present descriptive statistics and bivariate correlations for student and classroom measures, respectively. In the norm-referenced reading and spelling assessments, students performed close to the average range; however, averages for LWID and the Basic Reading Cluster were approximately two-thirds of a standard deviation above the national average, which could reflect more attention to decoding instruction in kindergarten. In the observational data, reading instruction occurred in 17.54 blocks per observation day, on average, which represents 87.70 minutes. There were more than three times as many blocks when reading instruction was observed as writing instruction, which occurred 5.21 blocks per day, on average, which represents 26.05 minutes. However, there was greater relative variation in the amount of writing instruction than reading. There were moderate to strong correlations among the reading and writing measures; however, gender and minority status were weakly correlated to all

measures (Table 1). At the classroom level, there were small-to-moderate correlations between skills writing instruction and both correct/copy and generative writing practice (respectively, r = .21; r = -.19). Alternately there was a moderate correlation between composing writing instruction and generative writing practice (r = .31).

Cross-Level Measurement Invariance

Multilevel latent variable analyses may allow unstandardized factor loadings to vary between levels or constrain loadings to be equal between levels (Marsh et al., 2012). The latter method, termed cross-level measurement invariance, facilitates construct interpretation by measuring multilevel latent variables by the same metric at all levels of analysis. Alternately, if cross-level measurement invariance does not hold, multilevel latent variables require a unique interpretation at each level due to inconsistent factor loadings.

We conducted two ML-CFAs to evaluate cross-level measurement invariance of latent spring reading achievement. Both models included an a priori correlated measurement residual between LWID and WA. Both ML-CFA models fit the data well and the Santorra-Bentler scaled chi-square difference test indicated no significant difference between the free and constrained factor loading models (TRd = .29, df = 3, p = .96). Accordingly, latent spring reading achievement met the requirements for constrained factor loadings and subsequent analyses were based upon the cross-level invariance model.

Analytically, cross-level measurement invariance allows us to interpret latent spring reading achievement as a common construct between the student- and classroom-levels. Computationally, cross-level measurement invariance constrains unstandardized factor loadings to be equivalent across levels. However, standardized factor loadings will be different between levels because standardization occurs at each level and is not based on an overall standard

deviation. To this end, standardized factor loadings presented in the multiple mediator model will not appear to be equivalent despite cross-level measurement invariance.

Multiple Mediator Model

We developed the multiple mediator model through a four-step model building process. First, we estimated a baseline model of latent spring reading achievement (see Model 1 in Table 3). Due to cross-level measurement invariance, latent spring reading achievement is a consistent construct at the student- and classroom-levels. As such, we were able to estimate a baseline model to calculate the ICC for latent spring reading achievement. Model 1 fit the data well at both the student- and classroom-levels (χ^2 = 4.73, RMSEA = .00, CFI = 1.00, TLI = 1.00, SRMR-W = .01, SRMR-B = .02) and the latent spring reading achievement ICC reveals 18 percent variance at the classroom level (ρ_{LSRA} = .18), which indicated adequate variance for multilevel analysis.

Second, we specified the student-level structural model (see Model 2 in Table 3). Fall vocabulary, spelling and basic reading achievement measures were included because of their relationship with overall reading success (NICHD, 2000; Snow et al., 1998), and handwriting fluency was included because of its relationship to letter knowledge, an important component of decoding (Bara, Morin, Alamargot, & Bosse, 2016). Gender and minority status were also included as control variables because both have been associated with reading achievement (NCES, 2013).

Model 2 fit the data well at both the student- and classroom-levels (χ^2 = 81.49, RMSEA = .08, CFI = .97, TLI = .95, SRMR-W = .03, SRMR-B = .02). Fall vocabulary, handwriting fluency, spelling, and basic reading were significantly associated with spring reading achievement (respectively, β = .08, p = .011; β = .09, p = .001; β = .15, p < .001; β = .72, p <

.001). Alternately, the effects of gender and minority status were not statistically significant (respectively, $\beta = .00$, $p = .983$; $\beta = -.08$, $p = .196$). Furthermore, Model 2 explained 86% of the student-level variation in spring reading achievement.

Third, we specified the instructional portion of the classroom-level structural model with school fixed effects (see Model 3 in Table 3). In addition to the variables included in Model 2, observed measures of skills and composing writing instruction, total reading instruction, and school fixed effects were included at the classroom level. Model 3 fit the data well at both the student- and classroom-levels ($\chi^2 = 157.14$, RMSEA = .06, CFI = .96, TLI = .95, SRMR-W = .03, SRMR-B = .03). Furthermore, Model 3 explained 49% of the classroom-level variation in spring reading achievement.

Neither skills nor composing writing instruction were statistically significantly associated with spring reading achievement (respectively, $\beta = -.06$, $p = .661$; $\beta = -.07$, $p = .630$). Additionally, the effect of total reading instruction was not statistically significant ($\beta = -.01$, $p = .966$). While not directly interpreted, school fixed effects were included to account for the non-independence of classrooms in schools. While the causal steps approach to mediation analysis requires a significant relationship between the independent and dependent variable prior to mediation analysis (Baron & Kenny, 1986; Judd & Kenny, 1981), this relationship may not be statistically significant in inconsistent mediation models (MacKinnon, Fairchild, & Fritz, 2007). When the direction of mediated relationships is inconsistent, the overall relationship between an independent and dependent relationship may be zero due to opposing mediational processes. To explore inconsistent mediational model, we specified a multiple mediator model to test the effect of writing instruction on reading achievement as mediated by writing practice.

Finally, we specified the mediated classroom-level structural model with school fixed effects (see Model 4 in Table 3; Figure 2). In addition to the variables included in Model 3, an observed measure of correct/copy writing practice was included as a mediator of both skills and composing writing instruction, and an observed measure of generative writing practice was included as a mediator of composing writing instruction. Model 4 fit the data well at both the student- and classroom-levels (χ^2 = 287.80, RMSEA = .07, CFI = .92, TLI = .90, SRMR-W = .03, SRMR-B = .07). Furthermore, Model 4 explained 59% of the classroom-level variation in spring reading achievement.

The direct and indirect paths from skills writing instruction through correct/copy writing practice did not explain statistically significant variation in spring reading achievement (respectively, β = -.11, p = .437; β = -.05, p = .310). Similarly, the direct and indirect paths from composing writing instruction through correct/copy writing practice did not explain statistically significant variation in spring reading achievement (β = -.16, p = .205; β = .02, p = .441). Alternately, the indirect path from composing writing instruction to spring reading achievement through generative writing practice was positive and statistically significant (β = .15, p = .024). Furthermore, the total indirect effect of composing writing instruction on spring reading achievement is positive and statistically significant (β = .17, p = .029). These findings are consistent with inconsistent mediation as the direct effect of composing writing instruction is in the opposite direction of the indirect effects.

Discussion

Previous research on the effects of writing interventions on reading has demonstrated that interventions targeting spelling, sentence writing, and composition involving student practice have affected students' reading outcomes (Graham & Hebert, 2011). However, there is little

empirical work on whether the writing instruction and practice that teachers typically provide in first grade are related to student reading achievement. In this study, we modeled the effects of typical writing instruction and student writing practice on reading achievement in first grade using ML-SEM.

The results for the first research question indicated that there were no direct effects of either skills or composing instruction on reading achievement. However, results from the second research question revealed that composing instruction had a positive and significant relationship with reading achievement when mediated by generative writing practice. These findings provide limited support for our conceptual framework and suggest that in typical first-grade classrooms, the impact of writing instruction on reading achievement depends on the type of writing instruction and the type of writing practice. These results are interpreted in the context of the two main hypotheses.

Direct Effects of Typical Writing Instruction

Our first hypothesis was that the effects of skills and composing writing instruction would have direct effects on reading achievement after controlling for students' fall achievement and reading instruction. Our hypothesis was not confirmed, and the results did not support our assumptions in the conceptual framework that writing instruction would have a direct effect on reading achievement.

One explanation for these results is likely related to the amount of typical first-grade writing instruction observed in our sample. Overall, relatively little writing instruction was provided. On average, classrooms had only 9.2 minutes (1.84 blocks) of skills instruction and 15.55 minutes (3.11 blocks) of composing instruction, and there was substantial variation around the average. These data signal that sustained writing instruction focusing on strengthening skills

and composing was not a consistent practice in most classrooms and did not meet current recommendations for the amount of daily writing instruction (Graham et al., 2012). This lack of consistency may explain why there were no direct effects for writing instruction found.

Another potential explanation for the null results was the limited curricular support for writing that teachers received. Nearly half of the teachers (46%) reported that they had no writing curriculum for their class. Many others indicated that writing was integrated into the reading curriculum (44%), and a small percentage (10%) used a supplementary curriculum designed to teach nonfiction writing (Coker et al., 2016). With little curricular guidance, teachers may have been uncertain about how to teach writing or about how to integrate it into their existing lessons.

A final reason why no direct effects of writing instruction were found may be related to student practice. In instructional research, student practice is a central component of academic interventions. The writing intervention studies relevant for this study included frequent opportunities for student practice (e.g., Berninger et al., 1998; Conrad, 2008; Craig, 2006; Sénéchal et al., 2012). Alternating writing instruction and opportunities for writing practice may have contributed to the success of those interventions in strengthening reading achievement. The importance of student writing practice was also recognized in an Institute of Education Sciences' practice guide, which listed providing time for students to practice the skills and strategies they are taught as one of its four recommendations (Graham et al., 2012).

Direct and Indirect Effects of Typical Writing Instruction and Student Writing Practice

In the second hypothesis, we acknowledged the importance of student writing practice by stating that the effect of writing instruction would be mediated by writing practice. Additionally, we hypothesized that the indirect effect of writing instruction through writing practice might

depend on both the type of instruction and the type of practice. We simultaneously fit indirect paths from composing instruction through both generative writing practice and correct/copy practice and a path from skills instruction through correct/copy writing practice (Model 4). Only the path from composing instruction through generative writing practice was positive and statistically significant.

These results provide partial support for our conceptual framework by signaling that the relationship between one type of writing instruction (composing) and reading achievement was mediated by a specific form of writing practice (generative). These results suggest that there are beneficial roles for specific types of writing practice and that at least generative writing practice may have special benefits when coupled with composing instruction.

Our findings add to the body of empirical research that has revealed relationships between writing and reading in first grade (Abbott, Berninger, & Fayol, 2010; Ahmed, Wagner, & Lopez, 2014; Lerkkanen et al., 2004; Shanahan, 1984). Furthermore, evidence of the relationships between writing instruction, practice, and reading achievement may also support theoretical approaches that state that reading and writing draw on shared knowledge (Fitzgerald & Shanahan, 2000).

These findings are also consistent with empirical research indicating that writing instruction and practice may have numerous benefits for reading achievement (Graham & Hebert, 2011). In a meta-analysis Graham and Hebert (2011) found that instruction in spelling and sentence writing improved the word reading skills of students across grades 1-7. These instructional studies did not isolate instruction from opportunities for student writing practice, so the effects also included the contribution of student practice. However, Graham and Hebert also estimated the unique impact of writing practice by analyzing studies that investigated whether

increasing student writing would impact reading comprehension. Even in the absence of writing instruction, writing practice strengthened reading comprehension in grades 1-6 with an average effect size of .35.

An important point is that the effect of composition instruction mediated by generative writing was found even after including control variables for students' fall achievement and for the amount of reading and writing instruction. In Models 3 and 4, students' fall reading, handwriting fluency, and spelling achievement, as well as the amount of reading instruction were controlled. It is notable that even with these controls at both the within- (student) and between-group (classroom) levels of the model, a significant indirect path was found for composing instruction through generative writing practice. One conclusion is that composing instruction combined with opportunities to generate text may be a promising combination of classroom practices for first graders.

Our findings are also noteworthy considering that teachers were not given any guidance on how to teach writing in this study. Despite relatively few curricular resources, the results suggest that some effective practices are currently in use. Conversely, sustained efforts to build teachers' use of evidence-based writing practices may be expected to demonstrate stronger relationships with student achievement.

In addition, these findings point to the complexity of the relationships between writing instruction and practice. It may be that combining specific types of writing instruction and practice yield differential student outcomes. While little existing research has explored how to pair different instructional foci with practice activities, researchers have found that providing opportunities to integrate handwriting (Berninger et al., 1997; Graham, Harris, & Fink, 2000)

and spelling instruction (Berninger et al., 1998; Berninger et al., 2000) with word and text writing is effective.

One somewhat surprising result is that the only significant and positive effect of writing instruction and practice was found when composing instruction was mediated by generative writing practice. The simple view of writing (Juel et al., 1986) and empirical investigations of early writing achievement have revealed strong relationships between transcription skills (i.e., handwriting and spelling) and both writing (Abbott et al., 2010; Jones & Christensen, 1999; Kent & Wanzek, 2016;) and reading achievement (Abbott et al., 2010; Graham & Hebert, 2011). Based on these earlier results, the findings of this study may appear unexpected. There may be several explanations for our findings.

First, it is possible that composing instruction coupled with generative writing practice has greater explanatory power because the complex demands these two activities make on young writers. When students receive instruction in composing, they may learn how to plan, draft, evaluate the text that has been written, revise, and edit their work. Then, when given opportunities to write open-ended texts, students must draw on linguistic knowledge to generate ideas, process knowledge to plan and revise text, transcription skills to encode individual words, syntactic knowledge to form sentences, and discourse knowledge to create larger textual units. The complex demands of learning to compose and creating new texts may strengthen writing skills, text generation capacity, writing knowledge, and perhaps even executive functions. If composing practices strengthen multiple sources of writing knowledge, they may contribute to reading achievement above the level of transcription skills.

In addition to its complexity, generative writing practice may also facilitate transfer to reading in ways that skills practice does not. When students compose texts, they apply their

phonological and orthographic knowledge to encode new words, often called invented spelling. Using invented spelling when composing differs from spelling practice, which typically has students writing specific words from a lesson rather than leveraging their knowledge to spell. Furthermore, as they use invented spelling while composing, students may build and refine their understanding of how sounds and letters function in words (Ouellette & Sénéchal, 2008), which may transfer to reading. In fact, various approaches to supporting students' invented spelling have been shown to improve word reading (Ouellette & Sénéchal, 2008; Ouellette et al., 2013; Sénéchal et al., 2012). Another way that generative writing may facilitate transfer to reading skills is through sentence writing. As students compose sentences that convey their ideas, they may deepen their understanding of how written language is used to represent thought. This experience may transfer to comprehending the meaning of texts.

Secondly, the nature of the measures used in this study may also help explain the findings. Since little is known about current practices in early writing instruction, our goal was to investigate the amount and type of writing instruction occurring in typical, first-grade classrooms. Our observers used a highly structured observation system to collect these data; however, measures of instructional quality were not included. It is possible that the quality of the spelling and handwriting instruction was not high enough to explain unique variance in reading achievement. There is little research on the quality of early writing instruction, but in at least one study, researchers found that teachers' responsiveness was predictive of first-graders' writing quality (Kim et al., 2013). Whether the quality of writing instruction would be predictive of reading achievement is a question that, to our knowledge, has not been addressed. Certainly, the impact of writing instructional quality on reading and writing achievement deserves attention.

Typical Reading Instruction as a Control Variable

In the models, the amount of reading instruction was included as a control variable at the classroom level. Since this analysis was focused on the effects of writing instruction and practice, we did not make a hypothesis about whether reading instruction would predict broad reading achievement; however, it seemed likely that it would. Since our primary interest was to control for the effect of reading instruction, the instructional variable included was the total amount of reading instruction. It may be that specific types of instruction and practice would be more strongly associated with achievement. In addition, the efficacy of specific forms of instruction and practice may also depend on student skill (e.g., Connor et al., 2011; Connor et al., 2004; Connor, Morrison, & Underwood, 2007). While those questions were not a part of this analysis, they deserve future research.

Student-Level Control Variables

When controlling for students' fall achievement, we included measures of reading, spelling, handwriting and vocabulary in the student-level model. Not surprisingly fall reading was strongly related to spring reading achievement ($\beta = .72$). In addition, spelling ($\beta = .15$), vocabulary ($\beta = .08$) and handwriting fluency ($\beta = .09$) were also significantly associated with reading achievement.

Limitations

The findings of this study are certainly influenced by the measures that were collected as well as the nature of observational research. The classroom measures relied on four, time-sampled observations of instruction and student practice and are therefore subject to various sources of error, as described by Raudenbush and Sadoff (2008). Our initial design attended to minimizing error in ways that were feasible given human and financial resources for conducting observational work. These included minimizing the number of observers, conducting whole-day

observations, and increasing the number of days observed to four from the three typically found in literacy research. However, bias may still be introduced through factors such as non-random assignment of observers to classrooms and the timing of classroom observations during the school year. Additionally, we have utilized average blocks as a measure of instruction and student practice across the school year. In using the average, however, we recognize that we lose information that would be helpful in further differentiating classrooms, and we also lose information about between-observation variability, which we have discussed elsewhere (Coker et al., 2016). These limitations should be acknowledged, and future research may improve upon our methods in several ways. Including more observations across the school year could increase the reliability of classroom measures. More fine-grained analyses could be done with classroom videos, which would enable coders to capture instruction in very small time segments and potentially to characterize the nature and amount of each student's writing practice with more precision. Furthermore, random assignment of observation dates and observers may be useful in reducing bias, if possible given the constraints of resources and context.

Another limitation of our measures was that observers coded the nature and total amount of writing instruction rather than rating the instructional quality. Perhaps with measures that are sensitive to the multiple dimensions of instructional quality, a more nuanced pattern of results would have emerged.

With only 50 unbalanced classrooms included in the sample, the models were underpowered, especially to detect a mediated effect. Post hoc Monte Carlo simulation (n = 10000 replications) indicated the study was underpowered to detect statistically significant indirect effects in Model 4 (*power* < .80; Cohen, 1990). With more classrooms and a balanced

sample of students in each classroom, other relationships might have reached statistical significance in our models.

All of the teacher participants volunteered to participate in the study. Since they were the teachers who expressed interest in the project, they might have been stronger writing instructors, or they might have had higher levels of writing instructional efficacy. It is important to note that even with a sample of volunteers there was no writing instruction observed in nearly a quarter of our observation days (Coker et al., 2016).

Although we included several theoretically relevant controls for students' achievement in the fall, additional measures would likely strengthen the model. Additional measures related to reading and writing achievement, such as phonological awareness, listening comprehension, working memory, and self-regulation, may be used to model both the direct and indirect effects on reading (Kim & Schatschneider, 2016; Vellutino, Tunmer, Jaccard, & Chen, 2007).

Future Directions

Potentially productive directions for future research include investigating the relationships between writing instruction, writing practice, and reading achievement with students at other grade levels. As students become more advanced readers and writers, it seems likely that the effects of writing instruction on reading achievement may change, particularly as students achieve fluency with word recognition and transcription.

Our analyses of student writing practice included two theoretically motivated types of practice—correct/copy and generative writing. In our first-grade sample, these were also the most prevalent forms of writing practice. However, future work with older students should explore additional forms of writing practice, such as writing about text. It is likely that this type

of writing practice will be more common in higher grades as students move from an emphasis on learning to write to writing to learn.

Another important extension from these results involves investigating the nature of composing instruction and generative writing practice. A more detailed understanding of this combination of writing instruction and practice may provide insight into how it explains reading achievement. These studies may involve intervention research that isolates and tests individual components of both composing instruction and generative writing practice to determine if specific parts or combinations are more effective than others. Finally, research on additional combinations of literacy instruction and student practice that contribute to students' reading and writing development is warranted.

Conclusions

The findings of the current study provide additional evidence of the importance of writing instruction for reading achievement (Graham & Hebert, 2011). Specifically, our models indicated that only composing instruction was related to reading achievement, and its effect was mediated by generative writing practice. The findings provide partial support for our conceptual framework and highlight the important role that specific forms of student writing practice may play in writing instruction.

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.

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

Adams, M.A. (1994). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adams-Boating, A. (2001). Second graders' use of journal writing and the effects on reading comprehension. Unpublished master's thesis, Keen University, Sydney, Australia.

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.

Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of Educational Psychology*, *106*(2), 419–434. http://doi.org/10.1037/a0035692

Bara, F., Morin, M.-F., Alamargot, D., & Bosse, M.-L. (2016). Learning different allographs through handwriting: The impact on letter knowledge and reading acquisition. *Learning and Individual Differences*, *45*, 88–94. https://doi.org/10.1016/j.lindif.2015.11.020

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

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606.

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders 20*(4), 65-84.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review, 8*(3), 193–218.

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., & 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. W., Vaughan, K. B., Abbott, R. D., Abbott, S. P., Rogan, L. W., Brooks, A., … Graham, S. (1997). Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition. *Journal of Educational Psychology, 89*(4), 652–666.

Berninger, V. W., Vaughan, K., Abbott, R. D., Brooks, A., Abbott, S. P., Rogan, L., … Graham, S. (1998). Early intervention for spelling problems: Teaching functional spelling units of

varying size with a multiple-connections framework. *Journal of Educational Psychology*, *90*(4), 587–605.

Berninger, V. W., Vaughan, K., Abbott, R. D., Brooks, A., Begay, K., Curtin, G., … Graham, S. (2000). Language-based spelling instruction: Teaching children to make multiple connections between spoken and written words. *Learning Disability Quarterly*, *23*(2), 117–135. http://doi.org/10.2307/1511141

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304-1312.

Coker, D. L., Ritchey, K. D., Uribe-Zarain, X., & Jennings, A. S. (2017). An analysis of first-grade writing profiles and their relationship to compositional quality. *Journal of Learning Disabilities,* Advance online publication. https://doi.org/10.1177/0022219417708171

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

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., … Schatschneider, C. (2011). Testing the impact of child characteristics * instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*(3), 189–221

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. http://doi.org/10.1037/0022-0663.96.4.682

Connor, C. M., Morrison, F. J., & Underwood, P. S. (2007). A second chance in second grade: The independent and cumulative impact of first-and second-grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading*, *11*(3), 199–233.

Conrad, N. J. (2008). From reading to spelling and spelling to reading: Transfer goes both ways. *Journal of Educational Psychology*, *100*(4), 869-878.

Craig, S. A. (2006). The effects of an adapted interactive writing intervention on kindergarten children's phonological awareness, spelling, and early reading development: A contextualized approach to instruction. *Journal of Educational Psychology*, *98*(4), 714.

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

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*(3), 339-362.

Denner, P. R., McGinley, W. J., & Brown, E. (1989). Effects of story impressions as a prereading/writing activity on story comprehension. *The Journal of Educational Research*, *82*(6), 320–326.

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

Ehri, L. C., & Wilce, L. S. (1987). Does learning to spell help beginners learn to read words? *Reading Research Quarterly*, *22*(1), 47. http://doi.org/10.2307/747720

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, *35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., … Wissel, S. (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. R., Herrera, S., Patscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. *Reading and Writing, 28*, 655-681.

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.

Frey, J. L. (1993). *Process Writing Approach and Its Effects on the Reading Comprehension of First-Grade Students in the Mississippi Delta*. Unpublished Dissertation, Delta State University.

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities*, *39*(5), 467–475. https://doi.org/10.1177/00222194060390050701

Good, R. H., & Kaminski, R. A. (Eds.). (2011). *DIBELS Next assessment manual*. Eugene, OR: Dynamic Measurement Group. Retrieved from http://dibels.org/next.html

Good, R.H. III, Kaminski, R.A., Dewey, E.N., Wallin, J., Powell-Smith, K.A., & Latimer, R.J. (2011). *DIBELS Next Technical Manual*. Eugene, OR: Dynamic Measurement Group, Inc. Retrieved from http://www.dibels.org.

Gough, P. and Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.

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., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, *94*(4), 669–686.

Graham, S., Harris, K. R., & Fink, B. (2000). Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers. *Journal of Educational Psychology*, *92*(4), 620.

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., & 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

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.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hsu, H.-Y., Kwok, O.-M., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research, 50*(2), 197-215.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

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.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*(5), 602-619. http://doi.org/10.1177/0193841X8100500502

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

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447.

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., Al Otaiba, S. A., 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. G., & Schatschneider, C. (2016). 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

Klein, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

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

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Mahwah, NJ: Lawrence Erlbaum Associates.

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

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*(1), 593-614. http://doi.org/10.1146/annurev.psych.58.110405.085542

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research.* (pp. 115–130). New York: Guilford Press.

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. http://doi.org/10.1027/1614-1881.1.3.86

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106-124.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259-284.

Muter V., Hulme C., Snowling M. J., Stevenson J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*, 665–681. http://doi.org/10.1037/0012-1649.40.5.665

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics (2013). *The nation's report card: Trends in academic progress 2012* (NCES 2013–456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Education Statistics (n.d.). *NAEP - 2015 Mathematics & Reading Assessments*. Retrieved from https://www.nationsreportcard.gov/reading_math_2015/#reading/acl?grade=4

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.* Washington, DC: Author. Retrieved from http://www.corestandards.org/the-standards

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: Report of the Subgroups* (NIH Publication 00-4754). Washington, DC: U.S. Department of Health and Human Services.

Oslund, E. L., Simmons, D. C., Hagan-Burke, S., Kwok, O.-M., Simmons, L. E., Taylor, A. B., & Coyne, M. D. (2015). Can curriculum-embedded measures predict the later reading achievement of kindergarteners at risk of reading disability? *Learning Disability Quarterly*, *38*(1), 3–14. http://doi.org/10.1177/0731948714524752

Ouellette, G., & Sénéchal, M. (2008). Pathways to literacy: A study of invented spelling and its role in learning to read. *Child Development*, *79*(4), 899–913. https://doi.org/10.1111/j.1467-8624.2008.01166.x

Ouellette, G., Sénéchal, M., & Haley, A. (2013). Guiding children's invented spellings: A gateway into literacy learning. *Journal of Experimental Education*, *81*(2), 261–279. https://doi.org/10.1080/00220973.2012.699903

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., Sadoff, S., (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness 1*(2), 138–154.

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

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16*, 583–601. http://doi.org/10.1080/10705510903203466

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48,* 115-136.

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

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Sénéchal, M., Ouellette, G., Pagan, S., & Lever, R. (2012). The role of invented spelling on learning to read in low-phoneme awareness kindergartners: A randomized-control-trial study. *Reading and Writing*, *25*(4), 917–934. http://doi.org/10.1007/s11145-011-9310-2

Shanahan, T. (1984). Nature of the reading-writing relation: An exploratory multivariate analysis. *Journal of Educational Psychology*, *76*(3), 466–477.

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

Snow, C., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

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

Storch, S. A. & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

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.

Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first-grade reading program: effects on several measures of reading. *Reading Research Quarterly*, *28*(3), 219–233. http://doi.org/10.2307/747995

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, *11*(1), 3–32.

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

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	10
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 Letter-Word ID*	.92	112.45	12.81	-.05	-.07	.39	.47	.77	.86	1			
8. WJ Word Attack*	.91	108.68	9.39	-.09	-.09	.29	.39	.70	.76	.80	1		
9. WJ Passage Comp*	.83	105.35	11.37	-.04	-.17	.51	.40	.63	.72	.76	.64	1	
10. DIBELS ORF	.95‡	73.37	34.35	.03	-.13	.37	.49	.69	.78	.81	.65	.71	1

Notes: $N = 391$; 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. *Mean and standard deviation of standard scores reported for interpretability; W scores used for WJ subtests and raw scores used for Oral Language (PPVT-4) to compute correlations and in subsequent analyses. †Interrater reliability. ††Composite measure; reported reliability is the lowest reliability of all component measures. ‡Lowest alternate-form and test-retest reliability reported by publisher as schools did not report item-level data.

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

Variable	M	SD	1	2	3	4	5	6	7	8	9
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. Generative Writing Practice	6.78	4.04	-.19	.31	.15	-.17	1				
6. WJ Letter-Word ID*	*	*	-.23	.13	.14	-.38	.47	1			
7. WJ Word Attack*	*	*	-.21	.14	.21	-.34	.36	.96	1		
8. WJ Passage Comp*	*	*	-.23	.08	.20	-.27	.39	.95	.91	1	
9. DIBELS ORF*	*	*	-.23	.15	.21	-.24	.44	.94	.91	.93	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: *Mediated Writing Instruction Models*

Variable	Model 1	Model 2	Model 3	Model 4
Within (Student)				
Female		.00(.04)	.00(.04)	.00(.04)
Minority Status		-.08(.07)	-.08(.07)	-.08(.07)
Vocabulary		.08(.03)*	.08(.03)*	.08(.03)*
Handwriting Fluency		.09(.03)***	.09(.03)***	.09(.03)***
Spelling		.15(.04)***	.15(.04)***	.15(.04)***
Basic Reading		.72(.04)***	.72(.04)***	.72(.04)***
Between (Classroom)				
Skills-Based Instruction				
Direct Path			-.06(.14)	-.11(.14)
Indirect Path (Correct/Copy)				-.05(.05)
Composition-Based Instruction				
Direct Path			-.07(.14)	-.16(.13)
Indirect Path (Correct/Copy)				.02(.03)
Indirect Path (Generative				.15(.07)*
Writing)				
Total Reading Instruction			-.01(.16)	.24(.21)
School Fixed Effects	No	No	Yes	Yes
R^2-Within	–	.86***	.86***	.86***
R^2-Between	–	–	.49***	.59***

Notes: standardized path coefficients, standard errors in parentheses. *$p < .05$, **$p < .01$, ***$p < .001$

Figure 1:
Theoretical Model of Relationships between Classroom Writing Instruction, Student Writing Practice and Spring Reading Achievement

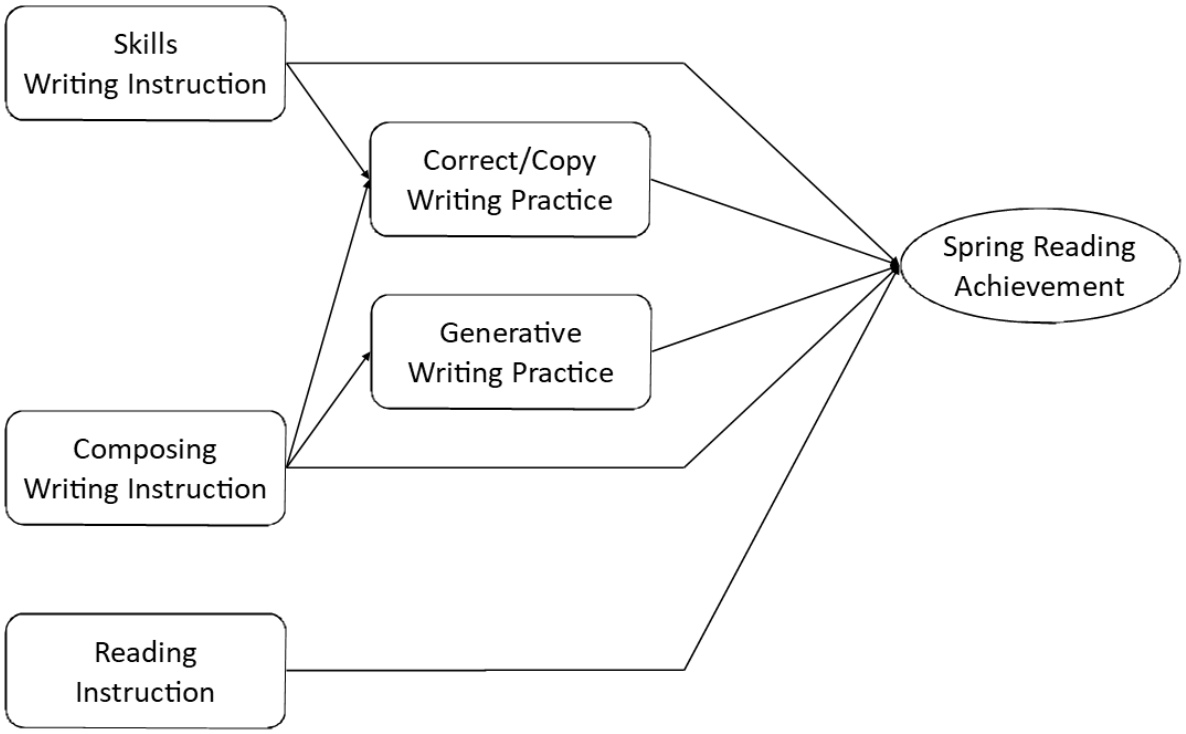
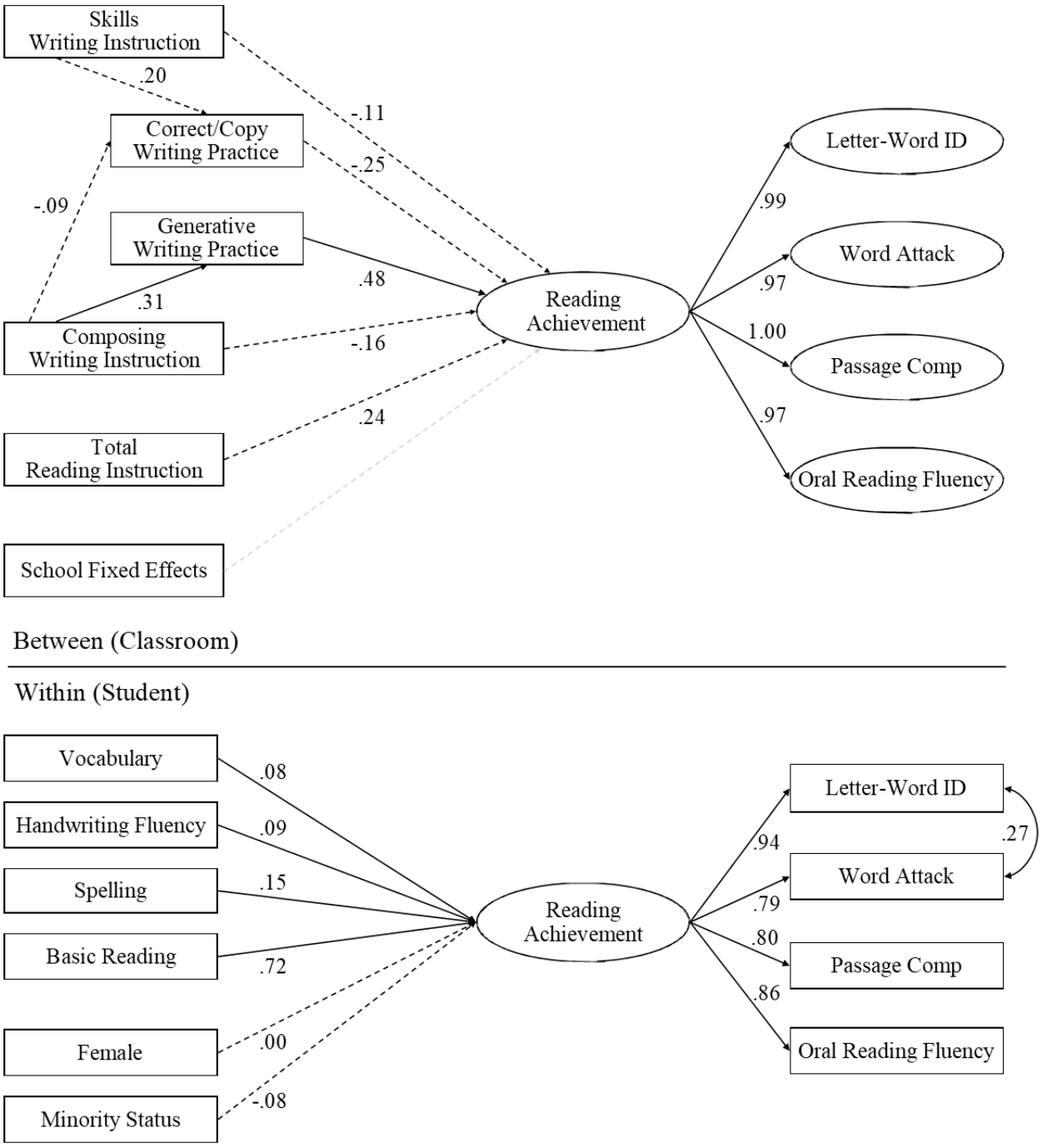


Figure 2: *Final Multiple Mediator Model (Model 4)*



Notes: numerical values indicate standardized path coefficients; solid lines indicate statistically significant paths while black dashed lines indicate non-statistically significant paths and the grey dashed line indicates a non-interpreted, fixed effects path. Standardized factor loadings are different at the within- (student) and between-group (classroom) levels despite cross-level measurement invariance due to standardization.