Writing Instruction Improves Students' Writing Skills Differentially Depending on Focal Instruction and Children: A Meta-Analysis for Primary Grade Students

Young-Suk Grace Kim,
University of California, Irvine

Dandan Yang,
University of California, Irvine

Marcela Reyes,
Irvine Valley College

Carol Connor
University of California, Irvine

Abstract

We investigated the effect of writing interventions on written composition for students in primary grades (K-G3) with a focus on whether effects vary as a function of different dimensions of composition outcomes (i.e., quality, productivity, fluency, and other), instructional focus (e.g., transcription, self-regulation strategies such as Self-Regulated Strategy Development [SRSD]), and student characteristics (i.e., initially weak writing skills). A total of 24 studies (number of effect sizes, $k = 166$; $N = 5589$ participants) met inclusion criteria. The overall mean effect size was moderate and positive ($ES = .31$) with some variation across the dimensions of composition: .32 in writing quality, .31 in writing productivity, .15 in writing fluency, and .34 in writing: other. SRSD had large and consistent effect sizes across the outcomes (.59 to 1.04) whereas transcription instruction did not yield statistically significant effects on any dimensions of composition due to large variation of effects across studies. Variation in instructional dosage (total length of instruction) did not explain variation in the effect sizes. Lastly, the average effect on writing quality was larger for writers with weaker writing skills compared to those with typical skills.

Keywords

Writing; meta-analysis; moderation; instruction; primary grade

1. Introduction

Writing is foundational in daily lives and academic achievement. Perhaps not surprisingly, writing (i.e., written composition) is an integral part of instruction as specified in the

Corresponding Author, Young-Suk Grace Kim, Professor, University of California, Irvine, youngsk7@uci.edu, Phone: (949) 824-2348.

Common Core State Standards (CCSS; National Governors Association & Council of Chief School Officers, 2010), which is widely adopted in the US. In primary grades, for example, students are expected to write in narrative, informational, and opinion genres for a range of discipline-specific tasks, purposes, and audiences. However, writing is one of the most challenging skills to develop because it involves a coordination and juggling of multiple processes, and draws on a number of language and cognitive skills (Hayes & Flower, 1981; Kim & Graham, 2021; Kim & Park, 2019; McCutchen, 2006). This is evident in the National Assessment of Educational Progress [NAEP] in the US, which has consistently shown that only about a quarter of school-aged students write at or above proficiency. In other words, the vast majority of students, three fourths, do not write with proficiency (National Center for Education Statistics [NCES], 2003, 2012). Although we acknowledge that the proportion of students meeting the "proficiency" standard depends on how stringent the criterion is, NAEP has consistently shown that the proportion of students who do not reach the proficiency level is greater in writing than in reading, which is approximately one third of school-aged students (see NCES, 2020) in the US. Then, our understanding of effective teaching of writing (writing instruction hereafter) is critical to address the needs of writing development. To this end, we conducted a meta-analysis of writing instruction studies that employed randomized controlled trial or quasi-experimental research designs for primary grade students (kindergarten to Grade 3).

Several meta-analyses have been conducted on the effect of writing instruction interventions in the last four decades. Some of these focused on a specific writing instructional approach such as the process approach to teach writing (e.g., Graham & Sandmel, 2011), self-regulation strategies (e.g., Graham, 2006), and the effect of reading instruction on students' writing quality (Graham et al., 2018; Stotsky, 1983). Other meta-analyses reviewed broader instructional approaches. For example, Graham and Perin (2007) reviewed studies for students in Grades 4 to 12 and found that several instructional practices such as sentence-combining (but not grammar instruction), modeling of good writing, summarization, and peer assistance were effective in improving writing quality. More directly relevant to the present meta-analysis is Graham, McKeown, Kiuhara, and Harris' (2012) review of studies of writing instruction, which focused on elementary school grade students (Grades 1 to 5). In this study, Graham et al. found that instruction on self-regulation strategies ($ES = 1.02$), transcription skills ($ES = .55$), and text structure ($ES = .59$) improved elementary-grade students' writing quality. The present study overlaps with that of Graham et al. (2012) in terms of grade span; and there was an overlap between the studies included in Graham et al. (2012) and in this study (58% or 14 out of 24 studies).

The present meta-analysis expands and differs from prior meta-analyses in two important aspects: by focusing on primary grade students *and* moderation of effects. Primary grades are a critical period when children learn and develop key foundational skills necessary for writing such as transcription skills (i.e., handwriting and spelling), language, cognition, and reading skills. Therefore, identifying effective writing instruction that builds strong foundational skills is critical to develop an important insight about effective practices in writing instruction during this crucial period. Longitudinal studies have shown that early literacy skills are related to later writing skills (Ahmed, Wagner, & Lopez, 2014; Juel, Griffith, & Gough, 1986; Kim, Al Otaiba, & Wanzek, 2015).

Beyond the overall average effect of writing instruction, we also systematically examined whether the effects differ by several important characteristics such as dimensions of written composition, the nature of writing instruction, student characteristics (typical vs. struggling writers), and dosage of writing instruction to understand for whom, what type of writing instruction, and in what outcomes writing instruction is effective. First, we systematically examined multiple dimensions of written composition. We examined the effects of writing instruction on writing quality, writing productivity (amount of writing), writing fluency (accuracy within a specified time), and other aspects of writing (e.g., writing conventions or number of transition words). Although writing quality is typically considered as the ultimate writing outcome, other dimensions of writing have been also widely examined and used in studies and education settings, particularly for novice and beginning writers in primary grades. Writing productivity (number of words or sentences), for example, is widely used as this can be easily measured and it is moderately to strongly related to writing quality for children in elementary grades (Abbott & Berninger, 1993; Kim, Al Otaiba, Sidler, Greulich, & Puranik, 2014; Kim, Al Otaiba, Wanzek, & Gatlin, 2015; Mackie & Docrkell, 2004; Olinghouse & Graham, 2009; Wagner et al., 2011). Similarly, writing fluency (e.g., the number of words or sentences constructed with a specified time) has been shown to have a strong relation to other outcomes such as grammatical complexity, linguistic experience, and writing quality (e.g., Chenoweth & Hayes, 2001; Johnson, Mercado, & Acevedo, 2012; Kim et al., 2015). Therefore, capturing effects of writing instruction on multiple outcomes beyond writing quality would reveal a more nuanced, precise picture of instruction effects.

Second, we examined differential effects as a function of the nature of writing instruction (e.g., transcription, self-regulation strategies). Studies vary in their focal target skills, and effects on different outcomes are likely to vary. Previous studies indeed suggest differential effects. Strategy instruction which typically includes instruction on multiple aspects (e.g., text structure, self-regulation strategies), for instance, appears to have a large effect compared to instruction that focuses on a single skill such as transcription, text structure, or sentence combining (Graham et al., 2012; Graham & Perin, 2007). However, this was on writing quality outcomes, and thus, it is not clear whether the nature of writing instruction has a differential impact on various dimensions of composition. In addition, we examined another feature of writing instruction—total length of writing instruction (i.e., dosage). Writing instruction varies in terms of total length of instruction, and therefore, it is an open question whether effects differ as a function of dosage of instruction.

Finally, we evaluated whether writing instruction effects differ by child characteristics (i.e., initial writing skill). Children bring their own skills to learning, and studies in reading have consistently shown that the effect of instruction varies depending on the child's needs (Connor et al., 2009, 2013). In this meta-analysis, we focused on the student's writing skill at pretest and examined whether instruction effects differ for students with weak versus typical writing skills. Considering such child-by-instruction interactions puts instruction at the center of literacy acquisition (Connor, 2016).

The theoretical foundation that guided the present inquiry on moderation was the direct and indirect effects model of writing (DIEW; Kim, 2020; Kim & Graham, 2021; Kim & Park, 2019). According to DIEW, overall writing development (typically examined

as writing quality) draws on a comprehensive set of skills and knowledge such as domain-general cognitions (e.g., working memory), language skills (e.g., vocabulary), transcription skills (spelling and handwriting/keyboarding), higher order cognitions and regulations (e.g., inferencing, perspective taking, monitoring, goal setting), topic and discourse knowledge, and socio-emotions. Importantly, however, DIEW posits differential or dynamic contributions of language, cognitive, and transcription skills to writing as a function of the dimensions of written composition. In other words, relative contributions of language and cognitive skills to writing differ depending on the dimensions of composition such as writing quality or productivity. Specifically, writing quality relies on multiple skills such as language, higher order cognitive skills (e.g., reasoning, perspective taking), and transcription skills because coherence of expressed ideas (i.e., writing quality) requires use of accurate and rich language and logical sequences of ideas as well as transcription skills (Kim et al., 2014; Kim & Graham, 2021). In contrast, writing productivity—the amount of writing—more heavily relies on transcription skills (spelling and handwriting; Kim et al., 2014; Kim & Graham, 2021). Then the effects of writing instruction likely vary depending on focal dimensions of written composition and depending on the nature of writing instruction. For instance, improving transcription skills via transcription instruction is expected to improve overall writing but its impact might be greater on writing productivity than writing quality because transcription would have more immediate effect on the amount of writing (writing productivity) whereas writing quality also relies on a greater number of skills in addition to transcription skills (Kim et al., 2014, 2021). To make an impact on writing quality, instruction on a more comprehensive set of skills and knowledge is necessary, and therefore, a multi-component instructional approach such as Self-Regulation Strategy Development (SRSD), which explicitly addresses text structure knowledge and self-regulation strategies (Harris & Graham, 2016; Harris, Graham, MacArthur, Reid, & Mason, 2011), is likely to have a greater effect than instructional approaches that target one specific skill (e.g., transcription).

The dynamic relations hypothesis as a function of development states that transcription skills have a large influence on writing during the initial phase of writing development due to its constraining role whereas higher order cognitive skills (e.g., reasoning) would exert a greater influence at a later phase (Kim, 2020; Kim & Park, 2019). Then, the effects of writing instruction might vary depending on child characteristics such that for students in the initial phase of writing development, addressing transcription skills might improve their written composition by reducing a large constraining role of transcription. To the best of our knowledge, the critical dynamics of instructional factors, composition skills, and child characteristics, as well as how these factors interact with the effectiveness of writing instruction have not been addressed in any of the existing meta-analyses for students in primary grades.

Our goal in the present meta-analysis was to systematically capture the effect of writing instruction for students in primary grades when writing and writing related skills are developing at a rapid rate. Of particular interest was differential effects of writing instruction as a function of the dimension of composition, instructional focus, and students' baseline writing skill. The following were specific research questions that guided the present meta-analysis.

1. What is the overall effect of writing instruction interventions on written composition skill for children in primary grades (kindergarten to Grade 3)?

2. Does the effect differ for various dimensions of composition (writing quality, writing productivity, writing fluency, and 'other') and as a function of instructional focus (e.g., SRSD, transcription, and mixed)?

3. Does the effect vary by the dosage of writing instruction?

4. Does the effect vary for writers with weaker initial writing skills versus those with typical writing skills on various dimensions of written composition?

5. Does the effect of instructional focus on various writing outcomes vary for writers with weaker writing skills versus those with typical writing skills?

We anticipated that writing instruction would have an overall positive effect on writing skill in line with previous studies (e.g., Graham et al., 2012). We also posited that the effects would vary depending on the dimensions of written composition and that the effects are likely larger for writers with weaker writing skills. We did not have a specific hypothesis regarding dosage of instruction due to lack of prior evidence. Importantly we hypothesized that the effects for writers with weaker writing skills might moderate instructional focus such that instruction that targets transcription skills might have a larger effect for writers with weaker writing skills in primary grades.

2. Method

2.1. Search Procedures and Inclusion Criteria

A search of studies was conducted using electronic databases, including ERIC, PsychINFO, ProQuest, and Google Scholar. Additionally, a manual search was conducted by reviewing references of the studies that met our criteria. We included various terms in line with inclusion and exclusion criteria below. Similar terms within criteria were separated with "OR" and included terms like *writing intervention, writing skills, motivation writing, writing process, writing routines, write frequently, writing goals, writing tools, writing feedback, writing knowledge, genre knowledge, discourse knowledge, transcription, sentence, vocabulary, spelling, drafting, revising, editing,* and *encoding.* Please see Appendix A for a comprehensive list of the search items.

Inclusion criteria were: (a) studies were published between January 2000 and April 2019; (b) students were in kindergarten to third grade; (c) the treatment group received specific target writing instruction whereas the control group received business-as-usual instruction, no instruction, or alternative instruction; (d) studies employed a randomized controlled trial or quasi-experimental design (pretest or regression discontinuity model); (e) instruction was provided in English and in a classroom setting during the regular academic school year; (f) sample sizes were no fewer than 20 students in total (at least 10 students per control and treatment group); (g) reported adequate information to calculate effect sizes for written composition outcome measures (if not, the primary author was contacted); (h) the intervention was conducted in regular school settings (after-school programs or tutoring programs in out-of-school settings were excluded).

2.2. Study Selection and Exclusion

Figure 1 presents a Prisma chart. The initial search procedure in electronic data bases yielded 1,367 articles. Manual search via the references of articles yielded 704 articles. After removing duplicates, a total of 1,229 articles were first screened by their titles and abstracts (first screen) and the resulting 152 studies were reviewed in the second round of screening. Studies were excluded at the second round of screening due to a number of reasons, including: the study design was not experimental or quasi-experimental (e.g., Little et al., 2010); effect sizes were combined with other grade levels that are beyond the scope of the current meta-analysis (e.g., Ukrainetz, 2000); the intervention was not conducted in a regular school setting (e.g., Lembke & Deno, 2003); the outcome measures did not target any compositional skills (Conrad, 2008); and lack of information to calculate effect size (e.g., Berninger et al., 2006). Twenty-three articles or 24 studies (see Table 1) met the inclusion criteria, which yielded 166 effect sizes ($N = 5589$ participants). These included unpublished work such as dissertations ($n = 5$) and reports ($n = 2$) along with published articles to mitigate publication bias.

2.3. Coding Procedures

The following characteristics were coded: a) sample demographic characteristics (e.g., sex), b) dimension of composition outcomes, c) nature of writing instruction (e.g., transcription, SRSD; intensity/dosage), d) student characteristics (students with weak or typical writing skills), and e) other study features (e.g., quality of study, publication status [published vs. not], research design [randomized vs. quasi-experimental]).

2.3.1.   Coding dimensions of composition outcomes—Regarding the dimension of composition outcomes, we first recorded the composition outcomes reported in each study (e.g., the Wechsler Individual Achievement Test Written Expression). Then, composition outcomes were further categorized into dimensions such as quality, productivity, writing fluency, and 'writing: other.' *Writing quality outcomes* included those that focus on overall quality (e.g., holistic scoring) or analytic scoring focusing on word choice, quality of ideas, and organization and structure. This was based on prior empirical evidence that these different aspects are best described together as quality (Kim et al., 2014, 2015). *Writing productivity* refers to the length of the essay (e.g., total number of words written, number of sentences, number of paragraphs, and number of correctly spelled words). *Writing fluency outcomes* included measures such as the writing fluency task of the Woodcock-Johnson Test of Achievement (e.g., Woodcock, McGrew, & Mather, 2001), where students are asked to write as many sentences using given words in a specified time (3 min). The *writing: other* category included a mixture of composition outcomes that did not fit any of the previous three categories and were measured sparsely and thus could not form their own categories. Examples include planning time (Harris et al., 2006), capitalization (Swain et al., 2007), letter sequencing (Rosenthal, 2006), perceptions about writing (Rosenthal, 2006), and composition time (Graham et al., 2005).

2.3.2.   Coding writing instruction—The nature of writing instruction was categorized as follows based on the focal content of instruction: Self-Regulated Strategy Development (SRSD), transcription, or mixed. *SRSD* provides explicit instruction on multiple aspects

such as goal setting, planning, and motivation strategies, and teaches text structure and transition or linking words (e.g., first, finally; Graham & Harris, 2003). SRSD is used as its own category because of a relatively large number of studies in the included studies. *Transcription* instruction focuses on spelling and handwriting (including keyboarding) skills, which are necessary for written composition (Berninger & Winn, 2006; Kim, 2020; Kim & Park, 2019). The '*mixed*' category was writing instruction that included approaches that did not fit either SRSD or transcription *and* did not have sufficient number of studies to examine as a moderator. Some studies in this category had a single instructional focus (e.g., adding writing assessment; Kozlow, 2004) while others had more than multiple foci (e.g., writing assessment and cross-age tutoring; Paquette, 2008). Finally, the student characteristics included students with typical writing skills and those with weak writing skills at baseline. This was determined by the authors in included studies using various screening criteria including writing assessment and teacher report.

In addition, dosage of writing instruction was coded. We recorded three dimensions of intervention intensity: number of weeks the intervention was provided to the students, how many sessions per week, and the length of each session. After collecting this information, we calculated the dosage of the intervention by multiplying the three dimensions. For example, if students in the treatment group received a writing intervention for 4 weeks, 5 sessions per week and 30 minutes per session, the dosage is 600 minutes (4*5*30). Interrater reliability among coders was 93%. Any discrepancies were resolved with discussion.

2.3.3. Study Quality—The quality of the included studies was assessed using a modified version of the Quality Assessment Tool for Quantitative Studies (National Collaborating Centre for Methods and Tools, 2008). We deleted questions that were not applicable to our study on education focus and focused on the following sections: component rating, study design, confounders, blinding, and withdrawals and drop-outs. A strong quality study met the following criteria: (a) the research question or objective was clearly stated; (b) the treatment and control samples were selected from the same or similar population; (c) inclusion and exclusion criteria for being in the study were pre-specified and applied uniformly to all participants; (d) for longitudinal studies, attrition was reported in terms of numbers and reasons per group; (e) the researchers were blinded to who received the treatment and control; (f) information randomization was clearly presented; (g) information on number of students is clearly stated; and (h) treatment and control pretest scores did not differ. These factors helped the coders determine and attribute a rating to the articles as weak, moderate, or strong. Thirteen studies (54%) were rated "strong" (those that met all the criteria or did not meet one criterion); nine studies (38%) were rated as "moderate" in quality (did not meet two or three criteria); and two studies were rated as "weak" (did not meet four or more criteria).

2.4. Meta-Analytic Procedures

Overall effects of the writing instruction on students' written composition were calculated by comparing the posttest score differences (and pretests when available) between the treatment and control groups. All students in the treatment groups received writing instruction but the nature of instruction for students in the control group varied across

studies (e.g., business-as-usual, or alternative instruction). When calculating effect sizes, we only focused on students' written composition. For the studies where students' writing skills were measured at various time points, only the posttest scores immediately after the intervention were included. If the study reported separate analytical scores on various dimensions of writing, we generated an overall effect size by averaging the standardized mean differences in each aspect of composition outcome between treatment and control groups (Lipsey & Wilson, 2001). For the other research questions, we examined four dimensions of written composition: quality, productivity, writing fluency, and writing: other.

2.4.1. Computation of the Effect Sizes—Cohen's d, a widely used effect size, indicating standardized difference between the two means was computed by dividing the mean score difference between treatment and control group by the pooled standard deviation (Lipsey & Wilson, 2001). Please see Appendix B for details. We calculated weighted effect sizes using the inverse of the study variance to multiply the effect sizes (Graham & Perin, 2007). Random effect models were used because the effects of writing instruction on students' composition outcomes vary based on different factors including students' demographics, instruction implementation and study methodology. Overall weighted mean effect size $\bar{g}$ was calculated using equation (5):

$$\bar{g} = \frac{\sum (w_i * g_i)}{\sum w_i}$$

5)

where w_i is the weight for each study and g_i is the effect size.

2.4.2. Heterogeneity of Effect Sizes—The Q test was used to examine the variation in effect size estimates between studies and the I^2 statistic was used to quantify the percent of variation attributed to true heterogeneity (Borenstein, 2009). Specifically, I^2 statistics allowed us to calculate the percentage of variation in effect sizes across studies that is not due to chance (Higgins, Thompson, Deeks, & Altman, 2003). The values of I^2 can range from 0 to 100% and higher values suggest higher level of heterogeneity.

2.4.3. Moderation Analysis and Robust Variance Estimation—We conducted moderation analyses to examine whether the variances in the effect of writing instruction could be explained by the differences in the dimensions of written composition, types of writing instruction, and student characteristic—either students with weak writing skills or typical writing skills. The overall effect of writing instruction was calculated using a weighted mean based on estimated marginal variances. Effect sizes within studies cannot be assumed to be independent from each other. In order to account for assumed correlated effect sizes within studies and to minimize the loss of information when clustering effect sizes, robust variance estimation was used to reanalyze the data (Hedges, Tipton, & Johnson, 2010). We further conducted sensitivity analysis to examine whether the effect sizes hold as the within-study correlation (rho) varies from 0 to 1. The weighted effect size and the sensitivity analysis were computed using Robumeta package (version 2.0; Fisher, Tipton, & Hou, 2017) in R Studio (version 1.1.447; RStudio, Inc., 2016).

2.4.4.   Publication bias—We used funnel plot to detect the potential publication bias on our meta-analysis research findings (see Appendix C). Egger's linear regression test (Egger, Smith, Schneider, & Minder, 1997) was performed to further examine the risk of publication bias by regressing the effect size estimate of a study against the precision of the study, indexed by its standard error.

3.   Results

3.1.   Research Question 1: Overall Effect of Writing Instruction

Using data from $k = 166$ ($N = 5589$), we calculated the standardized mean difference for students in treatment and control conditions. Figure 2 displays the distribution of effect sizes for all the 24 included studies, showing that, on average, students in the treatment condition scored about one-third of a standard deviation higher than those in the control condition ($ES = .31$, $SE = .09$, $p < .001$). The 95% confidence interval ranged from .14 to .49, suggesting that overall, writing instruction had a positive impact on students' composition skill.

We found a high degree of heterogeneity among the studies, $Q = 173.49$, $df = 23$, $p < .001$, with the I^2 statistic revealing that 87% of the total variance were attributed to between-study differences rather than within-study sampling error. We then examined the heterogeneity of the writing instruction effects on the four different dimensions of written composition: writing quality, writing productivity, writing fluency, and writing: other. Once again, there was a high degree of heterogeneity among the studies, with $Q = 216.86$, $df = 46$, $p < .001$, with the I^2 statistic revealing that 79% of the total variance could be attributed to between-study differences.

3.2.   Research Question 2: Effects for Different Dimensions of Composition Outcomes and Different Instructional Focus

When we examined the average effect sizes by composition outcomes, we found similar effect sizes in writing quality ($ES = .32$, $SE = .13$, $p = .02$, $k = 74$; see Figure 3a), writing productivity ($ES = .31$, $SE = .11$, $p = .01$, $k = 45$; see Figure 3b), and writing: other ($ES = .34$, $SE = .12$, $p = .004$, $k = 33$) with a smaller effect on writing fluency ($ES = .15$, $SE = .42$, $p = .72$, $k = 14$). Table 2 further shows average effect sizes for the four composition outcomes by instructional focus. SRSD had consistent positive effect sizes, specifically 1.04 in writing quality ($p < .001$, $k = 42$), .72 in writing: other (e.g., element score, which looks at the extent of inclusion of different elements of text structure, $p < .001$, $k = 10$), and .59 in writing productivity ($p < .001$, $k = 22$). Note that no writing fluency outcomes were reported in SRSD studies, and therefore, results are not available.

Transcription-focused writing instruction did not yield statistically significant average effect sizes due to large variation across the included studies. For example, in Graham et al.'s (2002) study, second graders who struggled with spelling were provided with spelling instruction with a large effect on writing fluency (.81). In contrast, Jung et al. (2017) implemented data-based instruction with a focus on handwriting and spelling with students at risk and those with learning disabilities, and found a large negative effect (–2.71). The

'mixed' category of writing instruction had positive average effect sizes, ranging from .23 in writing quality ($p = .08$, $k = 18$) to .58 in writing productivity ($p = .02$, $k = 9$).

3.3. Research Question 3: Differential Effects by Instruction Dosage

We examined whether the effect of writing instruction varies by instruction dosage (total length of the instruction). This information was not reported in 6 of the 24 studies, which yielded 22% missing data (37 effect sizes) on this variable. Our analysis results showed that the moderation effect of dosage was not statistically significant in either linear models ($\beta = .000$, $p = .77$, $k = 129$, $N = 18$) or nonlinear models ($\beta = .000$, $p = .95$, $k = 129$, $N = 18$), suggesting that the impact of writing instruction on students' writing outcomes did not differ by instruction dosage.

3.4. Research Question 4: Effects for Writers with Weak versus Typical Skills

To examine whether writing instruction has differential effects as a function of students' initial writing status before instruction (i.e., weak writing skills vs. typical writing skills), we fit meta-regression models for the four different composition outcomes where intercept refers to the average effect size for the reference group, students with typical writing skills. As shown in Table 3, although average effect sizes differed, they were not statistically significant due to large variation across studies (i.e., standard errors). For example, regarding writing quality, the average effect size for students with weak writing skills was .47 ($k = 51$) larger than those with typical writing skills (intercept = .08, $k = 23$), but this did not reach the conventional statistical significance ($p = .13$) due to large standard error (.31).

3.5. Research Question 5: Differential Effects by Instructional Focus

We also examined whether the nature of instruction influenced students with weak versus typical writing skills differently in different composition outcomes (Research Question 4) except mixed condition because these studies did not include students with weak writing skills. The reference condition in the analysis (i.e., intercept) was the effect of transcription instruction for students with typical writing skills. As shown in Table 3, the only statistically significant effect was for writing quality outcome—SRSD had a larger effect size than transcription instruction ($\beta = 1.32$, $p < .001$) and the effect of writing instruction was larger for students with weak writing skills ($\beta = .92$, $p < .001$). For writing productivity outcome, there was a trend that SRSD had a larger effect for students with weak writing skills ($\beta = .70$), but this did not reach the conventional statistical significance ($p = .06$).

3.6. Sensitivity Analysis and Publication Bias

As part of robustness checks, we examined whether effects differed as a function of several characteristics of the included studies (e.g., publication status, quality of study). The effect size for studies published in peer-reviewed journals appeared to be .25 larger than non-published papers, but this difference was not statistically significant ($Q = 1.42$, $df = 1$, $p = .23$). Similarly, the effect size for studies with randomized control design was .19 larger than for quasi-experimental design, but again the difference was not statistically significant ($Q = .77$, $df = 1$, $p = .38$). Lastly, we examined whether results differ by the nature of

control conditions (i.e., business-as-usual, no instruction, or alternative interventions), and no differences were found in any of the writing outcomes (ps > .55).

In order to verify the robustness of our findings generated from Comprehensive Meta-Analysis Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2014), alternative statistical analysis using Robumeta package in R Studio was conducted to reanalyze the data. Robust Variance Estimation Correlated Effects Model with Small-Sample Corrections yielded a similar overall significant effect size of writing instruction with $g = .33$, $SE = .11$, $p = .01$. Sensitivity analysis using various values of rho yielded the same intercept and standard error, which indicated that the effect is robust to different values of rho. The identical result from R Studio confirmed the robustness of our previous estimation using CMA in handling dependent effect sizes.

Results of the Egger's regression and Begg and Mazumdara rank correlation (Begg, & Mazumdar, 1994) indicate that we cannot reject the null hypothesis of no publication bias (Egger's regression: Intercept = 1.39, $SE = .86$, $p = .12$; Kendall's S statistics: $Tau = .09$, Z $value = .60$, $p = .55$).

4. Discussion

The primary goal of the present meta-analysis was to estimate the causal impact of writing instruction for children in primary grades, and to investigate how impact varies as a function of different dimensions of composition outcomes, instructional factors (focal content and dosage), and students' baseline writing skills. Primary grades are an important period for laying foundations for writing skills, and therefore, information on effective writing instruction during this period is imperative. Overall, we found that writing instruction has a moderate positive effect (.31), which is in line with previous meta-analyses which showed that writing instruction does make a difference (Graham et al., 2012; Graham & Perin, 2007).

Beyond the average effect, however, the present study revealed that the effects varied depending on outcome of interest. We also found that the average effects of writing instruction varied not only by dimensions of composition outcomes but also by the nature of writing instruction. SRSD, consistent with previous meta-analyses (Graham, 2006; Graham et al., 2012), had positive and large effect sizes for students in primary grades. However, the effect sizes varied for different composition outcomes, with the largest effect size in writing quality ($ES = 1.04$), but also a substantial effect size in writing productivity ($ES = .59$). The large effects of SRSD over the other instructional foci examined in the present meta-analysis may be due to the nature of the construct of writing. As previously noted, writing is a complex construct that is supported by an individual's language, cognitive, and transcription skills, knowledge, and social-emotional skills. As such, to make an observable difference in students' writing outcomes beyond target component skills, instructional approaches that address multiple skills likely are advantageous and yield more consistent positive results than instructional approaches that address discrete skills alone. SRSD is a multi-component writing instructional program that targets several skills important for writing development, such as text structure for different genres, knowledge (discourse knowledge and vocabulary),

and strategies for self-regulation (e.g., goal setting, self-assessment, self-reinforcement). The present findings suggest that multi-component approaches, such as SRSD, are likely needed to improve writing outcomes even for primary grade students.

Unlike our hypothesis, transcription instruction was not found to have statistically significant effects on any of the composition outcomes, including writing productivity. We anticipated a positive effect of instruction on transcription skills particularly on writing productivity, given its essential role in writing especially for beginning writers (Berninger & Winn, 2006; Kim, 2020; Kim & Park, 2019). As noted above, the null result generally stems from large variation in effect sizes across studies—many yielded positive results, but several had large negative results. The reasons for the inconsistencies include several aspects of study design that could not be systematically accounted for in the meta-analysis due to idiosyncrasies across studies. For example, Jung et al. (2017) taught spelling and handwriting fluency using data-based instruction to students in Grades 1 to 3, and found a positive effect on the number of words written ($ES = .30$). In contrast, Graham et al.'s (2002) study examined the effect of spelling instruction for second graders with writing difficulties, and found that students in the control condition (math instruction) produced longer stories than those in the spelling instruction condition in the posttest ($ES = -.42$) primarily due to three outliers in the control condition (Graham et al., 2002). However, effect sizes without the three outliers were not reported. A negative effect size ($ES = -.43$) was also found in Berninger et al. (2000), and this appears to be due to sample characteristics. In this study, a follow-up transcription instruction was provided to a group of children who were classified as slow responders in an earlier intervention whereas those in the control group were fast responders who were not provided with additional follow-up instruction. Therefore, although children in both groups were initially identified and coded as struggling writers, they were not equivalent in terms of their responses to instruction and these characteristics likely influenced the negative effect (slow responders wrote fewer words even after a follow-up instruction). Because these potential explanatory characteristics differed across studies in a non-systematic way, these could not be accounted for in our statistical analysis.

It should be noted that the present findings do not negate the importance of transcription skills in written composition or instruction of transcription as part of writing instruction. A previous meta-analysis with students in kindergarten to Grade 12 revealed a large effect of handwriting fluency instruction on writing quality (.84; Santangelo & Graham, 2016) although the effect of spelling on writing outcomes was not statistically significant (Graham & Santangelo, 2014). Furthermore, previous studies have shown that instruction on transcription skills does improve students' spelling and handwriting fluency (Berninger et al., 2000, 2002; Graham et al., 2002; Wanzek et al., 2017), which are necessary for written composition. What the present findings indicate is that the effects of transcription instruction alone on writing composition outcomes are inconsistent at least for students in primary grades. Future investigations are necessary to further elucidate whether instruction on transcription skills alone is sufficient enough to improve written composition skill or whether instruction on other aspects (language [vocabulary], sentence uses, knowledge [e.g., text structure], self-regulation strategies) in addition to transcription skills is necessary for detectable improvement on written composition. Moreover, it is important for future studies to further elucidate a potentially differential impact of handwriting fluency versus spelling

on writing outcomes for students in primary grades. In the present study, we combined handwriting fluency and spelling instruction into a single category rather than examining them separately due to a small number of studies that targeted these separately (e.g., only two studies targeted handwriting alone).

Another characteristic of writing instruction, dosage, did not yield differential effects. However, caution needs to be taken for interpreting these results for a couple of reasons. First, by nature of the meta-analysis, studies of different nature are examined together. However, addressing the question of instructional dosage rigorously requires an experimental design where the same writing instruction is implemented with variation of dosage. Second, the question of instructional dosage would interact with fidelity of implementation because greater dosage will not result in greater effect sizes if implementation fidelity is low. We could not examine these nuances because many studies included in the present study did not report implementation fidelity. The present findings indicate a need for a systematic understanding of the details about delivery of writing instruction such as dosage. Studies to date have primarily focused on effective delivery of instructional content (explicit teaching of target skills) to improve writing outcomes, but the instructional dosage question has not garnered much attention. As such, many important questions remain open, including whether (and if so, to what extent) effects of particular writing instruction vary as a function of instruction dosage; whether the effects of dosage of instruction vary depending on student characteristics (e.g., struggling writers); and whether the effects of the dosage of instruction vary depending on writing outcomes (quality vs. productivity). For instance, the instructional dosage required for handwriting fluency instruction versus for language instruction to improve writing quality likely varies because language is a large domain or unconstrained skill (Snow & Kim, 2007; Snow & Matthews, 2016) whereas handwriting fluency is a more constrained skill. In addition, the instructional dosage required within a target domain, say handwriting fluency, would be greater for students whose primary difficulty is with handwriting compared to students who do not have such a difficulty.

The present findings also underscore variation of effects on composition skills by students' baseline writing skill level and instructional focus. When we simply compared effect sizes between students with weak versus typical writing skills across the four dimensions of composition, no statistically significant differences were observed although there were differences descriptively. When we took the nature of instruction into account, however, we found that the effect size was large for writers with weaker writing skills in the writing quality dimension compared to typical peers. SRSD had a larger effect on writing quality compared to transcription instruction, but the effect of SRSD instruction did not vary by students' baseline writing skill. Although the hypothesis of the impact of transcription instruction on writing productivity and for writers with weaker writing skills was not supported in the present meta-analysis, findings do indicate differential effects. Therefore, for children in primary grades, although transcription skills are important, writing instruction that targets transcription skills alone without systematic teaching of other important skills (e.g., language, self-regulation, text structure) may not have a consistent impact on composition skills, including the theoretically most likely dimension of writing productivity. Overall, these findings allow us to develop a more nuanced and refined picture about what

type of writing instruction is effective for whom and for what writing outcomes, which is necessary for a more precise approach to individualizing writing instruction to meet students' needs.

5. Limitations and Future Directions

The results of the present study should be interpreted with the following limitations in mind. First, sample sizes in the vast majority of studies were small—the majority of the studies had 10 students to 62 students with an exception of three studies which included more than 100 students. Although the effect of sample size was taken into consideration by using a weighted mean in data analysis, estimates are not as robust with small sample sizes. Therefore, future studies of writing instruction with a larger number of participants are certainly needed. Second, our classification of the nature of writing instruction included three categories, SRSD, transcription, and mixed. Note that these categories were developed after examining the focal skills targeted in the included studies, and were not a priori determined. In other words, these reflect the status of the current literature on writing instruction for students in primary grades. The mixed category included a variety of instructional approaches with one study examining each approach, and therefore, future studies are necessary to evaluate their effect. Similarly, the "writing: other" category of the writing dimension included a variety of composition outcomes and therefore, it is difficult to draw any practical interpretation of the results in this composition outcome. Taken together, these results indicate a need for greater efforts that target multiple components to improve primary graders' writing skills. Although a multi-component instructional approach, SRSD, was shown to be effective in the current meta-analysis as well as in previous ones with older students (Graham et al., 2012), the present meta-analysis revealed that a limited number of studies systematically target multiple component skills of writing using approaches other than SRSD for primary grade students.

Furthermore, we classified writing outcomes into several categories based on prior evidence that certain aspects (e.g., word choice, idea quality, and structural organization) are best described as a common dimension (e.g., writing quality), which is associated with but dissociable from other dimensions (e.g., writing productivity and conventions; Kim et al., 2014, 2015). However, we recognize that differences in writing evaluation approaches such as holistic approaches, analytic approaches, and approaches that look at syntactic structures and writing productivity matter for deciding one's writing ability for adult writers (Schoonen, 2005; Van den Bergh et al., 2012) and beginning writers (Kim, Schatschneider, Wanzek, Gatlin, & Al Otaiba, 2017). Lastly, in the current meta-analysis we included the studies between 2000 and 2019 in order to reflect studies conducted in contemporary classrooms. Future replications where studies before the year 2000 are included are warranted.

Despite these limitations, the study points to important next steps in the study of writing instruction and how it affects students' skills. First, more research is needed. These findings overall indicate the importance of teasing out and considering student characteristics, the nature of instruction, and composition outcomes to understand the impact of writing instruction on individual students' writing achievement, aligned with DIEW (Kim, 2020;

Kim & Park, 2019). The present findings also support the importance of addressing multiple components in writing instruction, as employed by SRSD. Of course, instructional approaches that address multiple skills are not limited to SRSD, but SRSD had a sufficient number of studies to be classified as its own category. Future work on multi-component instruction in writing that is aligned with theoretical models and empirical evidence is needed to investigate the promise of efficacy and scalability so that writing skills improve for all students. Another crucial point we would like to reiterate is a need for systematic and detailed understanding about effective delivery of writing instruction—what instructional content, for whom, and under what conditions. Writing development is undergirded by multiple language and cognitive skills as specified by theoretical models. Studies, including the present study, previous meta-analyses, and individual empirical studies, have provided rich information about skills to target and their effects on writing outcomes. However, substantially less evidence is available about nuances that are necessary for practice—what to teach for which students in what manner (e.g., when, how long, and what activities). Future studies are warranted to address these important gaps in the field.

Acknowledgments

Appendix

Appendix A: Search terms used

Writ* OR "Writing intervention" OR "Writing Skills" OR "Motivat* writing" OR "Writing Process" OR "writing routines" OR write frequently OR "writ* goals" OR "Writ* Tools" OR "writ* feedback" OR "writing knowledge" or "genre knowledge" OR "discourse knowledge" OR transcription OR Sentence OR vocabulary OR spelling OR draft* OR revis* OR edit* OR encod*

AND

"Low read*" OR "low skilled reader" OR dyslexia OR DYSLE* OR "reading disability" OR dysgraphia OR "writing disability" OR "special education" OR "tier 2" OR "tier 3" OR "struggle to read" OR "learning disability" OR "severe learning disability" OR "specific learning disability" OR "co-morbid*" OR "executive function*" OR "neurodevelopmental disorder" OR "struggling read*" OR "weak read*" OR "poor read*"

AND

Kindergarten OR "first grade" OR "Grade one" OR "Grade 1" OR "1st grade" OR "second grade" OR "2nd grade" OR "Grade two" OR "Grade 2" OR "third grade" OR "3rd grade" OR "Grade 3" OR "Grade three" OR "elementary school" OR "age four" OR "age five" OR "age six" OR "age seven" OR "age eight" OR "age nine" OR "4 year old" OR "5 year old" OR "6 year old" OR "7 year old" OR "8 year old" OR "9 year old"

AND

"Random assignment" OR "RCT" OR "random control trial" OR "Randomized control trial" or "randomized controlled trial" OR experiment* OR causal OR affect OR effect OR mediation OR moderation OR "individual differences" OR interact* OR "aptitude by treatment"

AND

Poverty OR low-socioeconomics or "low SES" OR "parent* education" OR "free or reduced lunch" OR "free lunch" OR "reduced lunch" OR "FRL" OR "SES" OR age OR "grade level" OR minority OR Hispanic OR Latino OR Black OR "African American" OR "African-American" OR "Asian American" OR "Asian-American" OR "English language learn*" OR "English learn*" OR "limited English proficient" OR "non-native English speakers" OR bilingual OR "emerging bilingual students" OR "language minority" OR "English as a Second Language" OR "English for speakers of Other Languages" OR "dual language learn*" OR "EL" OR "ELL" OR "LEP" OR "ESL" OR "ESOL" OR "DLL" OR "executive funct*" OR "EF" OR "working memory" OR "inhibitory control" OR "shift*" OR "attention*" OR self-regulation OR genetics OR "genetics by instruction" OR "genetics by culture" OR "family history" or "Heredit*"

AND

intervention or treatment or training

Appendix B: Details about effect sizes

Pooled standard deviation was calculated using equation (1):

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{1}$$

where n_1 and n_2 are the sample sizes for the two groups and s_1 and s_2 are the sample standard deviations.

For studies in which only posttest scores were reported, the between-group mean score difference was calculated by subtracting the mean score on treatment posttest and mean score on control posttest using equation (2):

$$d = \frac{\overline{M}_2 - \overline{M}_1}{SD_{pooled}} \tag{2}$$

where $\overline{M}_1$ and $\overline{M}_2$ are the posttest mean scores.

For studies where both pretest and posttest scores were reported, the between-group mean score difference was calculated by subtracting the mean pretest scores from the mean posttest scores for both treatment and control groups first and then subtracting two mean difference between treatment and control group:

$$d = \frac{(\overline{M}_{post1} - \overline{M}_{pre1}) - (\overline{M}_{post2} - \overline{M}_{pre2})}{SD_{pooled}} \qquad (3)$$

To adjust for upward bias in Cohen's *d* associated with small sample sizes ($n < 20$), we transformed all effect sizes to Hedge's *g* (Hedges & Olkin, 1985) using the following equation:

$$g = \left[1 - \left(\frac{3}{4(n_1 + n_2) - 9}\right)\right] * d \qquad (4)$$

where *d* is referred to Cohen's *d*.

Appendix C: Funnel plot of standard error by Hedge's g.
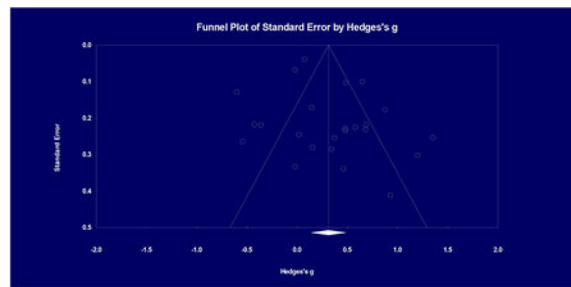
References

Abbott RD, & Berninger VW (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary and intermediate-grade writers. Journal of Educational Psychology, 85, 478–508. doi:10.1037/0022-0663.85.3.478

Ahmed Y, Wagner RK, & 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, 419–434. DOI: 10.1037/a0035692 [PubMed: 24954951]

Begg CB, & Mazumdar M (1994). Operating characteristics of a rank correlation test for publication bias. Biometrics, 1088–1101. doi:10.2307/2533446 [PubMed: 7786990]

* Berninger VW, Vaughan K, Abbott RD, Brooks A, Begayis 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, 117–135. doi:10.2307/1511141

* Berninger VW, Vaughan K, Abbott RD, Begay K, Coleman KB, Curtin G, Hawkins JM & Graham S (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. Journal of Educational Psychology, 94, 291. doi:10.1037/0022-0663.94.2.291

Berninger VW, Rutbert JE, Abbott RD, Garcia N, Anderson-Youngstrom M, Brooks A, & Fulton C (2006). Tier 1 and Tier 2 early intervention for handwriting and composing. Journal of School Psychology, 44, 3–30. doi:10.1016/j.jsp.2005.12.003

Berninger VW, & Winn WD (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In MacArthur C, Graham S, & Fitzgerald J (Eds.), Handbook of writing research (pp. 96–114). New York: Guilford.

Borenstein M, Hedges L, Higgins J, & Rothstein H (2013). Comprehensive Meta-Analysis Version 3. Englewood, NJ: Biostat.

* Carlisle AA (2016). The effect of a morphological awareness intervention on early writing outcomes (Doctoral dissertation, University of Missouri--Columbia).

Chenoweth NA, & Hayes JR (2001). Fluency in writing: Generating text in L1 and L2. Written Communication, 18(1), 80–98. 10.1177/0741088301018001004

Borenstein M, Hedges L, Higgins J, & Rothstein H (2014). Comprehensive Meta-Analysis Version 3. Englewood, NJ: Biostat.

Connor CM (2016). A lattice model of the development of reading comprehension. Child Development Perspectives, 10, 269–274. doi:10.1111/cdep.12200 [PubMed: 27867420]

Connor CM, Morrison FJ, Fishman B, Crowe EC, Al Otaiba S, & Schatschneider C (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading form first through third grade. Psychological Science, 24, 1408–1419. doi:10.1177/0956797612472204 [PubMed: 23785038]

Connor CM, Piasta SB, Fishman B, Glasney S, Schatschneider C, Crowe E, Underwood P, & Morrison F (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. Child Development, 80, 77–100. doi:10.1111/j.1467-8624.2008.01247.x [PubMed: 19236394]

Conrad NJ (2008). From reading to spelling and spelling to reading: Transfer goes both ways. Journal of Educational Psychology, 100(4), 869. 10.1037/a0012544

Cooper H, Hedges LV, & Valentine JC (2009). The handbook of research synthesis and meta-analysis. New York, NY: Russell Sage Foundation.

Effective Public Health Practice Project. (1998). Quality assessment tool for quantitative studies. Hamilton, ON: Effective Public Health Practice Project. Available from: https://merst.ca/ephpp/

Egger M, Smith GD, Schneider M, & Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. Bmj, 315, 629–634. doi:10.1136/bmj.315.7109.629 [PubMed: 9310563]

Fellner T, & Apple M (2006). Developing writing CBM writing and lexical complexity with blogs. The Jalt Call Journal, 2, 15–26.

Graham S (2006a). Strategy instruction and the teaching of writing: A meta-analysis. In MacArthur CA, Graham S, & Fitzgerald J (Eds.), Handbook of writing research (pp. 187–207). New York, NY: Guilford Press.

* Graham S, Harris KR, & Chorzempa BF (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. Journal of educational psychology, 94, 669. doi:10.1037/0022-0663.94.4.669

Graham S, & Harris KR (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In Swanson HL, Harris KR, & Graham S (Eds.), Handbook of learning disabilities (pp. 323–344). The Guilford Press.

Graham S, & Harris KR (2000). The role of self-regulation and transcription skills in writing and writing development. Educational psychologist, 35, 3–12. doi:10.1207/S15326985EP3501_2

*Graham S, Harris KR, & Fink B (2000). Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers. Journal of educational psychology, 92, 620. doi:10.1037/0022-0663.92.4.620

*Graham S, Harris KR, & Mason L (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. Contemporary Educational Psychology, 30, 207–241. doi:10.1016/j.cedpsych.2004.08.001

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

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

Graham S, & Sandmel K (2011). The process writing approach: A meta-analysis. Journal of Educational Research, 104, 396–407. doi:10.1080/00220671.2010.488703

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

Harris KR, & Graham S (2016). Self-regulated strategy development in writing: Policy implications of an evidence-based practice. Policy Insights from Behavioral and Brain
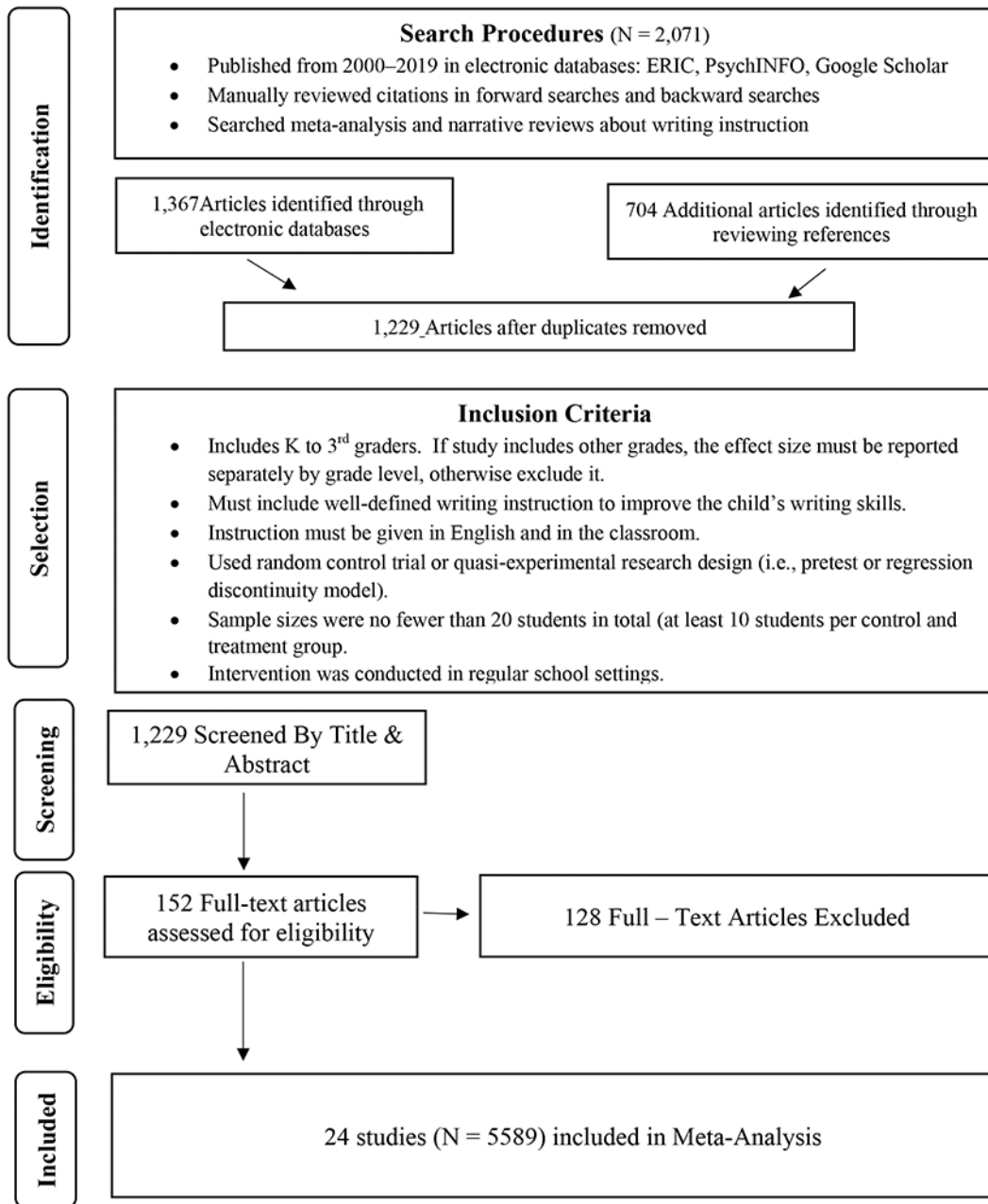
Sciences, 3, 77–84. Retrieved from http://bbs.sagepub.com/cgi/reprint/2372732215624216v1.pdf?ijkey=SgqI7iI7IIrq7ln&keytype=finite

Harris KR, Graham S, MacArthur C, Reid R, & Mason L (2011). Self-regulated learning processes and children's writing. Zimmerman B & Schunk DH (Eds.), Handbook of self-regulation of learning and performance (pp. 187–202). London, UK: Routledge Publishers.

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

*Harris K (2015). Practice-based professional development and Self-Regulated Strategy Development for Tier 2, at-risk writers in second grade. Contemporary Educational Psychology, 40, 5–16. doi:10.1016/j.cedpsych.2014.02.003

Hayes J & Flower L (1981). A cognitive process theory of writing. College Composition and Communication, 32, 365–387. doi:10.2307/356600

Hedges LV & Olkin I (1985). Statistical methods for meta-analysis. New York, NY: Academic Press.

Hedges LV, Tipton E, & Johnson MC (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1, 39–65. doi:10.1002/jrsm.5 [PubMed: 26056092]

Higgins JP, Thompson SG, Deeks JJ, & Altman DG (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327 (7414), 557. doi:10.1136/bmj.327.7414.557 [PubMed: 12958120]

Higgins JPT, Altman DG, & Sterne JAC (2017). Assessing risk of bias in included studies. In Higgins JPT, Churchill R, Chandler J, & Cumpston MS (Eds.), Cochrane handbook for systematic reviews of interventions version 5.2.0. Retrieved from www.training.cochrane.org/handbook.

*Hooper SR, Costa LC, McBee M, Anderson KL, Yerby DC, Childress A, and Knuth SB (2013). A written language intervention for at-risk second students: a randomization controlled trial of the process assessment of the learner lesson plans in a tier 2 response-to-intervention (RtI) model. Annals of Dyslexia, 63, 44–64. doi:10.1007/s11881-011-0056-y [PubMed: 21837551]

Johnson MD, Mercado L, & Acevedo A (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. Journal of Second Language Writing, 21(3), 264–282. doi: 10.1016/j.jslw.2012.05.011

*Jones D (2004). Automaticity of the transcription process in the production of written text. (Unpublished doctoral dissertation, University of Queensland, Australia).

Juel C, Griffith PL, & Gough PB (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. Journal of Educational Psychology, 78 (4). 243–255.

*Jung PG, McMaster KL, & delMas RC (2017). Effects of early writing intervention delivered within a data-based instruction framework. Exceptional Children, 83, 281–297. doi:10.1177/0014402916667586

Kim Y-SG (2020). Structural relations of language, cognitive skills, and topic knowledge to written composition: A test of the direct and indirect effects model of writing (DIEW). British Journal of Educational Psychology. doi:10.1111/bjep.12330

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

Kim Y-S, Al Otaiba S, & Wanzek J (2015). Kindergarten predictors of third grade writing. Learning and Individual Differences, 37, 27–37. doi:10.1016/j.lindif.2014.11.009 [PubMed: 25642118]

Kim Y-S, Al Otaiba S, Wanzek J, & Gatlin* B (2015). Towards an understanding of dimension, predictors, and gender gaps in written composition. Journal of Educational Psychology, 107, 79–95. doi:10.1037/a0037210 PMCID: PMC4414052 [PubMed: 25937667]

Kim Y-SG, & Graham S (2021). Expanding the direct and indirect effects model of writing (DIEW): Dynamic relations of component skills to various writing outcomes. Journal of Educational Psychology. 10.1037/edu0000564

Kim Y-SG, & Park S (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. Reading and Writing: An Interdisciplinary Journal, 32, 1319–1343. doi:10.1007/s11145-018-9913-y

Kim Y-SG, Schatschneider C, Wanzek J, Gatlin* B, Al Otaiba S (2017). Writing evaluation: Rater and task effects on the reliability of writing scores for children in Grades 3 and 4. Reading and Writing: An Interdisciplinary Journal, 30, 1287–1310. doi:10.1007/s11145-017-9724-6

*Kozlow M, Bellamy P (2004). Experimental Study on the Impact of the 6+1 Trait Writing Model on Student Achievement in writing. Portland, OR: Northwest Regional Educational Laboratory.

*Lane KL, Harris K, Graham S, Driscoll SA, Sandmel K, Morphy P, … Schatschneider C (2011). Sefl-regulated strategy development at tier 2 for second-grade students with writing and behavioral difficulties: A randomized control trial. Journal of Research on Educational Effectiveness, 4, 322–353. doi:10.1080/19345747.2011.558987

Lipsey MW, & Wilson DB (2001). Practical meta-analysis (2nd ed.). Thousand Oaks, CA: Sage Publications.

Little MA, Lane KL, Harris K, Graham S, Brindle M, & Sandmel K (2010). Self-regulated strategies development for persuasive writing in tandem with schoolwide positive behavioral support: Effects for second grade students with behavioral and writing difficulties. Behavior Disorders, 35, 157–179. 10.1177/019874291003500206

Mackie C, & Dockrell JE (2004). The nature of written language deficits in children with SLI. Journal of Speech, Language, and Hearing Research, 47, 1469–1483. doi:10.1044/1092-4388(2004/109)

McCutchen D (2006). Cognitive factors in the development of children's writing. In MacArthur CA, Graham S, and Fitzgerald J (Eds.), Handbook of Writing Research (pp. 115–130). New York, NY: The Guilford Press.

National Center for Education Statistics (2003). The nation's report card: Writing 2002 (NCES 2003–529). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Education Statistics (2012). The nation's report card: Writing 2011 (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Educational Statistics. (2020). https://nces.ed.gov/nationsreportcard/

National Collaborating Centre for Methods and Tools. (2008). Quality assessment tool for quantitative studies. Retrieved from http://www.nccmt.ca/resources/search/14

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards. Washington, DC: Author.

Olinghouse NG, & Graham S (2009). The relationship between discourse knowledge and the writing performance of elementary-grade students. Journal of Educational Psychology, 101, 37–50. doi:10.1037/a0013462

*Paquette KR (2008). Integrating the 6 1 writing traits model with cross-age tutoring: An investigation of elementary students' writing development. Literacy Research and Instruction, 48, 28–38. doi:10.1080/193880708s02226261

*Puma M, Tarkow A, and Puma A (2007). The challenge of improving children's writing ability: A randomized evaluation of Writing Wings. Institute of Education Sciences, US Department of Education.

*Puranik CS, Patchan MM, Lemons CJ, & Al Otaiba S (2017). Using peer assisted strategies to teach early writing: Results of a pilot study to examine feasibility and promise. Reading and Writing, 30, 25–50. doi:10.1007/s11145-016-9661-9

Roberts C (2002). The influence of teachers' professional development at the Tampa Bay Area Writing Project on student writing performance (Unpublished doctoral dissertation). University of South Florida, Tampa.

*Roberts TA, & Meiring A (2006). Teaching phonics in the context of children's literature or spelling: Influences on first-grade reading, spelling, and writing and fifth-grade comprehension. Journal of Educational Psychology, 98, 690–713. doi:10.1037/0022-0663.98.4.690

*Rosenthal B (2006). Improving elementary-age children's writing fluency: A comparison of improvement based on performance feedback frequency (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3242508)

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

Schoonen R (2005). Generalizability of writing scores: An application of structural equation modeling. Language Testing, 22, 1–30.

*Shorter LL (2001). Keyboarding versus handwriting: Effects on the composition fluency and composition quality of third grade students (Unpublished doctoral dissertation, University of South Alabama).

*Sinclair JP (2005). The effects of explicit instruction on the structure of urban minority children's academic writing (Doctoral dissertation, The Catholic University of America.)

Snow CE, & Kim Y–S (2007). Large problem spaces: The challenge of vocabulary for English language learners. In Wagner RK, Muse A, & Tannenbaum K, (Eds.), Vocabulary acquisition and its implications for reading comprehension (pp. 123–139). New York: Guilford Press.

Snow CE, & Matthews TJ (2016). Reading and language in the early grades. The Future of Children, 26 (1), 57–74.

Stotsky S (1983). Research on reading/writing relationships: A synthesis and suggested directions. Language Arts, 60, 627–642.

*Swain SS, Graves RL, & Morse DT (2007). Effects of NWP teaching strategies on elementary students' writing. Berkeley, CA: National Writing Project, University of California.

*Tracy B, Reid R, & Graham S (2009). Teaching young students strategies for planning and drafting stories: The impact of self-regulated strategy development. The Journal of Educational Research, 102, 323–332.

Ukrainetz TA, Cooney MH, Dyer SK, Kysar AJ, & Harris TJ (2000). An investigation into teaching phonemic awareness through shared reading and writing. Early Childhood Research Quarterly, 15(3), 331–355. doi: 10.1016/S0885-2006(00)00070-3

Van den Bergh H, De Maeyer S, Van Weijen D, & Tillema M (2012). Generalizability of text quality scores. In Measuring writing: Recent insights into theory, methodology and practice (pp. 23–32). Leiden-Boston: Brill.

Wagner RK, Puranik CS, Foorman B, Foster E, Tschinkel E, & Kantor PT (2011). Modeling the development of written language. Reading and Writing, 24, 203–220. doi:10.1007/s11145-010-9266-7 [PubMed: 22228924]

*Wanzek J, Gatlin B, Al Otaiba S, & Kim Y-SG (2017). The impact of transcription writing interventions for first-grade students. Reading & Writing Quarterly, 33, 484–499. doi:10.1080/10573569.2016.1250142 [PubMed: 28989267]

Woodcock RW, McGrew KS, & Mather N (2001). Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside.

Figure 1.
Study search procedures and selection criteria.

Meta-analysis Main Effect

Model	Study name	Statistics for each study				Hedges's g and 95% CI
		Hedges's g	Lower limit	Upper limit	p-Value	
	Shorter, 2001	-0.601	-0.855	-0.347	0.000	
	Hooper, 2013	-0.540	-1.059	-0.022	0.041	
	Berninger, 2000a	-0.422	-0.851	0.006	0.054	
	Carlisle, 2016	-0.358	-0.788	0.073	0.103	
	Kozlow, 2004	-0.019	-0.154	0.116	0.783	
	Paquette, 2008	-0.017	-0.671	0.637	0.959	
	Jung, 2017	0.021	-0.463	0.504	0.934	
	Puma, 2007	0.080	0.002	0.157	0.043	
	Graham, 2002	0.147	-0.189	0.483	0.392	
	Berninger, 2002	0.156	-0.393	0.706	0.577	
	Berninger, 2000b	0.344	-0.217	0.904	0.230	
	Roberts, 2006	0.376	-0.124	0.876	0.140	
	Puranik, 2017	0.464	-0.202	1.129	0.172	
	Sinclair, 2005	0.480	0.021	0.940	0.041	
	Rosenthal, 2006	0.482	0.032	0.931	0.036	
	Tracy, 2010	0.487	0.283	0.691	0.000	
	Wanzek, 2017	0.582	0.138	1.026	0.010	
	Swain, 2007	0.652	0.455	0.850	0.000	
	Graham, 2000	0.685	0.231	1.140	0.003	
	Lane, 2011	0.689	0.261	1.116	0.002	
	Graham, 2005	0.879	0.531	1.226	0.000	
	Jones, 2004	0.932	0.125	1.738	0.024	
	Harris, 2015	1.202	0.609	1.795	0.000	
	Harris, 2006	1.355	0.857	1.852	0.000	
Fixed		0.169	0.117	0.221	0.000	
Random		0.313	0.139	0.486	0.000	

Figure 2.

Forest plot of the meta-analysis main effect size of writing instruction ($N = 24$).

Main Effect for Writing Quality

Model	Study name	Statistics for each study				Hedges's g and 95% CI
		Hedges's g	Lower limit	Upper limit	p-Value	
	Shorter, 2001	-0.975	-1.422	-0.529	0.000	
	Hooper, 2013	-0.540	-1.059	-0.022	0.041	
	Carlisle, 2016	-0.212	-0.946	0.522	0.571	
	Graham, 2002	-0.028	-0.603	0.548	0.925	
	Paquette, 2008	-0.017	-0.671	0.637	0.959	
	Kozlow, 2004	0.028	-0.163	0.219	0.772	
	Puma, 2007	0.085	-0.024	0.195	0.127	
	Berninger, 2002	0.278	-0.503	1.060	0.485	
	Tracy, 2010	0.345	-0.005	0.695	0.053	
	Sinclair, 2005	0.371	-0.273	1.016	0.259	
	Graham, 2000	0.540	-0.094	1.175	0.095	
	Swain, 2007	0.693	0.413	0.974	0.000	
	Lane, 2011	0.775	0.162	1.388	0.013	
	Graham, 2005	1.170	0.548	1.792	0.000	
	Harris, 2015	1.202	0.609	1.795	0.000	
	Harris, 2006	1.978	1.184	2.773	0.000	
Fixed		0.161	0.083	0.239	0.000	
Random		0.315	0.061	0.569	0.015	

-3.00   -1.50   0.00   1.50   3.00

Figure 3a.

Forest plot of the effect sizes of the writing instruction on students' writing quality outcome ($N = 16$, $k = 74$).

Main Effect for Writing Productivity

Model	Study name	Hedges's g	Lower limit	Upper limit	p-Value
		Statistics for each study			
	Berninger, 2000a	-0.422	-0.851	0.006	0.054
	Graham, 2002	-0.224	-0.803	0.354	0.447
	Carlisle, 2016	-0.128	-0.861	0.604	0.731
	Berninger, 2002	0.037	-0.737	0.810	0.926
	Jung, 2017	0.212	-0.590	1.013	0.605
	Berninger, 2000b	0.344	-0.217	0.904	0.230
	Roberts, 2006	0.376	-0.124	0.876	0.140
	Wanzek, 2017	0.404	-0.215	1.024	0.201
	Tracy, 2010	0.443	0.091	0.794	0.014
	Rosenthal, 2006	0.562	-0.219	1.343	0.159
	Sinclair, 2005	0.593	-0.063	1.248	0.076
	Lane, 2011	0.607	0.011	1.203	0.046
	Graham, 2005	0.660	0.075	1.245	0.027
	Harris, 2006	0.952	0.313	1.590	0.003
Fixed		0.291	0.140	0.443	0.000
Random		0.305	0.090	0.520	0.005

Hedges's g and 95% CI

Figure 3b.

Forest plot of the effect sizes of the writing instruction on students' writing productivity outcome ($N = 14$, $k = 45$).

Main Effect for Writing Fluency

Model	Study name	Hedges's g	Lower limit	Upper limit	p-Value	Hedges's g and 95% CI
				Statistics for each study		
	Shorter, 2001	-1.166	-1.620	-0.711	0.000	
	Jung, 2017	-0.668	-1.577	0.240	0.149	
	Rosenthal, 2006	0.263	-0.506	1.032	0.503	
	Graham, 2002	0.725	0.131	1.319	0.017	
	Graham, 2000	0.838	0.187	1.489	0.012	
	Jones, 2004	0.932	0.125	1.738	0.024	
Fixed		-0.038	-0.300	0.225	0.780	
Random		0.148	-0.667	0.962	0.723	

Figure 3c.

Forest plot of the effect sizes of the writing instruction on students' writing fluency outcome ($N = 6$, $k = 14$).

Main Effect for Other Writing Outcomes

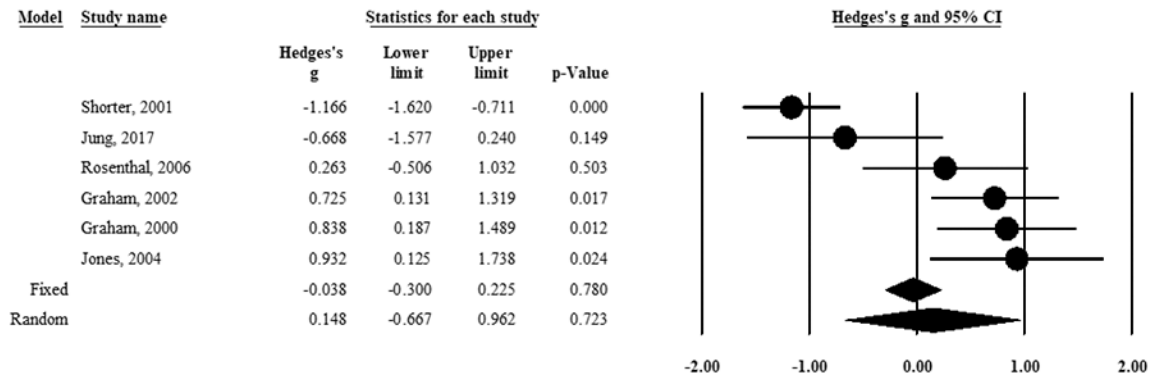| Model	Study name	Hedges's g	Lower limit	Upper limit	p-Value	Hedges's g and 95% CI
	Carlisle, 2016	-0.773	-1.545	-0.001	0.050	
	Kozlow, 2004	-0.066	-0.256	0.125	0.498	
	Puma, 2007	0.074	-0.035	0.183	0.184	
	Shorter, 2001	0.217	-0.204	0.638	0.313	
	Jung, 2017	0.375	-0.437	1.187	0.365	
	Puranik, 2017	0.464	-0.202	1.129	0.172	
	Swain, 2007	0.612	0.333	0.890	0.000	
	Rosenthal, 2006	0.630	-0.156	1.416	0.116	
	Tracy, 2010	0.681	0.323	1.038	0.000	
	Wanzek, 2017	0.769	0.133	1.405	0.018	
	Graham, 2005	0.838	0.238	1.438	0.006	
Fixed		0.163	0.081	0.244	0.000	
Random		0.335	0.106	0.563	0.004	

Figure 3d.
Forest plot of the effect sizes of the writing instruction on students' other composition outcome ($N = 11$, $k = 33$).

Table 1

Studies included in the meta-analysis

Citation	Grade Level	Sample Size	Nature of Design	Composition Outcomes	Instruction Focus	Students with weak writing skills
Berninger et al., 2000	3	93	Quasi-experimental	Productivity	Transcription	Yes
Berninger et al., 2000	3	48	Quasi-experimental	Productivity	Transcription	Yes
Berninger et al., 2002	3	96	RCT	Productivity; quality	Transcription	Yes
Carlisle, 2016	2 and 3	24	RCT	Quality; productivity; other	Transcription	Yes
Graham, Harris, & Chorzempa, 2002	2	60	RCT	Fluency; productivity; quality	Transcription	Yes
Graham, Harris, & Fink, 2000	1	38	RCT	Fluency; quality	Transcription	Yes
Graham, Harris, & Mason, 2005	3	72	RCT	Quality; productivity; other	SRSD	Yes
Harris, 2015	2	53	RCT	Quality	SRSD	Yes
Harris, Graham, & Mason, 2006	2	66	RCT	Quality, productivity	SRSD	Yes
Hooper et al., 2013	2	205	RCT	Quality	Transcription	Yes
Jones, 2004	1	381	Quasi-experimental	Fluency	Transcription	No
Jung, McMaster, & delMas, 2017	1, 2, and 3	48	RCT	Fluency; productivity; other	Transcription	Yes
Kozlow & Bellamy, 2004	3 and older	424	RCT	Quality; other	Other	No
Lane et al., 2011	2	49	RCT	Quality, productivity	SRSD	Yes
Paquette, 2008	2 and older	85	Quasi-experimental	Quality	Other	No
Puma, Tarkow, & Puma, 2007	3 and older	2405	RCT	Quality; other	Other	No
Puranik, Patchan, Lemons, & Al Otaiba, 2017	kindergarten	86	Quasi-experimental	Other	Transcription	No
Roberts & Meiring, 2006	1	39	RCT	Productivity	Transcription	No
Rosenthal, 2006	3	45	RCT	Productivity; fluency; other	Other	No
Shorter, 2001	3	91	Quasi-experimental	Fluency; quality; other	Transcription	No
Sinclair, 2005	3	36	Quasi-experimental	Quality, productivity	Other	No
Swain, Graves, & Morse, 2007	3 and older	941	Quasi-experimental	Quality	Other	No
Tracy, Reid, & Graham, 2009	3	126	RCT	Quality; productivity; other	SRSD	No
Wanzek, Gatlin, Al Otaiba, & Kim, 2017	1	78	RCT	Productivity; other	Transcription	Yes

RCT = Randomized Controlled Trial

Table 2

Effect sizes for writing quality, writing productivity, writing fluency, and other by instructional focus (SRSD, transcription, and mixed)

Variable	Writing Quality		Writing Productivity		Writing Fluency		Writing: Other	
	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value
SRSD	1.04(.28)	3.77/.00	.59(.13)	4.64/.00	NA	NA	.72(.16)	4.61/.00
Transcription	−.19(.25)	−.79/.43	.05(.13)	.42/.68	.13(.49)	.26/.80	.24(.23)	1.02/.31
Mixed	.23(.13)	1.75/.08	.58(.26)	2.26/.02	.26(.39)	.67/.50	.23(.15)	1.55/.12

SRSD = Self-regulated strategy instruction; Mixed = Instructional content varies

Table 3

Effect sizes in writing quality, writing productivity, writing fluency, and other for students with weak versus typical writing skills (top panel), and moderation by instructional focus and writing skill status (bottom panel)

Variable	Writing Quality		Writing Productivity		Writing Fluency		Writing: Other	
	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value	Parameter (s.e.)	Z value/P value
Intercept	.08(.22)	.37/.71	.47(.18)	2.63/.01	−.07(.45)	−.14/.89	.32(.14)	2.38/.02
Weak writing skill	.47(.31)	1.52/.13	−.25(.22)	−1.14/.25	.42(.64)	.65/.52	.04(.27)	.16/.87
Intercept	−.98(.23)	−4.29/.00	.38(.26)	1.48/.14	NA	NA	.29(.18)	1.58/.11
SRSD	1.32(.29)	4.57/.00	.07(.31)	.21/.83	NA	NA	.39(.26)	1.53/.13
Weak writing skill	.92(.27)	3.43/.00	−.41(.28)	−1.47/.14	NA	NA	−.08(.28)	−.29/.77
SRSD*Weak writing skill	−.06(.36)	−.17/.86	.70(.38)	1.85/.06	NA	NA	.24(.45)	.52/.60

SRSD = Self-regulated strategy instruction; Mixed = Instructional content varies