Reading for Writing: A Meta-Analysis of the Impact of Reading Interventions on Writing

Steve Graham
Learning Sciences Institute Australia at the Australian Catholic University in Brisbane
Arizona State University

Xinghua Liu
Shanghai Jiao Tong University

Brendan Bartlett
Australian Catholic University-Brisbane

Clarence Ng
Learning Sciences Institute Australia at the Australian Catholic University in Brisbane

Karen R. Harris
Learning Sciences Institute Australia at the Australian Catholic University in Brisbane
Arizona State University

Angelique Aitken, Ashley Barkel, and Colin Kavanaugh
Arizona State University

Joy Talukdar
Australian Catholic University-Brisbane

This meta-analysis examined if students’ writing performance is improved by reading interventions in studies (k = 54 experiments; 5,018 students) where students were taught how to read and studies (k = 36 investigations; 3,060 students) where students’ interaction with words or text was increased through reading or observing others read. Studies included in this review involved true- or quasi-experiments (with pretests) written in English that tested the impact of a reading intervention on the writing performance of students in preschool to Grade 12. Studies were not included if the control condition was a writing intervention, treatment students received writing
instruction as part of the reading intervention (unless control students received equivalent writing instruction), control students received a reading intervention (unless treatment students received more reading instruction than controls), study attrition exceeded 20%, less than 10 students were included in any experimental condition, and students attended a special school for students with disabilities. As predicted, teaching reading strengthened writing, resulting in statistically significant effects for an overall measure of writing (effect size [ES] = 0.57) and specific measures of writing quality (ES = 0.63), words written (ES = 0.37), or spelling (ES = 0.56). The impact of teaching reading on writing was maintained over time (ES = 0.37). Having students read text or observe others interact with text also enhanced writing performance, producing a statistically significant impact on an overall measure of writing (ES = 0.35) and specific measures of writing quality (ES = 0.44) or spelling (ES = 0.28). These findings provide support that reading interventions can enhance students’ writing performance.

**Keywords:** writing, reading, instruction, meta-analysis

Writing is one of the most essential skills that students must master if they are to be successful at school, work, and in everyday life. In school, writing is commonly used to promote and sharpen students’ learning (Bangert-Drowns, Hurley, & Wilkinson, 2004; Graham & Hebert, 2011b). At work, writing is an integral part of most white-collar and blue-collar jobs (Greene, 2000; National Commission on Writing in America’s Schools and Colleges, 2004, 2005), with forecasters predicting that this skill will become even more central to future job success (Kirsch, Braun, Yamamoto, & Sum, 2007). Socially, writing is part of the basic fabric of 21st century life, as texting, blogging, tweeting, emailing, Facebook posting and commenting, and other forms of creating text permeate everyday life.

Despite the importance of writing, many youngsters do not become skilled writers by the end of high school. The 2011 National Assessment of Educational Progress found that just 27% of 12th-grade students performed at or above the “proficient” level in writing (defined as solid academic performance), with similar results reported on all National Assessment of Educational Progress writing assessments conducted during the past 20 years (http://www.nationsreportcard.gov).

Concerns about students’ writing skills and observations that students spend little time writing or being taught how to write at school (e.g., Applebee & Langer, 2011; Brindle, Harris, Graham, & Hebert, 2016; Troia & Graham, 2016) have played a role in making writing a more central part of educational reform efforts. This was reflected in the 2010 Common Core State Standard (www.corestandards.org), where greater emphasis was placed on specifying what writing skills and processes students need to master at each grade level to prepare them to be successful in college, career, and life. It was also evident in analyses conducted by the Carnegie Corporation and independent scholars to identify effective instructional practices for teaching writing in schools (e.g., Graham & Perin, 2007; Graham, Harris, & Chambers, 2016; Graham, Harris, & Hebert, 2011; Graham, Harris, & Santangelo, 2015; Koster, Tribushinina, de Jong, & van den Bergh, 2015).

While efforts to identify effective instructional practices in writing have been fruitful (over 15 evidence-based practices have been identified to date; see Graham
et al., 2016), they are not complete. They have focused almost exclusively on directly teaching writing, but have not considered less direct methods such as the possible role of reading interventions as a means for improving writing. The meta-analysis presented here addresses this issue by examining if reading interventions enhance students’ writing performance. A previous meta-analysis (Graham & Hebert, 2011a, 2011b) demonstrated that teaching writing and increasing how much students wrote improved students’ reading performance. We think it is likely that parallel reading interventions (i.e., teaching reading and student reading) will enhance writing performance.

Why Should Reading Interventions Enhance Students’ Writing Performance?

Writers Gain Knowledge That Leads to Better Writing When Reading Is Taught

Reading and writing are quite similar, even if they are not identical. As Tierney and Shanahan (1991) noted, students draw on similar knowledge representations and cognitive processes when they read and write. A useful metaphor for illustrating this relationship is that reading and writing are “two buckets drawing water from a common well or two buildings built on a common foundation” (Shanahan, 2016, p. 195). Thus, specific reading interventions designed to enhance common sources of knowledge that writers draws on should lead to better writing performance.

Fitzgerald and Shanahan (2000) identified common knowledge sources that reading and writing share. One knowledge base includes domain knowledge. Readers draw on domain knowledge to understand what they are reading, while writers draw on this same source for ideas as they write. Another knowledge source involves meta-knowledge about written language. Readers and writers draw on what they know about the functions and purposes of written language as well as their knowledge about how writers and readers interact. This helps them interpret an author’s message and construct their own message for others to read. A third knowledge source is procedural knowledge. This includes knowledge about how to access information purposefully, set goals, question, predict, summarize, visualize, and analyze. Readers apply this type of knowledge to help them make sense of what is read, while writers use such knowledge to direct and craft what is written. The final knowledge source involves pragmatic knowledge of text attributes. This includes knowledge of the features of text, words, syntax, and usage. Writers and readers use such knowledge to decode/encode words and comprehend/construct sentences or larger units of text.

Consistent with the viewpoint that specific reading interventions lead to better writing, reading instruction designed to increase students’ knowledge about the functions and purposes of text should result in better writing, as it provides students with information they can use to guide their own writing and thinking about the expectations of the reader (Meyer et al., 2002). Moreover, teaching students comprehension strategies for reading and how the text they are reading is structured provides students with knowledge and strategies (e.g., goal setting, summarizing, monitoring, visualizing, and analyzing) that are essential to writing (Hayes, 1996).

Likewise, teaching phonological awareness, phonics skills, vocabulary, and reading fluency should enhance students’ spelling skills (Graham, 2000). Phonological awareness instruction is designed to help students learn that words are made of individual sounds—a skill that is essential when one tries to spell a
word part by part. Phonics instruction teaches students about the connections between letters and sounds—knowledge that can be used when trying to connect sound to letter(s), as is done when spelling. Vocabulary and reading fluency instruction typically involve repeated exposure to reading specific words. This instruction should make the spelling of these words more memorable. Making students better spellers through reading instruction may have added benefits, as it can reduce the interfering effect that spelling can have on other writing processes (McCutchen, 1995). For instance, when students have to stop and think about how to spell a word when writing, they may forget ideas or plans they are holding in working memory.

Writers Gain New Insights Into Writing by Reading and Observing Readers' Reactions

Reading and writing can be conceptualized in terms of writer–reader relations (Rubin, 1984). The purpose of both reading and writing is to communicate, and effective communication when reading or writing involves specific processes that inform each other (Shanahan, 2016). For example, an awareness of the author’s position when reading is important for making interpretation, as is a consideration of the reader’s perspective when writers produce text. As a result, specific reading interventions that provide students with the opportunity to gain new insights and awareness about writing or their audiences should lead to better writing.

To illustrate, readers may acquire important insights into writing, as they think about why an author used a particular word, phrase, sentence, or rhetorical device to deliver an intended meaning (Nelson & Calfee, 1998; Tierney & Shanahan, 1989). As students read words or text, they may evidence improved spelling performance as they are exposed to correct spelling and recode read words from print to sound (often referred to as self-teaching; see Share, 1999). Reading and analyzing what others write (MacArthur, 2016) as well as observing readers as they discuss material read or try to execute written task directions (Couzijn, 1999) should enhance writing performance, as these practices may make students more sensitive and knowledgeable about the elements of effective writing and the needs of their audience (MacArthur, 2016).

Purpose of the Current Meta-Analysis

This meta-analysis was designed to answer the following research questions for preschool to grade 12 students:

**Research Question 1:** Does teaching reading enhance writing performance?

**Research Question 2:** Does increasing students’ interaction with words or text through reading or observing others read enhance writing performance?

Affirmative answers to these two questions would provide support for the proposition that reading interventions improve students’ writing, providing additional evidence on how to teach writing effectively. Such findings would also provide partial support for the theoretical proposition that there is a bidirectional relationship between reading and writing (Shanahan, 2016), with reading
influencing writing and vice versa. Previous meta-analyses demonstrated that writing interventions enhanced reading performance (Graham & Hebert, 2011a, 2011b) and that combined reading/writing interventions improved both reading and writing (Graham et al., in press), but no analyses has examined the impact of reading interventions on writing.

We anticipated an affirmative answer to both of our research questions. These predictions are based on the logic underlying the two theoretical arguments presented earlier: (a) students draw on similar knowledge representations and cognitive processes when they read and write and (2) reading and writing involves specific processes that inform each other (Fitzgerald & Shanahan, 2000; Tierney & Shanahan, 1991). Accordingly interventions that enhance the shared knowledge bases readers and writers draw on or engage students in the act of reading or writing should have a positive impact on students’ literacy skills. Previous reviews demonstrated this was the case for writing interventions, as they improved students’ reading performance. As noted earlier, Graham and Hebert’s (2011a, 2011b) meta-analysis of 95 true- and quasi-experiments with students in Grades 1 to 12 found that writing and writing instruction produced statistically significant gains in word reading, fluency, and comprehension. Even more specifically, Graham and Santangelo (2014) reported that spelling instruction enhanced students’ reading skills (i.e., words reading and reading comprehension) in 20 true- and quasi-experiments conducted with students in kindergarten to Grade 12. While the impact of writing and writing instruction on reading does not ensure a reciprocal effect from reading to writing, it supports the possibility of such a relationship as predicted in this meta-analysis.

Correlational studies examining relationships between reading and writing provide additional support for the proposition that reading interventions can improve students’ writing performance (Nelson, 2008; Pearson, Brenner, & Long, 2012). If this is the case, then measures of reading and writing should be statistically related, and models testing reading and writing relations should show that the path from reading to writing is statistically significant. Reviews of selected literature by Shanahan and colleagues (Fitzgerald & Shanahan, 2000; Shanahan, 1984, 2006, 2016; Tierney & Shanahan, 1991) provide support for the first contention, showing that reading and writing measures share considerable variance, and this shared variance can be as high as 85% for word-level factors and 65% for text-level factors when multiple measures are used to assess each reading and writing construct (e.g., Berninger, Abbott, Abbott, Graham, & Richards, 2002). In terms of model testing, most studies demonstrate that bidirectional models of reading and writing relations provide a good fit to the obtained data (e.g., Berninger & Abbott, 2010; Shanahan & Lomaz, 1986), which is not inconsistent with this review’s proposition that reading influences writing. In one study, though a reading to writing model provided an even better fit to the data than a bidirectional model of reading and writing relations for word and text levels of language (Ahmed, Wagner, & Lopez, 2014).

While supportive, evidence demonstrating that writing and writing instruction improve reading performance as well as correlational evidence supporting reading and writing relations is not sufficient to establish that reading interventions improve students’ writing performance. A more powerful approach for demonstrating such
an effect is to directly test this proposition by systematically and comprehensively summarizing the impact of reading interventions on writing performance. This was the approach applied in this review. We conducted a meta-analysis that included experimental treatment studies to determine if reading instruction and increased interaction with reading strengthened students’ writing performance. Meta-analysis is particularly well suited to this task, as it provides an estimate of the effectiveness of a treatment “under conditions that typify studies in the literature” (Bangert-Drowns et al., 2004, p. 34).

To our knowledge, such an analysis has not been undertaken previously. This is not to say that prior reviews have not examined the outcomes of reading interventions on writing performance. They have, and they provide some support for the questions addressed in this meta-analysis. For example, Grabe and Zhang (2016) compiled a personally selected annotated bibliography that included studies examining reading interventions impact on writing, and drew general conclusions from the studies presented that reading and reading instruction enhanced writing performance. Similarly, a panel of scholar convened by the International Reading Association and the National Institute of Child and Human development (Pearson et al., 2012) presented a position statement that drew on selected literature, and they concluded that reading interventions improved students’ writing. Likewise, selective qualitative reviews of the literature by Nelson (2008), Shanahan (2016), and Stotsky (1983) made similar claims, and an even more selective qualitative review of the literature by Graham (2000) claimed that reading and reading instruction resulted in better spelling by students.

Each of these prior reviews evidenced three or more of the following limitations, as they failed to (a) comprehensively review the relevant literature, (b) focus on multiple writing outcomes, (c) examine the quality of obtained studies, (d) provide an empirical estimate of the obtained effects of reading interventions on writing performance, or (e) examine factors contributing to variability of obtained effects.

Our meta-analysis addressed all five of these limitations. We conducted a comprehensive search of the published and unpublished literature, examined the impact of reading interventions on multiple writing outcomes, systematically assessed the quality of obtained studies, computed average weighted effect sizes (ESs) for the reading interventions tested, and determined if excess variability in these effects was related to specific study factors.

In terms of variability in effects, Cheung and Slavin (2016) stress the importance of examining if methodological features of experimental studies are correlated with study outcomes. If they are correlated, it may indicate the presence of bias. Not identifying such bias may mislead researchers, educators, and policy makers about whether reading interventions strengthen writing. Such biases can also weaken the value of the evidence from this meta-analysis to support or not support theoretical positions on reading and writing relations.

To examine relations between ESs and methodological features of studies, we addressed five study features recommended by Cheung and Slavin (2016) and four of our own devising. The five recommended by Cheung and Slavin included (a) researcher made versus norm-referenced measures (the former may produce larger effects, as they are likely over aligned with the treatment); (b) smaller
versus larger sample studies (the former may yield larger effects, as small-scale studies are easier to control); (c) stronger versus weaker designs (e.g., quasi-experiments may produce larger effects than true-experiments, as selective factors may work in favor of the former); (d) elementary versus secondary students (the former may generate larger effects, as younger students tend to gain more during a year than older youngsters; Bloom, Hill, Black, & Lipsey, 2008); and (e) published versus unpublished studies (the former may yield larger effects, as reviewers are more likely to recommend publication for studies with larger effects).

The four additional study features examined were the following: (a) typically developing students versus students experiencing difficulty with literacy (the former is likely to produce larger effects, as typically developing students tend to gain more during a year than students experiencing school difficulties; Swanson, Harris, & Graham, 2013); (b) more instruction versus less instruction (the former is likely to produce larger effects, as students are provided more opportunity to learn); (c) researcher-delivered versus teacher-delivered instruction (the former is likely to produce larger effects, as researchers are more likely to deliver the treatment as intended; De Boer, Donker, & van der Werf, 2014); and (d) studies of higher quality versus ones of lower quality (studies that meet more indicators of study quality are likely to produce larger effects, as such studies are more tightly controlled).

The examination of the relations between these study features and the magnitude of the effects obtained in the studies reviewed not only provided a measure of whether our overall outcomes were biased but it also allowed us to identify specific methodological issues that exist in studies examining the impact of reading interventions on writing. This is important in terms of deciding how much confidence can be placed in the basic findings from this review and for making recommendations for future research.

Method

Search for Relevant Studies

Relevant studies were identified through six search strategies (all search strategies ended January 1, 2016, but extended backward as far as possible). First, electronic searches of the following data bases were conducted: ERIC, PsychINFO, ProQuest Dissertations & Theses Global, Linguistics and Language Behavior Abstracts, and EBSCOhost. These searches involved the following terms: (reading AND writing) AND ((research) OR (“control group” OR “comparison group” OR “quasi-experiment” OR “true experiment” OR randomized control design”)) NOT (“qualitative study” OR “case study” OR “action research” OR “single subject design” OR “descriptive study” OR “ethnography” OR “university” OR “higher education”).

Following the electronic searches, hand searches of 16 journals were conducted. We searched prominent reading (Annals of Dyslexia, Journal of Literacy Research, Reading Psychology, Reading Research Quarterly, and Scientific Studies of Reading) and writing journals (Reading and Writing, Research in the Teaching of English, Scientific Studies of Reading, and Written Communication) that publish quantitative studies, as we reasoned that these journals were likely to contain articles examining the impact of reading interventions on measures of

References from 42 previous reviews of the reading intervention literature were further examined, including the report of the National Reading Panel (National Institute of Child Health and Human Development, 2000). Nineteen prominent reading researchers were asked to identify relevant studies. A google search of studies referencing the investigations included in this review was also conducted to identify other possible studies. Finally, the reference lists of all obtained documents were searched to identify other possible studies.

Eligibility Criteria

Inclusion Criteria

Studies had to meet the following criteria to be included in this review: (a) involved students in preschool to 12th grade; (b) contained a reading intervention group; (c) applied a true- or quasi-experiment to test the intervention; (d) included at least one measure assessing writing that evaluated the impact of the reading intervention (quasi-experiments had to include a comparable pretest writing measure since students were not randomly assigned to conditions); (e) written in English; and (f) contained the data/statistics necessary to compute a weighted ES (or data/statistics were obtainable from the authors). The only exception to these inclusion criteria involved studies testing the self-teaching treatment where students read words or text to determine if spelling improved (Share, 1999). This area of research commonly employs a group comparison design where students serve as their own control (e.g., Kyte & Johnson, 2006). As a result, participants as own control designs along with true- and quasi-experiments were included for this reading treatment.

Exclusion Criteria

Studies were excluded for seven reasons. One, studies that were conducted in schools exclusively for students with disabilities (e.g., school for students with visual impairments, deafness, or behavioral/emotional difficulties) were not included, as the purpose of this review was to draw recommendations for students in regular school settings. Two, investigations were excluded if attrition exceeded 20% for the reading intervention condition. Attrition greater than 20% can bias findings (Dumville, Torgerson, & Hewitt, 2006). Three, investigations with fewer than 10 participants in any condition were excluded. Four, studies were excluded if the only writing outcome measure(s) involved filling in blanks on a graphic organizer or a cloze test, copying text if the purpose did not involve assessing handwriting legibility or fluency, or writing one’s name. While we included studies where students wrote individual words to assess skills like spelling, cloze tests and graphic organizers do not assess writing skills (at least this was the case in the studies examined in this review). Name writing was excluded as an outcome measure, as this skill is often taught by parents outside of the experimental situation.
Five, studies were excluded if the control condition received a reading treatment. The goal of this meta-analysis was to isolate the effects of reading or reading instruction as much as possible, not to compare one reading intervention to another. There was one exception to this rule. If the treatment and control conditions received the same amount of reading or reading instruction as part of their typical language arts program, but the experimental manipulation involved the treatment group receiving additional reading or reading instruction, the study was included.

Six, investigations comparing a reading intervention to a writing intervention were excluded. The only exception to this involved studies where both treatment and control received equivalent writing instruction, but the treatment condition also received a reading intervention. The purpose of this meta-analysis was not to compare the effectiveness of reading versus writing interventions, but to determine the impact of reading and reading instruction on writing performance.

Seven, studies that involved writing interventions as part of the reading treatment were excluded. The inclusion of such studies would draw into question the impact of reading or reading instruction on writing performance, as effects may have been due to writing instruction. For instance, we did not include studies assessing spelling that tested the effectiveness of a phonics reading program that conjointly taught spelling (e.g., Vadasy, Jenkins, & Pool, 2000) or studies that tested the effectiveness of a broader combined reading/writing program. This last exclusion criterion also applied to studies where students were explicitly taught the writing outcome as part of reading or the reading instruction treatment (e.g., Wijekumar et al., 2014).

There were two exceptions to this seventh exclusion criteria that involved eight studies. One, we included six studies where students in the treatment condition received a reading treatment and some writing instruction as part of the intervention, but students in the control condition received equivalent writing instruction but no reading instruction (Boyer & Ehri, 2011; DiCecco & Gleason, 2002; Fuchs et al., 2009; Levine, 2014; Sussman, 1998; Troyer, 1993). With these studies, we were still able to isolate the impact of the reading intervention on writing. Two, we included one study (Carro, 1999), where a relatively small amount of the phonological awareness intervention involved using letters in an Elkonian box to separate the sounds in words. This could be viewed as a spelling intervention (the control condition did not receive an equivalent amount of such instruction). However, for the writing outcome measure, we computed an ES was number of words students produced when writing a composition. A meta-analysis by Graham and Santangelo (2014) demonstrated that spelling instruction does not enhance students’ writing output.

Selection of Studies for Review

After removing duplicate listings, the six search strategies yielded 17,391 results. The electronic searches were conducted first and yielded 16,429 results. Subsequent journal searches produced 317 results, examination of references from previous reviews 123 results, queries to prominent researchers 13 results, google searches 48 results, and examination of references from obtained papers 462 results.
The title and abstract of the 17,301 obtained entries were read by the first and last authors of this review to determine if it met five of the six inclusion criteria specified earlier. Specifically, did the title and abstract provide evidence that the entry was an empirical intervention study written in English with students in preschool to Grade 12 involving a reading treatment as well as at least one writing assessment. We did not consider if the entry included relevant statistics for computing an ES, as this could not be determined from the title and abstract. Furthermore, we were less stringent about the need for evidence that clearly indicated a writing measure was administered, as we were concerned that relevant studies might be eliminated. Interrater agreement for this initial screening was 98%, with the first and last authors differing on 433 entries. Disagreements were resolved by the first author, who had 40 years of experience conducting literacy research. Two hundred and eleven of these disagreements (51%) centered on the inclusion of a writing assessment.

The initial screening resulted in the identification of 677 relevant documents that were then obtained (584 from the electronic searches and 83 from the other five search methods). Two trained graduate students read each document in full to determine if it met all inclusion and exclusion criteria (agreement was 92%). Disagreements were resolved by the first author. Only 89 of the 677 documents met all criteria. The most common reason for eliminating the 588 full documents that were reviewed was studies did not meet the criteria for design ($k = 127$). This was followed by: absence of a writing measure ($k = 121$), failed to meet the criteria for a reading intervention ($k = 101$), writing was included as part of the reading treatment ($k = 74$), compared two reading treatments ($k = 55$), quasi-experiments without a pretest ($k = 29$), compared a reading treatment to a writing treatment ($k = 26$), attrition greater than 20% ($k = 17$), involved participants who were not in preschool to 12th grade ($k = 15$), absence of information needed to compute an ES ($k = 12$), did not include at least 10 students in each condition ($k = 8$), and was conducted in a special school setting ($k = 3$).

It is important to note that some of the 91 documents contained multiple studies (i.e., Rosenthal & Ehri, 2008; Share, 2004; Vaughn et al., 2006), whereas other documents contained maintenance data for another experiment in this review (i.e., Byrne & Fielding-Barnsley, 1993, 1995; Byrne, Fielding-Barnsley, & Ashley, 2000; Cirino et al., 2009). Other studies included a control and two different reading treatments (Bhattacharya & Ehri, 2004; Couzijn & Rijlaarsdam, 2004; Crowhurst, 1990, Korat, Levin, Ben-Shabt, Shenor, & Bokovza, 2014; Lumbelli & Paoletti, 2004; Moore & MacArthur, 2012), increasing the number of comparisons that were possible.

Coding

Each study was coded for the following descriptors: (a) publication type (journal article, dissertation/thesis, conference presentation, book chapter, and other), (b) grade level (or age level if grade not available), (c) type of student (full range of a classroom or students experiencing difficulty with literacy or learning), (d) number of participants, (e) length of intervention (i.e., number of sessions), (f) gender, (g) race, (h) socioeconomic status (SES), (i) location (urban, suburban, rural), (j) country, (k) language, (l) who delivered the treatment (teachers or
researchers), (m) professional development (provided or not provided), and (n) type of assessment (norm-referenced, researcher designed). In addition, a brief written description of the intervention and control condition(s) was developed and the necessary statistics were recorded.

Each study received a score for nine quality indicators. A score of 1.0 was assigned if the condition in the parentheses following each indicator below were met. Otherwise a score of 0 was assigned. The indicators were (a) high-quality design (true experiment); (b) not an N of 1 design (more than two groups or classes in each condition); (c) teacher effects controlled (teachers randomly assigned to condition or taught in each); (d) attrition (not greater than 10%); (e) differential attrition (differences in attrition between each conditions was 5% or less); (f) pretest equivalence (treatment and control conditions demonstrated equivalence on pretest writing measure as mean scores for each condition did not differ by more than the smallest standard deviation for the two conditions at pretest; true-experiments that did not include a pretest were credited with meeting the criterion); (g) no floor or ceiling effects at pretest or posttest (mean score for a writing measure was not more than one standard deviation from the lowest and highest scale for the measure); (h) measures were reliable (a reliability coefficient of .70 or greater was reported for the writing measure; norm-referenced measures were assumed to be reliable); and (i) treatment fidelity (evidence of treatment fidelity provided).

For studies with multiple writing measures, the quality scores for pretest equivalence, floor/ceiling effects, and reliability were adjusted for the number of measures that met the criterion (e.g., the score was .50 when two out of four measures were reliable). For each study, a summative score for proportion of quality indicators met was calculated. This was done by summing the quality indicator points in a study, dividing by nine.

Each study was coded independently by two trained graduate assistants. The two graduate students resolved their differences. Any initial disagreement between the two trained coders was examined by the first author and resolved. Across all coded items, reliability was 94% (it was never below 90% for any single item).

Categorizing Studies

The first author read each study that met inclusion/exclusion criteria to determine which question the investigation addressed and to develop a brief written summary of the reading treatment tested. Based on these initial descriptions, studies were then reread by the first author, and the treatment categories and descriptions for each question were refined. A third reading was undertaken to finalize the placement of each study. The goal of this process was to create groupings of studies that isolated specific reading treatments relevant to each research question. After this third categorization, the second author independently categorized all studies using the descriptors (a brief written summary of each treatment) developed and refined by the first author. Reliability between the first and second author was 97%; disagreements were resolved through discussion.

The categorization process resulted in the following treatment categories for Research Question 1: (a) phonological awareness instruction, (b) phonics instruction, (c) vocabulary instruction, (d) reading fluency instruction, (e)
comprehension instruction, and (f) multicomponent reading instruction (this treatment involved reading instruction at both the word and comprehension level). For Research Question 2, the following treatment categories were identified: (a) self-teaching where students read words or text and effects on spelling was assessed (Share, 1999); (b) more reading where the amount of text students read was increased (this did not include studies that only assessed spelling, as it would overlap with self-teaching); (c) observing readers as they engage in reading or use text to carry out an activity; (d) reading and analyzing another person’s text; and (e) reading model text to emulate it when writing. Two studies did not specifically address any of the three questions posed. This included a study by Ross (2004) that assessed the effectiveness of a treatment involving formative assessment, where teachers assessed students’ reading performance using running records. Teachers used these assessments to plan their reading instruction. It also included an investigation by Daki and Savage (2010) testing the impact of a brief mental health program focused on students’ reading difficulties.

A study could be placed in more than one treatment category, as some studies included more than one tested intervention or multiple studies. For instance, Moore and MacArthur (2012) tested the effectiveness of two reading interventions, observing readers and reading and analyzing text.

Analysis

Basic Procedures

Effect sizes were computed by subtracting the posttest mean for the control group from the posttest mean for the treatment group and dividing by the pooled standard deviation of the two groups at posttest. Pretest differences between the treatment and control condition were accounted for by subtracting the mean pretest score of each group from the respective mean posttest score (for true-experiments without pretest scores no adjustment were made). The same procedures were used to calculate an ES for maintenance scores, again adjusting for possible pretest differences. For participants as own control studies (e.g., Kyte & Johnson, 2006), students’ mean control performance was subtracted from their mean treatment performance and divided by the pooled standard deviation.

Prior to calculating some ESs, a procedure recommended by Neurri and Greenberg (Cortina & Nouri, 2000) was used to average the performance of two or more groups in each condition when necessary (e.g., statistics reported separately by grade; Lie, 1991). It was also necessary in some cases to estimate missing standard deviations from the statistics reported by the study authors (e.g., Whitehurst et al., 1994). All quasi-experiments included in this meta-analysis assigned schools, classes, or groups of students to conditions (e.g., Carro, 1999), but examined student-level effects. As a result, we applied procedures recommended by Hedges (2007) to adjust for clustering effects. This involved the use of imputed intraclass correlation estimates applied in previous meta-analysis of writing outcomes (see Graham, Kiuhara, McKeown, & Harris, 2012; Graham & Santangelo, 2014). All ESs were adjusted for small sample size bias (i.e., Hedges' g; Hedges, 1982).
**Average Weighted Effect Size**

Statistical analyses were conducted with Comprehensive Meta-Analysis (Version 3). A weighted random-effects model (weighted to take into account sample size by multiplying each ES by its inverse variance) was employed to answer the research questions. For all average weighted ESs, we provided a corresponding confidence interval (CI) and level of statistical significance. In addition, two measures of variability were computed: $Q$ and $F$. The $Q$ statistic determines if variability in an average weighted ES exceeds sampling error alone (Shadish & Haddock, 2009). $F$ was calculated as a second complimentary measure of homogeneity, as it is less sensitive to sample size than $Q$. The $I^2$ statistic was interpreted using Higgins and Green’s (2011) guidelines: 0% to 40% = might not be important; 30% to 60% may represent moderate heterogeneity; 50% to 90% may represent substantial heterogeneity; and 75% to 100% considerable heterogeneity. To avoid inflating sample size, only one ES per study was used to compute an average weighted ES (Gleser & Olkin, 2009).

A shifting unit of analysis approach was applied (Cooper, 1998). With this approach, the effects associated with a specific writing measure (e.g., writing quality, spelling, words written) were coded as if they were independent estimates of the reading treatments effects for each question. For instance, if a reading treatment included a measure for writing quality, spelling, and words written, three different ESs were calculated. In addition, an overall writing ES was computed by averaging these three effects (if a study included just one writing measure, it served as the overall writing score). For each question, an average weighted ES was computed for overall writing as well as for each separate writing measure. In addition, an average weighted ES was computed for each specific treatment pertinent to each research question. In each of these analyses, each study contributed a single effect to the analysis. This shifting unit of analysis approach retains as many data points as possible for each investigation, while not violating the assumptions that effects are independent in each analysis.

It should also be noted that in studies that included multiple measures for a single writing construct (e.g., writing quality), the ESs were aggregated to provide a single estimate of the effect. For instance, if a study contained two spelling measures, the ESs for both measures were computed and averaged.

**Outliers**

Before computing average weighted ESs for our research questions, we examined if any single ES was exerting undue influence in terms of magnitude of effect or number of participants. Using Tukey’s (1977) definition of an outlier score as falling three times the interquartile range above the 75th percentile or below the 25th percentile of the distribution of scores, we identified four outliers that required adjustment for Research Question 1. Three studies had an ES greater than the obtained Tukey outlier of 2.06 (i.e., Armand, Lefrançois, Baron, Gomez, & Nucke, 2004; Martins & Silva, 2006; Rosenthal & Ehri, 2008). One study (Kegel & Bus, 2012) had 27 more students than the obtained Tukey outlier of 270.
participants. The ESs for the first three studies were adjusted accordingly, as was the standard error for the fourth study.

**Moderator Analyses**

When homogeneity in ESs for writing overall scores exceeded sampling error alone (as represented by a statistically significant $Q$ statistic), moderator analyses were conducted to determine if the excess variability could be accounted for by identifiable differences between studies. Study characteristics examined in these analyses included: sample size, type of experiment (true-experiment vs. quasi-experiment plus participants as own control), grade level (preschool vs. elementary vs. secondary), publication status (published vs. unpublished studies), type of assessment (norm-referenced vs. researcher designed), instructor (teacher vs. research team), overall quality of the study (based on the quality indicators described earlier), type of students (typical students vs. those experiencing challenges learning to read/write), number of instructional sessions. Moderator analyses tested the individual and combined effects of these study characteristics.

**Publication Bias**

We applied two methods to determine if publication bias existed in the studies used to answer each question in this meta-analysis. Examination of publication bias was limited to writing overall at posttest for each question. There were not enough studies available to estimate reliably possible bias at maintenance. These procedures (Borenstein, Hedges, Higgins, & Rothstein, 2009) applied included the following: (a) creating a funnel plot of precision and examining it using Duval and Tweedie’s (2000) Trim and Fill procedure and (b) conducting a Begg and Mazumdar rank correlation test (this procedure tests the interdependence of variance and ESs).

**Results**

The results of our meta-analysis are summarized in four tables. Table 1 presents the average weighted ES at posttest for overall writing score (i.e., all measures of writing averaged in a study) by question and specific reading treatments. For each averaged ES, Table 1 provides the following additional information: statistical significance of the obtained ES (i.e., is it greater than no effect), CI for each ES, and two measures ($Q$ and $I^2$) assessing degree of heterogeneity. Table 2 includes this same data by type of measure (e.g., writing quality, spelling) and research question. Table 3 presents this same type of information at maintenance by research question for writing overall. Table 4 includes information about individual studies. ESs at posttest and maintenance are for writing overall.

**Research Question 1: Does Teaching Reading Improve Students’ Writing Performance?**

Fifty-four experiments involving 5,018 preschool to Grade 12 students were located (see Table 4) that examined whether teaching students how to read enhanced the writing performance of students in preschool to Grade 12 (two studies only assessed maintenance; Byrne & Fielding-Barnsley, 1991, had maintenance papers published in 1993, 1995, and 2000; Ecalle et al., 2009, assessed
Most of the studies were with younger students (60% of studies with Grade 1 to 6 students, and 27% of studies with preschool to kindergarten children). Seventy-nine percent of studies were published articles.

The 54 experiments were published in a variety of countries including the United States (49%), Canada (23%), the United Kingdom (9%), Spain (4%), Netherlands (4%), Australia (2%), France (2%), Hong Kong (2%), Norway (2%), Portugal (2%), and Taiwan (2%). Reading instruction involved English orthography in slightly more than 80% of investigations. Almost one-half of the studies involved students who experienced difficulties mastering literacy. In the studies that provided information on gender (54%), race (20%), and SES (33%), 56% of students were boys, 74% were from minority groups, and 89% of youngsters were from either poor or poor to median income households/communities. Caution needs to be exercised in interpreting race and SES data, as the reporting of such data may have occurred more often when the study was conducted in specific communities.

Most of the experiments focused on teaching phonological awareness (39%) and phonics (28%), with the remaining studies involving comprehension instruction (22%), multicomponent interventions (7%), reading fluency (2%), and vocabulary (2%). Phonological awareness studies only involved students in

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Studies</th>
<th>ES</th>
<th>95% CI</th>
<th>p</th>
<th>Q</th>
<th>F²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1</td>
<td>Overall</td>
<td>52</td>
<td>0.57 [0.43, 0.70]</td>
<td>&lt;.001</td>
<td>180.90***</td>
<td>71.81</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>20</td>
<td>0.69 [0.43, 0.96]</td>
<td>&lt;.001</td>
<td>80.90***</td>
<td>76.52</td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>14</td>
<td>0.39 [0.26, 0.53]</td>
<td>&lt;.001</td>
<td>17.20</td>
<td>24.21</td>
</tr>
<tr>
<td></td>
<td>COMP</td>
<td>12</td>
<td>0.66 [0.33, 0.99]</td>
<td>&lt;.001</td>
<td>31.78**</td>
<td>65.39</td>
</tr>
<tr>
<td></td>
<td>MULTI</td>
<td>4</td>
<td>0.47 [-0.13, 1.07]</td>
<td>.126</td>
<td>24.41***</td>
<td>87.71</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>Overall</td>
<td>36</td>
<td>0.35 [0.23, 0.46]</td>
<td>&lt;.001</td>
<td>67.94**</td>
<td>48.51</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>15</td>
<td>0.28 [0.08, 0.48]</td>
<td>.007</td>
<td>33.08**</td>
<td>57.68</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>9</td>
<td>0.29 [0.14, 0.43]</td>
<td>.001</td>
<td>8.56</td>
<td>18.20</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>8</td>
<td>0.62 [0.34, 0.91]</td>
<td>&lt;.001</td>
<td>12.90***</td>
<td>45.73</td>
</tr>
<tr>
<td></td>
<td>R&amp;A</td>
<td>5</td>
<td>0.43 [0.21, 0.65]</td>
<td>&lt;.001</td>
<td>4.50</td>
<td>11.18</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>3</td>
<td>0.15 [-0.08, 0.51]</td>
<td>.32</td>
<td>1.05</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. ES = effect size; CI = confidence interval; PA = phonological awareness instruction; PH = phonics instruction; COMP = comprehension instruction; MULTI = multicomponent intervention; ST = self-teaching (read words/text to determine impact on spelling); MR = more reading; OR = observing readers; R&A = reading and analyzing text; MODEL = reading text to emulate it. The sum of PA, PH, COMP, and MULTI studies is 50 and does not equal the 52 studies included in Overall for Research Question 1 as Overall includes a vocabulary as well as a reading fluency intervention study.

*p < .05. **p < .01. ***p < .001.
preschool, kindergarten, or the elementary grades. The remaining studies were all conducted with students in Grades 1 to 6, except for four comprehension studies, two phonics studies, and one multicomponent intervention (see Table 4).

The overall quality of studies was generally strong: 74% were true-experiments, over 90% evidenced no issues with attrition, 89% included more than a single group or class in each condition, 69% controlled for teacher effects, and pretest equivalence, reliability, and floor/ceiling problems were not evident for 91%, 74%, and 69% of writing measures, respectively. The most notable weakness in study quality involved treatment fidelity. This was established in just 44% of studies.

**Effects at Posttest**

Ninety-four percent of the 52 studies testing Research Question 1 produced positive effects at posttest (see Table 4). The 52 studies yielded a statistically significant average weighted ES of 0.57 for overall writing (see Table 1). Heterogeneity in the obtained effects of these 52 studies exceeded sampling error alone (see $Q$ statistic Table 1), with 72% of the variance in effects attributable to

<table>
<thead>
<tr>
<th>Measures</th>
<th>Studies</th>
<th>ES</th>
<th>95% CI</th>
<th>$p$</th>
<th>$Q$</th>
<th>$I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>40</td>
<td>0.56</td>
<td>[0.41, 0.71]</td>
<td>&lt;.001</td>
<td>148.31***</td>
<td>73.70</td>
</tr>
<tr>
<td>Writing quality</td>
<td>12</td>
<td>0.63</td>
<td>[0.33, 0.93]</td>
<td>&lt;.001</td>
<td>36.67**</td>
<td>64.13</td>
</tr>
<tr>
<td>Words</td>
<td>5</td>
<td>0.37</td>
<td>[0.07, 0.68]</td>
<td>.017</td>
<td>2.22</td>
<td>00.00</td>
</tr>
<tr>
<td>Research Question 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing quality</td>
<td>21</td>
<td>0.44</td>
<td>[0.31, 0.57]</td>
<td>&lt;.001</td>
<td>28.06</td>
<td>28.71</td>
</tr>
<tr>
<td>Spelling</td>
<td>15</td>
<td>0.28</td>
<td>[0.08, 0.48]</td>
<td>.007</td>
<td>33.08**</td>
<td>57.68</td>
</tr>
</tbody>
</table>

*Note. ES = effect size; CI = confidence interval.

* $p < .05$. ** $p < .01$. *** $p < .001$. 

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Studies</th>
<th>ES</th>
<th>95% CI</th>
<th>$p$</th>
<th>$Q$</th>
<th>$I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>10</td>
<td>0.37</td>
<td>[0.14, 0.58]</td>
<td>&lt;.001</td>
<td>6.15</td>
<td>43.36</td>
</tr>
<tr>
<td>Research Question 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>5</td>
<td>0.15</td>
<td>[−0.04, 0.34]</td>
<td>.13</td>
<td>4.55</td>
<td>00.00</td>
</tr>
</tbody>
</table>

*Note. ES = effect size; CI = confidence interval.*
<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>$N$</th>
<th>Design</th>
<th>Grade</th>
<th>Ss</th>
<th>Number of sessions</th>
<th>Ins</th>
<th>Subject area</th>
<th>Measure</th>
<th>Study quality</th>
<th>ES post</th>
<th>ES follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armand et al. (2004)</td>
<td>PA</td>
<td>202</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td>Ball and Blachman (1991)</td>
<td>PA</td>
<td>59</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>35</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.83</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Boyer and Ehri (2011)</td>
<td>PA</td>
<td>60</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>11</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
<td>1.67</td>
<td>0.31</td>
</tr>
<tr>
<td>Bradley and Bryant (1985)</td>
<td>PA</td>
<td>65</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>396</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Byrne and Fielding-Barmsley (1991, 1993, 1995); Byrne et al. (2000)</td>
<td>PA</td>
<td>103</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>12</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Carro (1999)</td>
<td>PA</td>
<td>23</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Words</td>
<td>0.22</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Castiglioni-Spalten and Ehri (2003)</td>
<td>PA</td>
<td>45</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>6</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Davidson and Jenkins (1994)</td>
<td>PA</td>
<td>40</td>
<td>0</td>
<td>P</td>
<td>R</td>
<td>60</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>Defior and Tudela (1994)</td>
<td>PA</td>
<td>36</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>126</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.63</td>
<td>0.36</td>
<td>−0.02</td>
</tr>
<tr>
<td>Hatcher, Hulme, and Ellis (1994)</td>
<td>PA</td>
<td>61</td>
<td>1</td>
<td>E</td>
<td>R</td>
<td>100</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.83</td>
<td>1.43</td>
<td>−0.11</td>
</tr>
<tr>
<td>Hecht and Close (2002)</td>
<td>PA</td>
<td>76</td>
<td>1</td>
<td>P</td>
<td>G</td>
<td>126</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td>Lie (1991)</td>
<td>PA</td>
<td>208</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>96</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.44</td>
<td>0.5</td>
</tr>
<tr>
<td>Martins and Silva (2006)</td>
<td>PA</td>
<td>29</td>
<td>0</td>
<td>P</td>
<td>R</td>
<td>10</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td>Nunes, Bryant, and Olsson (2003)</td>
<td>PA</td>
<td>292</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>60</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Ouellette and Sénéchal (2008)</td>
<td>PA</td>
<td>46</td>
<td>0</td>
<td>P</td>
<td>R</td>
<td>20</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Ouellette and Sénéchal (2010)</td>
<td>PA</td>
<td>40</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>40</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
<td>−0.09</td>
<td>−0.06</td>
</tr>
<tr>
<td>Rogowsky (2010)</td>
<td>PA</td>
<td>80</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>45</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>1.00</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Sénéchal, Ouellette, Pagan, and Lever (2012)</td>
<td>PA</td>
<td>40</td>
<td>0</td>
<td>P</td>
<td>R</td>
<td>40</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Thomson, Leong, and Goswami (2013)</td>
<td>PA</td>
<td>24</td>
<td>1</td>
<td>E</td>
<td>R</td>
<td>30</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Van Otterloo, van der Leij, and Henrichs (2009)</td>
<td>PA</td>
<td>42</td>
<td>0</td>
<td>P</td>
<td>R</td>
<td>50</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.41</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>$N$</th>
<th>Design Grade</th>
<th>Number of sessions</th>
<th>Ins</th>
<th>Subject area</th>
<th>Measure</th>
<th>Study quality</th>
<th>ES post</th>
<th>ES follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeung, Siegel, and Chan (2013)</td>
<td>PA</td>
<td>76</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>60</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
</tr>
<tr>
<td>Apel and Diehm (2013)</td>
<td>PH</td>
<td>97</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>40</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
<tr>
<td>Beech, Pedley, and Barlow (1994)</td>
<td>PH</td>
<td>24</td>
<td>1</td>
<td>P</td>
<td>G</td>
<td>15</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
</tr>
<tr>
<td>Bhattacharya and Ehri (2004)</td>
<td>PH</td>
<td>40</td>
<td>0</td>
<td>S</td>
<td>R</td>
<td>6</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
</tr>
<tr>
<td>Ecalle, Magnan, and Calmus (2009)</td>
<td>PH</td>
<td>28</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>10</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
<tr>
<td>Fuchs et al. (2009)</td>
<td>PH</td>
<td>269</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>90</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>Kegel and Bus (2012)</td>
<td>PH</td>
<td>297</td>
<td>0</td>
<td>P</td>
<td>G</td>
<td>17</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>Lovett et al. (1994)</td>
<td>PH</td>
<td>62</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>35</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
</tr>
<tr>
<td>Lovett et al. (2008)</td>
<td>PH</td>
<td>160</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>26</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.94</td>
</tr>
<tr>
<td>Lovett et al. (2000)</td>
<td>PH</td>
<td>85</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>70</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
</tr>
<tr>
<td>Lovett and Steinbach (1997)</td>
<td>PH</td>
<td>122</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>20</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
<tr>
<td>Morris et al. (2012)</td>
<td>PH</td>
<td>134</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>70</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>O'Shaughnessy and Swanson (2000)</td>
<td>PH</td>
<td>30</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>126</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>Peyton (2002)</td>
<td>PH</td>
<td>99</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
</tr>
<tr>
<td>Reutzel, Fawson, and Smith (2005)</td>
<td>PH</td>
<td>144</td>
<td>0</td>
<td>S</td>
<td>R</td>
<td>4</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
</tr>
<tr>
<td>Weber and Henderson (1989)</td>
<td>PH</td>
<td>31</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>60</td>
<td>T</td>
<td>LA</td>
<td>WO</td>
<td>0.36</td>
</tr>
<tr>
<td>Chang, Sung, and Chen (2002)</td>
<td>COMP</td>
<td>60</td>
<td>1</td>
<td>E</td>
<td>G</td>
<td>35</td>
<td>T</td>
<td>C</td>
<td>Qual</td>
<td>0.78</td>
</tr>
<tr>
<td>Crowhurst (1991)</td>
<td>COMP</td>
<td>50</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>25</td>
<td>T</td>
<td>LA</td>
<td>WO</td>
<td>0.60</td>
</tr>
<tr>
<td>Crowhurst (1990)</td>
<td>COMP</td>
<td>46</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>25</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.78</td>
</tr>
<tr>
<td>DiCecco and Gleason (2002)</td>
<td>COMP</td>
<td>24</td>
<td>0</td>
<td>S</td>
<td>R</td>
<td>20</td>
<td>T</td>
<td>C</td>
<td>WO</td>
<td>0.89</td>
</tr>
<tr>
<td>Jampole, Konopak, Readence, and Moser (1991)</td>
<td>COMP</td>
<td>37</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>10</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.67</td>
</tr>
<tr>
<td>Levine (2014)</td>
<td>COMP</td>
<td>37</td>
<td>1</td>
<td>S</td>
<td>R</td>
<td>23</td>
<td>T</td>
<td>LA</td>
<td>Element</td>
<td>0.67</td>
</tr>
<tr>
<td>Study</td>
<td>Treatment</td>
<td>N</td>
<td>Design</td>
<td>Grade</td>
<td>Ss</td>
<td>Number of sessions</td>
<td>Ins</td>
<td>Subject area</td>
<td>Measure</td>
<td>Study quality</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>----</td>
<td>--------------------</td>
<td>-----</td>
<td>--------------</td>
<td>---------</td>
<td>---------------</td>
</tr>
<tr>
<td>Mason, Davison, Hammer, Miller, and Glutting (2013)</td>
<td>COMP</td>
<td>50</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>42</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.86</td>
</tr>
<tr>
<td>Meyer et al. (2002)</td>
<td>COMP</td>
<td>40</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>50</td>
<td>R</td>
<td>C</td>
<td>Qual</td>
<td>0.94</td>
</tr>
<tr>
<td>Russell (2005)</td>
<td>COMP</td>
<td>40</td>
<td>1</td>
<td>S</td>
<td>R</td>
<td>90</td>
<td>R</td>
<td>LA</td>
<td>Qual</td>
<td>0.67</td>
</tr>
<tr>
<td>Sjostrom and Chou Hare (1984)</td>
<td>COMP</td>
<td>19</td>
<td>1</td>
<td>S</td>
<td>G</td>
<td>20</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.56</td>
</tr>
<tr>
<td>Troyer (1993)</td>
<td>COMP</td>
<td>173</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>30</td>
<td>T</td>
<td>C</td>
<td>Qual</td>
<td>0.78</td>
</tr>
<tr>
<td>Vancil (2009)</td>
<td>COMP</td>
<td>57</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>30</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.67</td>
</tr>
<tr>
<td>Jiménez et al. (2010)</td>
<td>Multi</td>
<td>241</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>126</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
</tr>
<tr>
<td>Lovett, De Palma, and Frijters (2012)</td>
<td>Multi</td>
<td>188</td>
<td>1</td>
<td>S</td>
<td>R</td>
<td>70</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.83</td>
</tr>
<tr>
<td>Vaughn et al. (2006)–English</td>
<td>Multi</td>
<td>90</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>Vaughn et al. (2006)–Spanish</td>
<td>Multi</td>
<td>79</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>1.00</td>
</tr>
<tr>
<td>Cirino et al. (2009)–English</td>
<td>Multi</td>
<td>209</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>120</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.72</td>
</tr>
<tr>
<td>Cirino et al. (2009)–Spanish</td>
<td>Multi</td>
<td>209</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>120</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.72</td>
</tr>
<tr>
<td>Chenault, Thomson, Abbott, and Berninger (2006)</td>
<td>Fluency</td>
<td>20</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>20</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.08</td>
</tr>
<tr>
<td>Korat et al. (2014)</td>
<td>Vocab</td>
<td>100</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>4</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
</tr>
<tr>
<td>Research Question 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams (2009)</td>
<td>ST</td>
<td>60</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>20</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
<tr>
<td>Aram and Biron (2004)</td>
<td>ST</td>
<td>59</td>
<td>1</td>
<td>P</td>
<td>G</td>
<td>66</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.83</td>
</tr>
<tr>
<td>Bhattacharyya and Ehri (2004)</td>
<td>ST</td>
<td>40</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>35</td>
<td>T</td>
<td>LA</td>
<td>Spelling</td>
<td>0.72</td>
</tr>
<tr>
<td>Collins (1980)</td>
<td>ST</td>
<td>220</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>75</td>
<td>T</td>
<td>C</td>
<td>Spelling</td>
<td>0.67</td>
</tr>
<tr>
<td>De Jong and Share (2007)</td>
<td>ST</td>
<td>56</td>
<td>2</td>
<td>E</td>
<td>G</td>
<td>3</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
</tr>
<tr>
<td>Korat et al. (2014)</td>
<td>ST</td>
<td>100</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>4</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
</tr>
<tr>
<td>Kyte and Johnson (2006)</td>
<td>ST</td>
<td>32</td>
<td>2</td>
<td>E</td>
<td>G</td>
<td>2</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
<tr>
<td>Rosenthal and Ehri (2008) EXP1</td>
<td>ST</td>
<td>20</td>
<td>2</td>
<td>E</td>
<td>G</td>
<td>9</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>$N$</th>
<th>Design Grade</th>
<th>Ss</th>
<th>Number of sessions</th>
<th>Ins</th>
<th>Subject area</th>
<th>Measure</th>
<th>Study quality</th>
<th>ES post</th>
<th>ES follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosenthal and Ehri (2008) EXP2</td>
<td>ST</td>
<td>32</td>
<td>2</td>
<td>E</td>
<td>8</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.78</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>Share (1999) EXP3</td>
<td>ST</td>
<td>36</td>
<td>1</td>
<td>E</td>
<td>2</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Share and Shalev (2004)</td>
<td>ST</td>
<td>160</td>
<td>2</td>
<td>E</td>
<td>1</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Share (2004) EXP1</td>
<td>ST</td>
<td>36</td>
<td>2</td>
<td>E</td>
<td>3</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.56</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>Share (2004) EXP2</td>
<td>ST</td>
<td>64</td>
<td>2</td>
<td>E</td>
<td>1</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.07</td>
<td>-0.05</td>
</tr>
<tr>
<td>Share (2004) EXP3</td>
<td>ST</td>
<td>32</td>
<td>2</td>
<td>E</td>
<td>2</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.67</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Sussman (1998)</td>
<td>ST/MR</td>
<td>43</td>
<td>0</td>
<td>E</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>WO</td>
<td>0.89</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Crowhurst (1990)</td>
<td>MR</td>
<td>46</td>
<td>0</td>
<td>E</td>
<td>25</td>
<td>R</td>
<td>LA</td>
<td>Qual</td>
<td>0.78</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Koskinen et al. (2000)</td>
<td>MR</td>
<td>120</td>
<td>1</td>
<td>E</td>
<td>147</td>
<td>T</td>
<td>LA</td>
<td>Words</td>
<td>0.78</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Lee and Schallert (2015)</td>
<td>MR</td>
<td>183</td>
<td>1</td>
<td>S</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.72</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Neuman (1999)</td>
<td>MR</td>
<td>128</td>
<td>1</td>
<td>P</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.56</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Norris (2008)</td>
<td>MR</td>
<td>88</td>
<td>1</td>
<td>S</td>
<td>80</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.67</td>
<td>-0.21</td>
<td></td>
</tr>
<tr>
<td>Tsang (1996)</td>
<td>MR</td>
<td>96</td>
<td>0</td>
<td>S</td>
<td>126</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.78</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Whitehurst et al. (1994)</td>
<td>MR</td>
<td>167</td>
<td>0</td>
<td>P</td>
<td>156</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.67</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Whitehurst et al. (1999)</td>
<td>MR</td>
<td>206</td>
<td>0</td>
<td>P</td>
<td>198</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.56</td>
<td>0.13</td>
<td>0.15</td>
</tr>
<tr>
<td>Couzijn and Rijlaarsdam (2004)</td>
<td>OR</td>
<td>41</td>
<td>0</td>
<td>S</td>
<td>3</td>
<td>R</td>
<td>C</td>
<td>Qual</td>
<td>0.72</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Couzijn (1999)</td>
<td>OR</td>
<td>60</td>
<td>0</td>
<td>S</td>
<td>6</td>
<td>T</td>
<td>LA</td>
<td>WO</td>
<td>0.7</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Crasnich and Lumbell (2004)</td>
<td>OR</td>
<td>60</td>
<td>0</td>
<td>S</td>
<td>9</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.56</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>Lumbelli, Paoletti, and Frausin (1999)</td>
<td>OR</td>
<td>28</td>
<td>0</td>
<td>E</td>
<td>15</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.67</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Rijlaarsdam, Couzijn, Janssen, Braaksma, and Kieft (2006)</td>
<td>OR</td>
<td>40</td>
<td>0</td>
<td>S</td>
<td>2</td>
<td>R</td>
<td>C</td>
<td>Qual</td>
<td>0.78</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>Couzijn and Rijlaarsdam (2005)</td>
<td>OR/R&amp;A</td>
<td>90</td>
<td>0</td>
<td>S</td>
<td>2</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.80</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Lumbelli and Paoletti (2004)</td>
<td>OR/R&amp;A</td>
<td>28</td>
<td>1</td>
<td>E</td>
<td>3</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.67</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Treatment</td>
<td>N</td>
<td>Design</td>
<td>Grade</td>
<td>Ss</td>
<td>Number of sessions</td>
<td>Ins</td>
<td>Subject area</td>
<td>Measure</td>
<td>Study quality</td>
<td>ES post</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-----------</td>
<td>----</td>
<td>--------</td>
<td>-------</td>
<td>----</td>
<td>--------------------</td>
<td>-----</td>
<td>--------------</td>
<td>---------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>Moore and MacArthur (2012)</td>
<td>OR/R&amp;A</td>
<td>87</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>2</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.89</td>
<td>0.13</td>
</tr>
<tr>
<td>Holliway and McCutchen (2004)</td>
<td>R&amp;A</td>
<td>154</td>
<td>0</td>
<td>S</td>
<td>G</td>
<td>3</td>
<td>R</td>
<td>C</td>
<td>Qual</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>Philippakos and MacArthur (2016)</td>
<td>R&amp;A</td>
<td>145</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>3</td>
<td>R</td>
<td>LA</td>
<td>WO</td>
<td>0.80</td>
<td>0.42</td>
</tr>
<tr>
<td>Knudson (1993)</td>
<td>Model</td>
<td>117</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>14</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.89</td>
<td>0.32</td>
</tr>
<tr>
<td>Knudson (1989)</td>
<td>Model</td>
<td>76</td>
<td>0</td>
<td>S</td>
<td>G</td>
<td>14</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.78</td>
<td>0.26</td>
</tr>
<tr>
<td>Knudson (1991)</td>
<td>Model</td>
<td>76</td>
<td>0</td>
<td>S</td>
<td>G</td>
<td>14</td>
<td>T</td>
<td>LA</td>
<td>WO</td>
<td>0.89</td>
<td>-0.02</td>
</tr>
<tr>
<td>Other studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ross (2004)</td>
<td>Assess</td>
<td>73</td>
<td>0</td>
<td>E</td>
<td>G</td>
<td>32</td>
<td>T</td>
<td>LA</td>
<td>Qual</td>
<td>0.78</td>
<td>0.40</td>
</tr>
<tr>
<td>Daki and Savage (2010)</td>
<td>Counsel</td>
<td>14</td>
<td>0</td>
<td>E</td>
<td>R</td>
<td>32</td>
<td>R</td>
<td>LA</td>
<td>Spelling</td>
<td>0.89</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

Note. N = number of participants; Design (0 = true-experiment, 1 = quasi-experiment, 2 = participants as own control); Grade (P = preschool, E = elementary, S = secondary); Ss = type of student (G = typical students, R = students experiencing difficulties with literacy or learning); Ins = instructor (T = teacher, R = member of research team); study quality (based on proportion of quality indicators met by a study); PA = phonological awareness instruction; PH = phonics instruction; COMP = comprehension instruction; MULTI = multicomponent intervention; ST = self-teaching (read words/text to determine impact on spelling); MR = more reading; OR = observing readers; R&A = reading and analyzing text; MODEL = reading text to emulate it; Fluency = treatment for reading fluency; Vocab = treatment for teaching reading vocabulary; Assess = treatment involving reading assessment; Counsel = counseling treatment for reading difficulties; Qual = writing quality; WO = multiple measures of writing constructs; Elements = structural elements of composition.
between-study factors (see $I^2$ statistic in Table 1). The average weighted ES of 0.57 may underestimate the impact of reading instruction on writing performance, as the Trim and Fill method produced four imputed ESs to the right of the funnel plot, resulting in an adjusted average weighted ES of 0.64. This possibility of publication bias was further supported by a statistically significant correlation between study effects and variance in effects ($p < .003$ for the Begg and Mazumdar rank correlation test).

Of the studies testing the impact of teaching reading on writing, 12 of them included a measure of writing quality. Positive effects were reported in 92% of these studies. The average weighted ES of 0.63 for these 12 studies was statistically significant (see Table 2). Heterogeneity in the obtained effects of these 12 studies exceeded sampling error (see $Q$), with 64% of the variance in effects due to between-study factors (see $I^2$).

Forty studies included a spelling measure, with all but two of them yielding a positive effect (95%). The average weighted ES of 0.56 for these 40 studies was statistically significant (see Table 2). Heterogeneity in the obtained effects of these 40 studies exceeded sampling error (see $Q$), with 74% of the variance in effects due to between-study factors (see $I^2$).

Five investigations measured writing output, with each study producing a positive effect. The average weighted ES of 0.37 was statistically significant (see Table 2). Heterogeneity was not statistically significant. An average weighted ES was not computed for other writing measures, as they were assessed infrequently (i.e., structural elements in three studies, organization and mechanics in two studies, and vocabulary in one study).

In terms of specific treatments, positive effects were reported in all but one study involving phonological awareness (95%). A statistically significant average weighted ES of 0.69 was obtained (see Table 1). Heterogeneity in the obtained effects of the 20 studies testing the teaching of phonological awareness exceeded sampling error alone (see $Q$), with 77% of the variance in effects attributable to between-study factors (see $I^2$). Spelling was the main writing outcome in 18 of these 20 studies, and the average weighted ES for spelling was a statistically significant 0.68 (95% CI [0.55, 0.81]). The $Q$ statistic was statistically significant; 78% of the variance was due to between-study factors.

For phonics instruction, all studies produced a positive effect, and each study measured spelling (Weber & Henderson, 1989, measured spelling and writing mechanics). The average weighted ES of 0.39 for writing overall was statistically significant (see Table 1). The $Q$ statistic was not statistically significant (see Table 1). When spelling was the single focus of the analysis, a statistically significant ES of 0.41 (95% CI [0.27, 0.55]) was obtained. Again, heterogeneity was not statistically significant.

All studies testing the effects of reading comprehension instruction on writing overall produced positive effects, resulting in a statistically significant average weighted ES of 0.66 (see Table 1). Heterogeneity in the obtained effects of the 12 comprehension instruction studies exceeded sampling error alone (see $Q$), with 65% of the variance in effects due to between-study factors (see $I^2$). Writing quality was an outcome in 10 of the 12 comprehension studies, producing a statistically significant ES of 0.72 (95% CI [0.37, 1.07]). Heterogeneity was again
statistically significant, with 68% of the variance in ESs due to between-study factors. In the five comprehension studies that assessed writing output, a statistically significant ES of 0.38 (95% CI [0.07, 0.70]) was obtained; heterogeneity in ESs was not statistically significant.

The writing outcome for the four studies testing multicomponent reading interventions was spelling. Three of these studies yielded positive effects. The average weighted ES was 0.47, and it was not statistically significant (see Table 1).

**Effects at Maintenance**

Eleven studies examining the impact of teaching reading on writing included a maintenance writing measure. Six involved phonological awareness instruction, two phonics instruction, two multicomponent interventions, and one comprehension instruction (see Table 4; Follow-up column). For follow-up assessments by Byrne and Fielding-Barnsley (1993, 1995) and Byrne et al. (2000) and Ecalle et al. (2009), ESs were averaged across time points. In addition, one study (Cirino et al., 2009) included an experiment with English language learners. This study was not used in the analysis of maintenance effects, as all other studies were conducted with first language learners.

Seventy percent of maintenance assessments conducted with first language learners produced positive effects, resulting in a statistically significant average weighted ES of 0.37 for writing overall (see Table 3). Heterogeneity was not statistically significant (see \( Q \)). All but one study involved spelling as the outcome measure. When the analysis just focused on spelling, a statistically averaged weighted ES of 0.36 (95% CI [0.14, 0.58]) was obtained, and again heterogeneity was not statistically significant.

**Moderator Analysis**

For Research Question 1, moderator analyses were conducted to determine if excess variability in ESs for writing overall were related to study characteristics. When individual study characteristics were tested, type of experiment (true-experiment ES = 0.46 and quasi-experiment ES = 0.99; \( Q \) [between] = 6.83, degrees of freedom \([df]\) = 1, \(p < .01\)) and grade level (preschool ES = 0.61, elementary ES = 0.46, and secondary = 0.57; \( Q \) [between] = 9.30, \(df = 2, p < .02\)) were statistically significant moderators. The average weighted ESs at all three grade levels and each type of experiment were statistically greater than zero.

There was no statistical difference by type of publication (published vs. unpublished studies; \(p = .31\)), assessment (norm-referenced vs. researcher designed; \(p = .25\)), sample size (\(p = .27\)), instructor (teacher vs. research team; \(p = .94\)), overall quality of the study (\(p = .21\)), type of student (typical students vs. those experiencing challenges learning to read/write; \(p = .85\)), or number of instructional sessions (\(p = .67\)).

When all variables were entered into the moderator analyses, they accounted for 32% of the variance in ESs (\(Q = 28.04, df = 10, p < .002\)). Five variables made a unique contribution to predicting variability in ESs after controlling for the other study characteristics. These were number of sessions (\(p = .048\)), type of experiment (\(p = .049\)), publication (\(p = .03\)), sample size (\(p = .01\)), and elementary versus secondary (\(p = .048\)).
Research Question 2: Does Increasing Students’ Interaction With Words or Text Through Reading or Observing Others Read Enhance Student Writing?

Characteristics of Studies

For Research Question 2, 36 studies were located (see Table 4) that involved 3,060 students in preschool to Grade 12. Four studies (Couzijn & Rijlaarsdam, 2005; Lumbelli & Paoletti, 2004; Moore & MacArthur, 2012; Sussman, 1999) had more than one treatment pertinent to Research Question 2. This increased k to 40 when we tested the effectiveness of the five identified specific treatments (e.g., self-teaching, reading and analyzing text). Three studies included in Research Question 2 also contributed an experiment to Research Question 1 (Bhattacharya & Ehri, 2004; Crowhurst, 1990; Korat et al., 2014), as they included a separate treatment testing the effectiveness of teaching reading.

The majority of studies addressing Research Question 2 involved younger students (58% of the studies were conducted with students in Grades 1 to 6, and 11% of studies were conducted with preschool to kindergarten children). Seventy-eight percent of the studies were published articles. The 36 experiments were conducted in a variety of countries including the United States (47%), Israel (19%), Netherlands (14%), Italy (6%), Canada (6%), Hong Kong (3%), Slovenia (3%), and South Korea (3%). Reading instruction involved English orthography in slightly more than half of the investigations (58%). Eighty-three percent of studies involved typically developing students. In studies providing information on gender (42%), race (32%), and SES (24%), 48% of students were boys, 63% were from minority groups, and 78% of youngsters were from either poor households or communities. As with Research Question 1, caution needs to be exercised in interpreting race and SES data, as the reporting of such data may have occurred more often when studies were conducted in specific communities.

Most of the experiments focused on self-teaching (38%) or more reading (23%), with the remaining studies involving observing readers (20%), reading and analyzing text (13%), and reading model text to emulate it (8%). Self-teaching studies were conducted almost exclusively with elementary grade students (93%; see Table 4), whereas more reading studies included students from preschool to high school. Observing readers involved students only in 5th to 12th grade, whereas students in reading and analyzing text were in fourth to ninth grade. Reading to emulate model text included students in Grades 3 to 8.

The overall quality of studies was generally strong: 60% were true-experiments (19% were quasi-experiments and 21% involved a participant as own control design), over 90% evidenced no issues with attrition, 100% included more than a single group or class in each condition, 72% controlled for teacher effects, and pretest equivalence and reliability were not an issue for 87%, and 61% of writing measures, respectively. The most notable weakness in study quality involved treatment fidelity. This was established in only 13% of studies. In addition, floor/ceiling problems were evident in just 47% of writing assessments.

Effects at Posttest

Eighty-nine percent of the 36 studies produced positive effects at posttest (see Table 4). These 36 studies yielded a statistically significant average weighted ES of
0.35 (see Table 1). Heterogeneity in the obtained effects of these 36 studies exceeded sampling error alone (see $Q$), with 49% of the variance in effects attributable to between-study factors (see $I^2$). Publication bias did not appear to be a problem, as the Trim and Fill method imputed no ESs to the left or the right of the funnel plot and the correlation between study effects and variance in effects (as assessed with the Begg and Mazumdar rank correlation test) was not statistically significant.

Of the 36 studies, 21 of them included a measure of writing quality. Positive effects were reported in 95% of these studies. The average weighted ES of 0.44 for these 21 studies was statistically significant (see Table 2). Heterogeneity of ESs did not statistically exceed sampling error (see $Q$). Fifteen studies included a spelling measure. The results for these studies are presented below, as they all tested the self-teaching treatment. No other writing measures occurred in more than two investigations for Research Question 2.

In terms of specific interventions, 15 studies tested the self-teaching hypothesis effect on spelling. Positive effects were reported in 87% of these investigations, resulting in a statistically significant average weighted ES of 0.28 (see Table 1). Heterogeneity in the obtained effects of these 15 studies exceeded sampling error (see $Q$), with 58% of the variance in effects due to between-study factors (see $I^2$). In eight of the studies (e.g., Share, 2004), students acted as their own controls, receiving different numbers of exposures to words read in isolation or context (e.g., two vs. six exposures; De Jong & Share, 2007). When the outcomes from these studies were statistically compared with the other seven studies using a true- or quasi-experimental design, there was no statistically significant difference by type of experiment ($Q = 1.29, df = 1, p = .26$).

The treatment in nine studies involved more reading (none of these studies involved a spelling measure). Seventy-eight percent of these investigations produced a positive effect. The average weighted ES of 0.29 for writing overall for these nine studies was statistically significant (see Table 1). Heterogeneity in the obtained effects of these nine studies was not statistically significant (see $Q$). When we just examined writing quality, the ES of 0.31 (95% CI [0.14, 0.47]) was statistically significant, and heterogeneity in effects did not statistically exceed sampling error ($Q = 7.68, p = .36$).

Eight studies tested the effectiveness of observing readers. All eight studies produced a positive effect, resulting in a statistically significant average weighted ES of 0.62 for writing overall (see Table 1). Heterogeneity in the obtained effects of these nine studies was not statistically significant (see $Q$). When we just examined writing quality, the ES of 0.67 (95% CI [0.44, 0.91]) tested in eight studies was statistically significant, and heterogeneity in effects did not statistically exceed sampling error ($Q = 8.69, p = .28$).

In five studies, students read and analyzed text produced by other students. All of these investigations resulted in a positive effect, producing a statistically significant average weighted ES of 0.43 for writing quality (see Table 1). The $Q$ statistic was not statistically significant (see Table 1).

There were three studies examining reading model text with the idea of emulating it. Two of these studies (67%) produced a positive effect. Collectively, the three studies produced an average weighted ES of 0.15 for writing overall. This ES was not statistically significant (see Table 1). All three of these studies included
a measure of writing quality. The results were the same when just these measures were included in the analyses, resulting in an ES of 0.22 (95% CI [−0.08, 0.51]) that was not statistically significant.

**Effects at Maintenance**

Only 5 of the 36 studies in Research Question 2 included a maintenance writing measure. (see Table 4; Follow-up column) The average weighted ES for these studies for writing overall was 0.15, and this effect was not statistically significant (see Table 3). Only two of the five studies produced positive results. The results were almost identical when writing quality (assessed in four studies) was employed as the writing outcome (ES = 0.15; CI [−0.12, 0.42]).

**Moderator Analysis**

Moderator analyses were conducted to determine if excess variability in ESs testing Research Question 2 were related to study characteristics. None of the individual study characteristics when entered into the analysis alone accounted for a statistically significant amount of variance (all ps > .11). Likewise, when all variables were entered together, they did not account for a statistically significant amount of variance in ESs (Q = 7.32, df = 11, p = .77).

**Discussion**

This meta-analysis examined if students’ writing performance is enhanced by (a) teaching reading or (b) increasing students’ interactions with words and text through reading or observing others read. This analysis allowed us to determine if specific reading interventions are responsible for improvements in students’ writing. It also provided evidence about one aspect of the theoretical proposition that there is a bidirectional relationship between reading and writing (Shanahan, 2016). This theoretical proposition rests in part on the assumption that reading affects writing. Previous meta-analysis provided support for two other assumptions underlying this theory (i.e., writing affects reading and reading and writing conjointly affect each other), demonstrating that writing interventions improved reading performance (Graham & Hebert, 2011a, 2011b; Graham & Santangelo, 2014) and combined reading/writing interventions improved both reading and writing performance (Graham et al., in press).

**Does Teaching Reading Improve Students’ Writing Performance?**

While reading and writing are related but not identical skills (Grabe & Zhang, 2016; Langer, 1986; Langer & Applebee, 1987; Pearson et al., 2012; Shanahan, 2016), teaching readings improved students’ writing performance. When preschool to high school students were taught how to read, 19 out of every 20 studies produced a positive effect at posttest, resulting in meaningful improvements on a composite measures of overall writing performance as well as on specific measures of writing quality, spelling, and writing output. More specifically, phonological awareness, phonics, and reading comprehension instruction strengthened students’ writing performance. Furthermore, in the 20% of studies that assessed the impact of reading instruction over time, observed writing performance gains were maintained at meaningful, but slightly lower levels than what had been
observed at posttest (one study found positive effects six years after reading instruction was delivered; Byrne et al., 2000). Most of the maintenance studies (73%) involved phonological awareness and phonics instruction.

It is important to note that considerable confidence can be placed in our finding that teaching reading improves students’ writing performance. First, almost three quarters of the included studies were true-experiments, and each study, on average, met almost 80% of the other indicators of study quality (e.g., reliable measures, no floor or ceiling effects). Second, while quasi-experiments yielded larger effects than true experiments and there were differences in effects by grade, true-experiments still yielded a statistically significant and meaningful effect, as did studies with preschool, elementary, and secondary students. Furthermore, magnitude of effects was not related to type of student taught (typical students vs. those experiencing difficulty in school), type of assessment (researcher-devised vs. norm-referenced), or instructor (teachers vs. research team members). Collectively, all nine methodological features examined in this review accounted for 32% of the variance in effects, with type of experiment, grade, publication, sample size, and number of sessions making a unique contribution to predicting variability in effects. Even so, both types of publication yielded statistically significant and meaningful effects (published studies = 0.44 and unpublished studies = 0.37), and the associations between magnitude of effects and sample size and number of sessions were small (−0.12 and −0.15, respectively). Third, the analysis of the effects of teaching reading involved a reasonable number of studies (54) across a variety of writing measures. As a result, researchers, educators, policy makers, and material developers can be confident that reading instruction strengthens writing.

To contextualize the effects of reading instruction on writing performance obtained in this review, we draw on two recent meta-analyses. Graham and Harris (2018) found that for students in Grades 1 to 12, the average weighted ESs for writing quality for 11 writing treatments ranged from 1.26 for strategy instruction to −0.17 for grammar instruction, with a median average-weighted ES of 0.50. The ES for writing quality of 0.63 obtained in this review compares quite favorably to other treatments designed to improve the quality of students’ text. Likewise, Graham and Santangelo (2014) reported that spelling instruction produced an average weighted ES of 0.54 for spelling performance. This ES is slightly less than the ES of 0.56 obtained in this meta-analysis.

*Does Increasing Students’ Interaction With Text Enhance Students’ Writing Performance?*

Increasing students’ interaction with text improved their writing performance. When preschool to high school students were directed to interact with words or text or observe others doing so (e.g., watching a student read text to carry out a science experiment), 18 out of every 20 studies produced an immediate positive effect, resulting in meaningful improvements on a composite measures of overall writing performance as well as specific measures of writing quality and spelling. More particularly, reading individual words, increasing the amount of reading students did, reading and analyzing text produced by others, and observing other readers interact with text enhanced students’ performance on one or more measure of writing (mostly involving spelling or writing quality).
The long-term impact of these effects, however, is not certain as writing gains were not maintained over time. Interpretation of this finding must be tempered by the fact that it is based on a small number of studies ($K = 5$; see Table 3). In addition, with one exception (i.e., Whitehurst et al., 1999), all five of these studies involved very brief treatments. In the two experiments by Share (2004), students received no more than four new exposures to words read, whereas the two studies by MacArthur and colleagues (Moore & MacArthur, 2012; Philippakos & MacArthur, 2016) involved no more than three instructional sessions. Because such short treatments are unlikely to produce lasting effects (Pressley, Graham, & Harris, 2006), additional research involving more robust and extended treatments are needed before drawing a conclusion about the long-term impact on writing of increasing students’ interactions with text or observing others do so.

With the exception of the outcome for maintenance, confidence can be placed in the finding that interacting with words or text and observing others interacting with text enhances students’ immediate writing performance. Sixty percent of the included studies testing the immediate impact of these treatments were true-experiments, and each study on average met close to 70% of the other indicators of study quality (e.g., controlled for teacher effects, no differential attrition). None of the nine methodological features tested were statistically related to the magnitude of study effects either individually or collectively, suggesting that none of these features biased the overall outcomes obtained in our analyses. Furthermore, this analysis involved enough studies (36) to establish reliable point estimates. In one way, the generality of these findings were not as broad as they were for the impact of teaching reading (Research Question 1). In another way, they were broader. Only 17% of studies testing students interacting with words or text or observing others interact with text involved youngsters experiencing difficulties with literacy learning (studies where reading was taught were almost equally divided between the two types of students). In contrast, studies testing the impact of reading and observing readers involved a broader set of orthographies (58% involved English orthography), but 80% of studies testing reading instruction involved English orthography.

To contextualize the effects of students’ increased interaction with text examined in this review, we again draw on the Grade 1 to 12 meta-analysis conducted by Graham and Harris (2018). They found that increasing how much students wrote resulted in an ES of 0.24 for writing quality, whereas we found that increased interaction with text with a reading focus resulted in an ES of 0.44 for writing quality in this review.

**Theoretical Implications**

The findings from this study provided support for one of the assumption underlying the theoretical proposition that there is a bidirectional relationship between reading and writing (Shanahan, 2016). This theoretical proposition is based in part on the assumption that reading affects writing. Our meta-analysis supports this assumptions by demonstrating that reading interventions enhance students’ performance on multiple writing measures.

It is important to note, that the findings form our meta-analysis did not provide evidence on other important assumptions underlying bidirectional relationship theories of reading and writing, including the assumptions that writing affects
reading and reading and writing conjointly affect each other. However, evidence from other meta-analyses supported these two assumptions showing that writing interventions improved reading performance (Graham & Hebert, 2011a, 2011b; Graham & Santangelo, 2014) and combined reading/writing interventions improved both reading and writing performance (Graham et al., in press). Taken together, the accumulated evidence from the current and previous analyses support the theoretical propositions underlying a bidirectional model of reading and writing relations (Shanahan, 2006, 2016).

Possible Limitations and Attempted Solutions

Our meta-analysis involved systematically aggregating and analyzing the findings from individual studies to answer specific research questions. Before considering implications for research or practice, it is important to note that the value of any conclusion drawn in this review depends on a variety of factors such as the quality of studies reviewed. We applied multiple techniques to address this issue. One, we specified how much confidence can be placed in our findings based on the number and quality of studies analyzed. We were able to draw conclusions about the quality of studies for two reasons. One, we applied relatively stringent criteria for the studies included in our review. We did not include investigations in this review if the study design was not rigorous, attrition was greater than 20%, fewer than 10 students participated in each instructional condition, a meaningful writing assessment was not administered, or a major flaw was evident that confounded interpretation of the findings. Two, when a study was included, we examined it to assess a range of quality indicators including differential attrition, number of instructional groups in each condition, teacher effects, pretest differences between conditions, reliability of measures, ceiling and floor effects, and treatment fidelity. With the exception of establishing treatment fidelity, this examination showed that the studies included in this review were methodologically sound in general.

It is worth noting that experimental design (true-experiment vs. quasi-experiment) was statistically related to variability in ESs for studies involving reading instruction, but a total score for the other quality indicators was not statistically significant in this analysis. One possible reason for this is that we did not examine each quality indicator separately (as was done for experimental design). However, a subsequent post hoc analysis did not find any statistically significant relationships with variability in ESs and the other quality indicators. A second and more plausible reason that such relationships were not observed is that study selection criteria were so stringent that it significantly reduced variability in the quality of studies reviewed. In any event, examining each study included in this review provided information to help us determine the confidence that can be placed in our findings and to make recommendations for future research (see Implications for Research).

The value of the conclusions drawn in this review also depend on other factors such as similarity of outcome measures, who participated in included studies, and number of missing studies. To address the first factor (similarity of outcome measures), we applied a shifting analyses approach, where we summarized ESs based on a combined effect for all writing measures combined as well as for individual
measures such as writing quality, spelling, and writing output. The second factor (who participated in included studies) was tackled by specifying the grade levels and types of students for each conclusion that was drawn. The third factor (missing studies) was addressed through the application of multiple search strategies (electronic searchers, hand searches of journals and published reviews, contacting authors, and conducting a google search of included studies) and the use of two post hoc analytic strategies (i.e., Trim and Fill and Begg and Mazumdar rank correlation test) to gauge the likelihood of publication bias.

**Implications for Research**

An important function of a literature review, including meta-analysis, is to provide suggestions for future research based on the analyses of existing research and its strengths and weaknesses. Despite our multiple efforts to locate studies that isolated the effects of reading interventions on students’ writing performance, we located less than 100 studies that met our inclusion criteria. Why were more studies not located given the hundreds-of-thousands of investigations involving reading conducted previously (National Institute of Child Health and Human Development, 2000)? One reason why such studies are not more frequent is that reading researchers do not commonly include measures of writing when testing reading treatments. This was clear in the over 17,000 abstracts and almost 700 papers we reviewed. Given the presumed theoretical importance of reading interventions as a means for enhancing writing, we encourage reading researchers and research funding agencies to make writing assessments a common feature of reading intervention research.

Another reason why we did not locate more studies is that writing was part of the reading intervention in about one out of every seven studies we examined. We were not able to include these studies, as control students did not receive an equivalent amount of writing instruction (our purpose in conducting the meta-analysis was to isolate the effects of reading interventions). This most frequently occurred in studies testing phonological awareness and phonics instruction, as spelling instruction was often an integral part of such treatments. It was also relatively common in comprehension instructional studies where writing was included as a way to facilitate students’ understanding of text. We hope researchers will be more specific about the role of writing in reading intervention studies in the future, and encourage scientists to design experiments that test the unique contribution of both reading and writing on a broad array of literacy outcomes in such studies.

It is further striking that only four specific interventions testing the impact of reading or reading instruction on writing performance were investigated in 10 or more studies. This included phonological awareness, phonics, and comprehension instruction as well as reading words or text to determine if spelling was enhanced (i.e., self-teaching treatment). Only one of these treatments (phonological awareness) was tested in at least 20 studies. Consequently, research is needed to provide additional verifications that the four treatments identified above enhance writing. But even more important, the impact of other reading treatments on writing, such as reading fluency instruction, vocabulary instruction, more reading, teaching students about the functions and structure of text, reading and analyzing text, and observing readers interact with text, need to be more fully tested. In addition,
multicomponent reading interventions or reading model text to emulate it did not statistically enhance writing performance in this review. More research is needed to test these treatments as they involved four and three studies, respectively. There is also a need to conduct more research with secondary students, as only 20% of all investigations involved these adolescent learners.

Furthermore, the studies in this review that focused on teaching reading mostly concentrated on phonological awareness, phonics, and reading comprehension. Additional research is needed to determine if reading fluency as well as reading vocabulary instruction improves students’ writing performance. Moreover, we located no studies investigating if reading interventions designed to help students unravel and explore author’s intentions facilitates writing performance. Additional research is needed to test these and other reading interventions.

In the studies included in this meta-analysis, writing outcomes were mostly limited to spelling performance and measures of writing quality. Future research needs to expand its focus to include other product measures (e.g., textual organization and cohesion, writing vocabulary, structural elements) as well as measures of writing knowledge, process, and motivation. As Shanahan (2016) has noted, reading and the teaching of reading should influence multiple and possibly diverse aspects of writing performance.

An especially critical goal for future studies examining the impact of reading interventions on writing is to determine if obtained effects are maintained over time. In a relatively small number of studies, we found that this was the case for investigations that involved the teaching of reading skills and strategies, but not for studies that examined the effectiveness of interacting with words or text or observing others do so. Additional research is needed to verify these findings, as well as investigate means for promoting such maintenance.

While many of the studies we reviewed provided the theoretical underpinnings for why the tested reading intervention should improve reading performance (this was the primary purpose of the majority of the studies in this meta-analysis), they usually did not provide a theoretical justification for why the reading treatment should strengthen writing performance. Such theoretical framing needs to become part of the rationale for all reading intervention studies that include a writing measure in the future.

Finally, while the studies reviewed in this meta-analysis were relatively solid methodologically (e.g., they were generally of higher quality than studies in Graham & Hebert, 2011a, 2011b, examining the impact of writing interventions on reading), there is room for improvement. This includes consistently measuring if reading treatments were delivered as intended (this was only done 31% of the time), eliminating ceiling/floor effects for writing measures (done 64% of the time), ensuring writing measures are reliable and teacher effects are controlled (done 69% of the time in both cases), and conducting true-experiments (done 71% of the time). Just as importantly, researchers need to do a better job of providing information about participants, as information about participants’ race, SES, and gender were provided in one-half of the studies or less. These methodological and reporting issues are not just a problem in the research examined here, but in educational research more broadly (Graham, 2015).
Implications for Practice

This meta-analysis provides support for the premise that teaching reading as well as increasing students’ interaction with text, either directly through some form of reading or indirectly through observing readers, enhances students’ writing performance. These findings provide support for the proposition that reading can and should be part of our instructional efforts to improve students’ writing.

As noted at the start of this article, many effective practices for teaching writing have been identified, with most of these focusing on increasing how much students write or directly teaching writing skills, processes, or strategies (e.g., Graham et al., 2015). Even so, the identification of practices that can strengthen students’ writing performance is incomplete, as demonstrated by our review. Our findings support the assertion that reading and reading instruction should be part of a well-balanced writing instructional writing program. These findings have important implications for policy makers in how they envision the shape and structure of 21st century literacy programs, how teachers enact classroom writing instruction, how instructional designers build new writing programs, and how researchers construct the writing treatments they plan to test. Reading and reading instruction needs to be part of these endeavors.

Despite the positive impact of reading interventions on students’ writing performance, it would be a mistake to assume that writing does not have to be taught directly—that all we need to do is to teach reading and increase students’ interaction with text. First, there is considerable evidence that writing is improved by teaching it (e.g., Graham et al., 2012; Graham & Perin, 2007; Pearson et al., 2012). Second, this meta-analysis did not test if reading treatments are as effective as writing treatments. In fact, we eliminated studies where the reading treatment included writing instruction, unless an equivalent amount of writing instruction was provided to controls (K = 6) or treatment students received a small amount of writing instruction that should not have influenced the writing outcome measure (K = 1). Third, the studies included in this review examined important but a relatively limited set of possible writing outcomes. As a result, the strongest conclusion that can be drawn from this meta-analysis is that reading interventions are one of many means for improving students’ writing, but the available evidence does not support the contention that reading interventions should replace writing interventions.

References

References marked with an asterisk indicate studies included in the meta-analysis.


**Authors**

STEVE GRAHAM is the Warner Professor in the Mary Lou Fulton Teachers College at Arizona State University and a research professor at the Learning Sciences Institute Australia at the Australian Catholic University in Brisbane; email: steve.graham@asu.edu. His research focuses on the following issues in writing: development, assessment, instruction, practice, evidence-based practices, and theory. He also examines the links between writing, reading, and learning. His research in writing has involved typically developing writers, students with special needs, and students whose families live in poverty.

XINGHUA LIU is currently a lecturer of applied linguistics at Shanghai Jiao Tong University, China; email: liuxinghua@sjtu.edu.cn. His research focuses on the cognitive processes of written production and corpus-based linguistic analysis of written discourses. His research deals with both first and foreign language writing, academic writing, and writing produced by abnormal population.

BRENDAN BARTLETT is professor of education at the Australian Catholic University in Brisbane; email: brendan.bartlett@asu.edu. His research in literacy is on how people construct ideas in texts they encounter or create, how they communicate, remember and make sense of such ideas, connections between quality of literacy and successful post-school transition, and how young people with poor histories of schooling may be retrieved as willing and effective communicators.

CLARENCE NG is an associate professor at the Learning Sciences Institute Australia (LSIA) at the Australian Catholic University in Brisbane; email: clarence.ng@acu.edu.au. He is the research director for the Learning, Learning Diversity and Reforming Classroom Practices concentration at LSIA. His current research focuses on the learning and engagement of diverse student groups with a particular interest in promoting reading engagement for disadvantaged students.

KAREN R. HARRIS is the Warner Professor in the Mary Lou Fulton Teachers College at Arizona State University and a research professor at the Learning Sciences Institute
Graham et al.

Australia at the Australian Catholic University in Brisbane; email: karen.r.harris@asu.edu. Developer of the Self-Regulated Strategies Development (SRSD) instructional model, most of her research involves students and teachers in low income or title 1 schools. Her current research focuses on continued refinement of SRSD and validation of strategies integrating reading and writing, as well as practice based professional development for SRSD for general and special education teachers.

ANGELIQUE AITKEN is a doctoral student in the Mary Lou Fulton Teachers College at Arizona State University; email: angelique.aitken@asu.edu. Her research interests include writing motivation, choice in writing, and writing for students with disabilities.

ASHLEY BARKEL is a doctoral student in the Mary Lou Fulton Teachers College at Arizona State University; email: ashley.barkel@asu.edu. She specializes in special education and English language learners. Ashley’s father, an elementary teacher, inspired her love and passion for exploring, learning, and teaching. Her research interests include intervention research for students who struggle with learning, writing instruction, self-regulation, and professional development for teachers.

COLIN KAVANAUGH was a doctoral student in the Mary Lou Fulton Teachers College at Arizona State University at the time this meta-analysis took place. He is now a high school teacher in New York. His research interest involve literacy and secondary students with special needs.

JOY TALUKDAR is research executive officer at the Institute for Positive Psychology and Education, Australian Catholic University; email: joy.talukdar@acu.ed.au. His doctoral research in the field of teacher education was awarded with a Dean’s Commendation for Doctoral Research Thesis Excellence. His research focuses on various aspects of teacher education: initial and continuing; assessment in teacher education; standards, quality, and evidence in teacher education; teacher pedagogical practice; teacher preparation; and teaching and teacher quality.