

Can Reading Self-Efficacy Be Modified? A Meta-Analysis of the Impact of Interventions on Reading Self-Efficacy

Norman J. Unrau

California State University, Los Angeles

Robert Rueda

University of Southern California

Elena Son

Hankuk University of Foreign Studies

Joshua R. Polanin

American Institutes for Research

Rebecca J. Lundeen and Alison K. Muraszewski

University of Southern California

Studies of interventions' impact on reading self-efficacy have been conducted since the 1980s. The purpose of this project was to conduct a systematic review of these studies because the primary studies often yielded divergent results. Included studies entailed an intervention, addressed reading specifically, and reported explicit pre- and postintervention measures of reading self-efficacy. Subjects were students in elementary grades through college. The results of a systematic search and screening procedure found 30 studies in which 2,300 subjects received treatments of various kinds while 1,957 were in control or comparison groups. A meta-analysis of three subsets of study designs revealed that each subset generated a significant effect size: treatment–control ($g = 0.24$, 95% confidence interval [CI] [0.10, 0.39]); treatment–comparison ($g = 0.44$, 95% CI [0.04, 0.84]); pretest–posttest ($g = 0.36$, 95% CI [0.16, 0.57]). Significant heterogeneity was found and modeled using moderator analyses conducted on several variables. The results indicated that significant moderators of effect sizes included grade level, number of sources shaping reading self-efficacy, a reading self-efficacy measurement index, and journal publication. In studies that measured the impact of the intervention on reading comprehension, its relationship with

reading self-efficacy was analyzed revealing a strong correlation between the two constructs. Discussion includes an exploration of the importance of these findings to future policy, practice, and research on the design of reading self-efficacy measurement instruments and on interventions that utilize major sources of experiences shaping reading self-efficacy.

KEYWORDS: reading, self-efficacy, interventions, meta-analysis

Many aspects of motivation and their relationship to achievement have been investigated. However, we believe that self-efficacy, as a motivational construct, warrants deeper investigation as a dimension of motivation having an impact on reading. While correlational research on self-efficacy has been plentiful and includes several meta-analyses of self-efficacy and its relationship with educational outcomes (Holden, Moncher, Schinke, & Barker, 1990; Multon, Brown, & Lent, 1991; Richardson, Abraham, & Bond, 2012), less abundant are intervention studies that attempt to modify students' self-efficacy beliefs. Intervention studies drive the field forward because of insights gained about their impact on self-efficacy or educational outcomes, such as reading comprehension. Of those intervention studies designed to investigate impact on self-efficacy, many have shown promising significant effects (Guthrie, McRae, & Klauda, 2007). Furthermore, intervention studies yield knowledge about the design and implementation of interventions that are likely to improve educational outcomes. These understandings can then guide educators in the development and selection of educational interventions that are more likely to generate robust educational results, such as improvements in motivation and reading comprehension, based on empirical findings.

The intent of our inquiry was to carry out a meta-analytic review of intervention studies that were grounded in self-efficacy theory and that measured changes in reading self-efficacy. We also aspired to explain variations in the magnitude of an intervention's effectiveness through the exploration of factors across studies, such as study design, intervention features, methods of measurement, and grade level of subjects. Where studies included information about the impact of an intervention on reading comprehension, we explored the relationship between change in reading self-efficacy and change in reading comprehension. The knowledge gained from our investigation will inform subsequent research into reading self-efficacy and the design of future interventions that could foster reading development.

What Is Reading Self-Efficacy?

Self-efficacy has been defined by Bandura (1986) as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (p. 391). In other words, self-efficacy may be conceived as a personal belief about what an individual is capable of learning or doing by means of organizing and carrying out actions that lead to a successful outcome. Because self-efficacy for reading is the form of self-efficacy we investigated in this study, we define it as readers' perceptions of competence in their ability to successfully complete reading tasks (Chapman & Turner, 1995; Guthrie & Coddington, 2009).

Self-efficacy has been demonstrated to exert a profound influence on student motivation for learning, self-regulation, and performance (Pajares, 1996; Schunk & Pajares, 2009). Bandura (1986, 1997) noted that the beliefs people have about their capability, for example, as readers, serve as better predictors of their behavior than what they actually accomplish. Therefore, self-perceptions exert an enormous influence on how people engage their skills and knowledge as readers. Previous research found positive relationships between students' self-efficacy beliefs and reading performance across a range of developmental levels (Chapman & Turner, 1995; Mills, Pajares, & Herron, 2007; Wigfield & Guthrie, 1997). Research also suggests that readers are more likely to demonstrate effort and persistence in reading a text if they believe in their capacity to comprehend it successfully (Solheim, 2011; Waleff, 2010). In other words, readers with high self-efficacy in reading engage in more reading-related activities.

Self-efficacy is one of a cluster of self-processes that have received attention in investigations of motives that drive and control our actions and learning, such as self-concept, self-worth, self-esteem, academic competence beliefs, and outcome expectations (Linnenbrink-Garcia & Patall, 2016). While many of these constructs share conceptual features with self-efficacy, they are distinct from it. Although self-efficacy and self-concept share some similarities, such as perceived competence and multidimensionality, a reader's self-efficacy beliefs are not the same as that reader's self-concept. Self-concept refers to an individual's collective self-perceptions, whereas self-efficacy is more specific to domains, tasks, and beliefs about how an individual will perform on context-specific tasks in specified domains, such as comprehending an editorial on immigration on an English exam. Self-efficacy, in fact, acts as a precursor to the development of self-concept (Bong & Skaalvik, 2003). Self-efficacy is also future-oriented and malleable, whereas self-concept is oriented toward the past and is characterized by its relative stability.

A Theory of Change for Reading Self-Efficacy

Within social cognitive theory, Bandura (1986, 1997) has provided a general overall model of development and change in self-efficacy. His triadic reciprocal causation model identifies three variables (internal personal factors, behavioral patterns, and environmental influences) whose bidirectional interactions explain human functioning. Therefore, altering one part of the system should produce alterations in the others. For example, to improve student functioning, teachers can implement strategies to enhance personal factors, such as cognitive or motivational processes, boost behavioral competencies through self-regulation, or alter the environment by modifying classroom structures. Interventions to modify a person's reading self-efficacy beliefs, a personal factor in the triadic model, could influence behavior in the form of increased reading engagement that would, in turn, affect the classroom environment. Those interventions might be designed to affect one or more sources of self-efficacy beliefs.

According to Bandura (1997), there are four major sources of self-efficacy: mastery experiences, vicarious experience, verbal and social persuasion, and emotional and physiological states. Mastery experiences related to reading might include teaching students research-based reading strategies that enhance their

comprehension and consequently lead to successful experiences. Vicarious experience includes modeling reading strategies that demonstrate an approach to reading that improves efficiency and effectiveness. Verbal and social persuasion might include offering supportive feedback to students who demonstrate an effective application of a reading strategy. Emotional and physiological states could be exemplified by a struggling reader whose high levels of anxiety when asked to read aloud in class are interpreted as low confidence in completing the task successfully. The development of academic self-efficacy beliefs has been extensively researched (Schunk, Hanson, & Cox, 1987), including the investigation of differential influences arising from the four major antecedents with mastery experiences being the most powerful source (Usher & Pajares, 2008) but with the indication that major sources promoting self-efficacy's growth vary over time (Phan & Ngu, 2016).

While we have a general theory of change in reading self-efficacy, we lack research that has synthesized studies using a range of interventions to discover what sources of reading self-efficacy are present in those interventions and the magnitude of their effects. It can be hypothesized that the more major sources of reading self-efficacy targeted, the greater the change in reading self-efficacy registered. While no investigation of motivational interventions targeting different sources of reading self-efficacy beliefs emerged in our review of the literature, we included the number and type of sources of self-efficacy beliefs incorporated in an intervention as moderators to explore their effect in the present study.

The Challenges of Measuring Self-Efficacy

The mismeasurement of self-efficacy in research has been previously documented. Schunk and Pajares (2009) have pointed out that researchers contribute to the problem by assessing self-efficacy at levels that are too specific or not specific enough, by using items closer in nature to self-concept than to self-efficacy and by using assessments that are not consistent with criterion tasks. However, Bandura (2006) has described procedures that are likely to increase the correspondence between measures of self-efficacy and the outcome(s) of interest to a researcher as reflected in a study's intervention. Items in a self-efficacy measure should be tailored to fit the domain of skills, knowledge, or behaviors under investigation and correspond to the range of task demands inherent in that domain. To the best of our knowledge, however, no one has compared across studies the methods researchers have used to demonstrate that any instrument or features present in an instrument affect outcomes differently.

Additionally, instruments developed and used to measure readers' self-perceptions have not always been grounded in deep theorizing about constructs under investigation, be they self-efficacy, self-concept, or other self constructs. Furthermore, some measurement instruments have titles that obscure what the instrument may actually measure. For example, the *Motivation for Reading Questionnaire* (MRQ; Wigfield & Guthrie, 1997) includes items said to measure reading self-efficacy, and the *Motivation to Read Profile* (MRP; Gambrell, Palmer, Codling, & Mazzone, 1996) includes items said to measure self-concept as a reader, but these items are identical. In the MRQ, students indicate on a scale from 1 to 4 the degree to which a statement is very different from them to a lot

like them, such as “I am a good reader.” On the MRP, one of four options that students can select following the stem “I am _____” is “a good reader.” The other options include “a poor reader,” “an OK reader,” and “a very good reader.” Because instruments like the MRP that purport to measure self-concept as a reader include items very similar to those items appearing in instruments reported to measure reading self-efficacy, attention needs to be given to individual items in an instrument and the constructs they measure rather than to the aspect of self they are said to measure.

Background on Interventions to Influence Reading Self-Efficacy

Motivation drives learning and reading. Motivated readers read more and more effectively. Researchers (Unrau & Schlackman, 2006; Wang & Guthrie, 2004) have found high correlations between motivation for reading, amount of reading, and measures of reading achievement. Intrinsic motivation, such as that arising from self-efficacy, propels deeper reading and comprehension (Vansteenkiste, Lens, & Deci, 2006). Unfortunately, researchers have repeatedly found that students’ motivation in general declines as they progress from elementary school through high school (Eccles, Lord, & Buchanan, 1996). That deterioration also characterizes motivation for reading from Grades 4 through the high school years (Gottfried, Fleming, & Gottfried, 2001; Unrau & Schlackman, 2006). Knowing that motivation for reading is fundamental to students’ development as readers, we could benefit from understanding how to sustain or expand students’ motivation for reading and reading self-efficacy.

The focus of the present analysis was on the impact of interventions designed to influence reading self-efficacy. Three types of studies with interventions were included: interventions with treatment and control groups, interventions with treatment and comparison groups, and interventions with a pre–post research design but no control group.

Interventions to influence self-efficacy beliefs could target several major sources for self-efficacy simultaneously, target a single source, or not target any major source. For example, one of the more common interventions (Schunk & Rice, 1991) involved modeling reading strategies by an expert (i.e., vicarious experience) and providing feedback on progress toward goals (i.e., persuasion). Interventions that integrate multiple sources of self-efficacy beliefs may be more effective than those that focus on one source alone (Souvignier & Mokhlesgerami, 2006).

A significant number of studies have been conducted that measure the impact of interventions on reading self-efficacy. Many of these studies have also shown a relationship between reading self-efficacy and reading comprehension. In fact, several of them were designed primarily to measure the impact of their intervention on reading comprehension with reading self-efficacy as a parallel concern of importance. However, to the best of our knowledge, no attempt has been made to synthesize studies with different interventions in order to gain a deeper understanding of the average gains in reading and reading self-efficacy.

Previous Reviews of Self-Efficacy: Correlational and Causal Approaches

Earlier meta-analyses have been conducted to investigate the relationship between self-efficacy and a variety of educational variables, although not reading.

Holden et al. (1990) conducted one of the earliest meta-analyses examining the relationship of self-efficacy to subsequent behavior in children under 16 years of age and found a mean effect size of 0.334 across the 26 studies included in their analysis. Multon et al. (1991) found an average correlation between self-efficacy beliefs and academic performance of .38 when analyzing 36 studies. Richardson et al. (2012) reviewed research on the antecedents of university students' GPA and found that both academic self-efficacy and performance self-efficacy had medium-sized and strong correlations with GPA, respectively. Prior meta-analyses investigating self-efficacy have indicated small to medium-size correlations and medium-size effect sizes with diverse constructs, such as academic performance and procrastination (Steel, 2007). Most of these meta-analyses have utilized correlational studies. However, none of these earlier meta-analyses examined the impact of various interventions on reading self-efficacy specifically, indicating a need for the present research.

Only one meta-analysis included the investigation of the impact of a single type of intervention on self-efficacy, namely, Concept-Oriented Reading Instruction (CORI). Guthrie et al. (2007) found that five studies, in which CORI was compared to other treatments, generated a mean self-efficacy effect size of 0.49. A second meta-analysis involving motivational interventions that also merits review here (Lazowski & Hulleman, 2015) found that interventions focused on motivation generally had a moderate impact on students' performance, behavior, or motivation with an average effect size of $d = 0.49$ (95% confidence interval [CI] = [0.43, 0.56]). Although Lazowski and Hulleman (2015) investigated the impact of interventions based on 15 motivational theories, including self-efficacy, they did not find any studies whose interventions were based solely on self-efficacy theory. In summary, our knowledge about the impact of interventions on self-efficacy is quite limited, a limitation our synthesis is designed to address.

Current Systematic Review

Researchers have found that self-efficacy serves as a motivating engine in specific domains, such as reading, that self-efficacy beliefs are malleable, and that reading self-efficacy specifically can be influenced through well-designed interventions. Our knowledge of the overall impact of interventions on reading self-efficacy based on systematic reviews is very limited. Other gaps in our knowledge remain. We do not know what features of these interventions, such as the number or type of major sources in the interventions shaping self-efficacy, have the most impact on reading self-efficacy. Nor do we know if other study characteristics make a difference in intervention outcomes, characteristics such as a study's design, grade level of subjects, prior reading performance of subjects, random assignment of subjects to experimental and control groups, an intervention's fidelity of implementation, and whether or not a study was published. We also have a limited understanding of the relationship between reading self-efficacy and reading comprehension, a limitation we also address.

Through this meta-analysis, answers to the following questions will be generated to reduce many of the existing gaps in our knowledge:

1. What is the magnitude of the impact of intervention studies that target reading self-efficacy as an outcome?
2. Does the type of intervention used in a study, more specifically its number of major sources affecting self-efficacy, influence reading self-efficacy as an outcome?
3. Which moderators, other than type of intervention, such as grade level, prior reading performance, study design, fidelity of implementation, and quality of self-efficacy measurement instrument, significantly affect reading self-efficacy outcomes?
4. For those studies that included measures of reading comprehension, what relationship, if any, exists between posttreatment self-efficacy measures and posttreatment reading comprehension?

With answers to these questions, we could increase the likelihood of enhancing reading self-efficacy, as well as reading comprehension, through the development and implementation of more powerful, optimum interventions. Currently, there are no meta-analytic reviews of reading self-efficacy interventions and limited information regarding what factors might influence the effectiveness of those interventions.

Methodology

Study Search and Identification

The purpose of this project was to review intervention studies in reading development with specific attention to reading self-efficacy as an outcome. To be included in the analysis, each study had to (a) be an intervention; (b) address reading specifically; (c) focus on students' reading self-efficacy, as a primary or secondary target; (d) provide sufficient information to calculate an effect size; (e) include explicit pre- and postintervention measures of reading self-efficacy; (f) be published in a peer-reviewed journal, as a doctoral dissertation, or in the National Reading Conference (NRC) or Literacy Research Association (LRA) Yearbooks; and (g) be published within the timeframe of 1980 to 2015. Studies that were duplicates (e.g., dissertations that were published in peer-reviewed journals) were excluded.

We searched keyword, title, abstract, and heading using the following words in combination: *intervention*, and *reading*, and *self-efficacy*. We searched the following online databases: PsycINFO, ERIC, PsycARTICLES, PsycCRITIQUES, MLA International Bibliography, Dissertations & Theses @USC, Proquest Dissertation & Theses Full Text, Proquest Dissertations and Theses A&I: The Humanities and Social Sciences Collections, Proquest Dissertations & Theses: UK & Ireland, annotated bibliographies of *Research in the Teacher of English* (RTE, 1990–2012). In addition, reference sections of key articles and studies as well as the AERA (American Education Research Association) online repository were hand-searched.

Applying Study Inclusion and Exclusion Criteria

The first three authors as a group reviewed each article and dissertation to determine if the study met the inclusion criteria. When we agreed as a group that

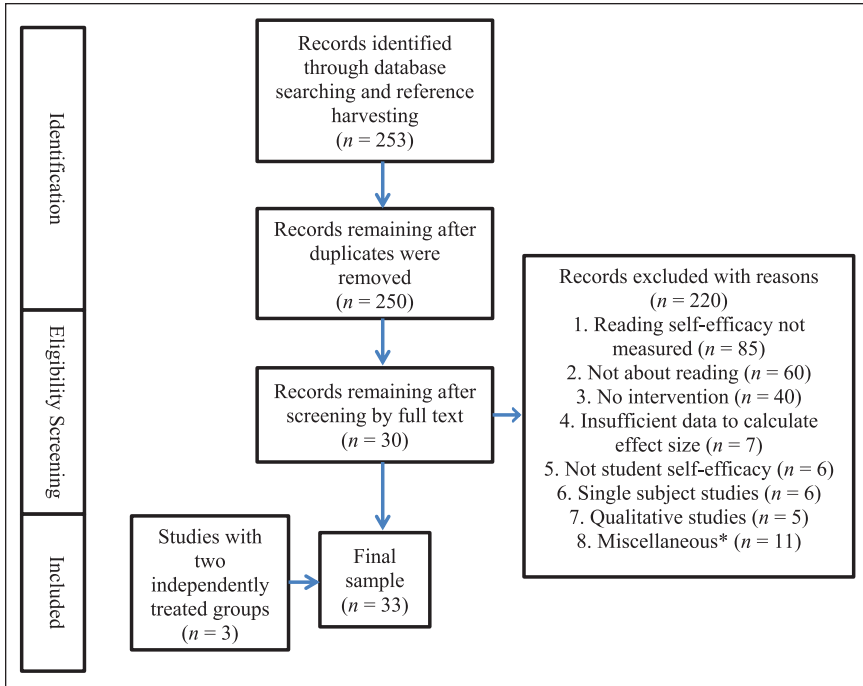


FIGURE 1. *Flow of study selection through different phases of review.*

*Miscellaneous excluded records that included correlational studies, studies that were not research oriented, studies that did not use students as subjects, studies that could not be located, and book reviews.

a study could be used in our pool, we added that study to our list of eligible studies. If a study did not meet our criteria for inclusion, we recorded our rationale for its exclusion (see Figure 1). More specifically, after a study passed the abstract screening stage, we reviewed items that made up the study’s measurement instrument, including any items that measured self-perceptions other than self-efficacy. For a study to be included, a consensus was reached.

In the process of selecting studies that would be included in our analysis, we found that researchers occasionally used instruments said to measure self-concept even though items in these instruments measured self-efficacy. As Pajares (1996) has observed, at domain-specific levels of generality, self-concept and self-efficacy beliefs may be empirically similar. However, the construct of self-efficacy that grounded the present analyses was that articulated by Bandura (1997, 2006). To deal with inconsistencies, we decided to include studies in our pool if they explicitly referred to reading self-efficacy as a measured variable of central importance to the study or if they provided theoretical grounding in self-efficacy and if items in the instrument used to measure reading self-perceptions covered self-efficacy.

We induced eight categories of exclusion and provide exemplars of the kinds of manuscripts that were excluded for particular categories. The largest excluded

category consisted of studies in which reading self-efficacy was not measured. It is exemplified by a study whose author (Ballard, 2007) presented self-efficacy as central to the study's theoretical framework and decided to use the term self-confidence instead of self-efficacy because the author believed that a suitable measure of self-efficacy could not be located for the age range of the subjects in the study. A reading attitude survey with items such as "How do you feel about going to a bookstore?" was used to measure self-confidence. We could not identify any items in the survey that measured reading self-efficacy based on Bandura's definition, and therefore, the study was excluded. Sixty studies were excluded because they measured self-efficacy in a domain other than reading, such as electronic information searching (Ren, 2000). On closer examination, 40 studies, such as Voorhees (2011), did not include an intervention. In seven instances, we could not extract sufficient self-efficacy data to calculate an effect size, as occurred when reading self-efficacy was integrated with other variables to form a composite measure, such as intrinsic motivation to read (Guthrie, Van Meter, McCann, & Wigfield, 1996). In six studies, we found that self-efficacy of subjects other than that of students was measured, as occurred in a study that measured mothers' self-efficacy (Horowitz, 2004). Six studies turned out to be single subject studies, such as that of Chandler (2012), which were excluded, as were five qualitative studies, such as that of Martin (2010). In contrast to these problematic measurements of self-efficacy, there are those that presented measurements with items that included specific passages for students with specific questions and asked students to rate their capacity to answer questions like those being asked (Schunk & Rice, 1989, 1991, 1993).

Coding

We developed two forms and procedures for the extraction and coding of data relevant to our meta-analysis that coders could follow. To develop the coding forms, we began with a "standardized" coding format (Lipsey & Wilson, 2001), and as we progressed through initial reviews of the studies and through several iterations of the forms, we selected and edited items to match those deemed essential to the parameters of our meta-analysis. One of the coding forms was for quantitative self-efficacy intervention studies with a control or a comparison group, and the other was for quantitative self-efficacy intervention studies using within-group pre-post measurement and no control group. These two coding forms were needed because of major differences between the two categories of studies, including data essential for calculation of effect sizes.

After drafting the two coding forms, we tested them to confirm that they would be thorough, sufficient, and useable by our coders. While testing the two coding forms with actual studies, we made further refinements to each form. The coding forms went through more than a dozen reviews before being deemed ready for use by two independent coders. The coders, who had recently completed their doctorates in education, were trained in the use of the two coding forms and given opportunities to read and code sample studies. Any items on either form that required further explanation or clarification were subsequently discussed. The coders then worked independently using the coding forms to code data from each study to be used in the meta-analysis. Several discrepancies in coding specific items for

studies arose. To address these discrepancies, we developed a document listing problematic coding and asked the coders to revisit specific items on the coding sheet for several studies. For example, an item on the coding sheet asked the coder to report the total number of sessions for the intervention condition including assessment (if given). The coders reported different totals. After revisiting the studies independently and reporting recalculations, the coders' discrepancies were resolved. The coders' agreement rate, which is found by dividing the number of observations agreed upon by the total number of observations, was 90%.

Effect Size Calculation

We used Hedges' *g*, a standardized mean difference between two groups, as an index of effect size for our meta-analysis, based on the index's capacity to correct and reduce the bias arising from small sample sizes (Glass, McGaw, & Smith, 1981; Hedges, 1981). Our meta-analysis included 20 studies that had treatment sample sizes under 40. Separate effect sizes were calculated for each of the three categories of studies in our sample.

The three categories were based on the type of study design the researcher(s) used: (a) treatment–control, (b) treatment–comparison, and (c) within-group pre–post. The effect sizes of these different study designs reflect the contrast made within each study design because of its inherent characteristics. Treatment–control effect sizes reflect what would occur to control group subjects should they receive the treatment provided. Treatment–comparison effect sizes reflect what would happen to the comparison group of subjects if they were to receive a comparable form of treatment. Pretest–posttest effect sizes reflect their own inherent qualities. For a study to fall into the treatment–control category, researchers had to have described and used a distinct control group. Subjects in these control groups received a traditional or standard curriculum. For studies in the treatment–comparison category, researchers identified and administered a comparison treatment against which the primary treatment was contrasted. Subjects in these comparison groups did not receive a traditional or standard curriculum but a different treatment from that which the researchers considered the primary focus of their inquiry. The primary treatment groups were commonly identified as the group hypothesized to outperform a designated comparison treatment group. For the third type of effect size, pretest–posttest designs, the same group of students was measured prior to and after the intervention was completed.

The procedures used to calculate effect sizes varied depending on the three types of studies. For two-group designs, Hedges' *g* was calculated by subtracting the mean of the control group (or in some cases, the appropriate comparison condition) from the mean of the treatment condition and dividing the difference by the average of the two groups' standard deviations (Lipsey & Wilson, 2001). When a study failed to provide means and standard deviations, we used a given *p* value or *t* statistic to estimate the effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). A positive *g* resulting from the calculation signified that subjects receiving a specified treatment gained more in reading self-efficacy than subjects in a control or comparison group.

For pretest–posttest designs, we drew on Dunlap, Cortina, Vaslow, and Burke (1996), who found that the size of the pretest–posttest correlation can affect the

effect size and, therefore, should be included in the calculation. Although none of the studies in our pool of studies included pre–post correlations, we obtained data from four researchers (Chirchick, 2009; Gavigan, 2010; McCrudden, Perkins, & Putney, 2005; Taboada Barber, et al., 2015). That enabled us to calculate an average pretest–posttest correlation ($r = .70$) that was used across other studies for which pre–post correlations were unobtainable. We tested for difference in effect sizes by running sensitivity analyses with alternative correlations (e.g., .60 and .90) to confirm that the various estimates did not result in significant differences. Hedges' g was calculated by subtracting the mean of the pretreatment scores from that of the posttreatment scores and dividing the difference by the average of the pre- and posttreatment standard deviations (Lipsey & Wilson, 2001). A positive g resulting from the calculation signified that subjects' level of reading self-efficacy increased as a result of the intervention they received.

Data Analyses

We used Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2006) software to run the data analyses. Two synthesis models are available. Fixed-effects models are based on the assumption that one true effect is inherent in all the studies included in the meta-analysis, where random-effects models assume that more than one true effect exists and that the effect sizes emerging from an analysis have arisen from a population of effects with varying values (Borenstein et al., 2009). We hypothesized that the effect sizes derived from a population of effects rather than one true effect and, therefore, chose to use and report a random-effects model. All effect sizes were weighted by inverse variance based on sample sizes (Borenstein et al., 2009). For each type of effect size (treatment–control, treatment–comparison, and pretest–posttest) and for all design types combined, we calculated an independent estimation of the average effect size based on interventions in each study.

Following analysis of the three separate categories, we combined all study designs into a common group because researchers who designed each and all included studies had a common stated purpose: to evaluate the impact of a treatment on subjects' reading self-efficacy. The combination of all study designs also provided us with a sensitivity test to compare each category of study design and its moderators against the combined set of all studies (Borenstein et al., 2009). However, the findings from these analyses should be viewed as exploratory.

Moderators

To understand the scope of the intervention effects and key variables that were important in our research and review of the literature, we conducted a series of moderator analyses that explored effect size heterogeneity. Moderators are variables in a study that are likely to have an impact on the outcome of the effect size calculations, and the moderators we focused on are described in the paragraphs below. Except for the Duration of Treatment Index and the effect sizes of reading comprehension outcomes that were continuous variables, these moderators were all categorical. We conducted a test of effect size heterogeneity using the Q formulas provided by Borenstein et al. (2009). Comprehensive Meta-Analysis calculated the Q statistic and its p value, as well as Hedges' g and confidence

intervals for categorical variables included in moderator analyses. Moderators included (a) grade level of subjects (e.g., elementary or other); (b) reading performance of the sample prior to treatment (e.g., struggling readers, non-struggling readers, or mixed); (c) number of major sources in an intervention shaping reading self-efficacy; (d) type of major source in an intervention shaping self-efficacy; (e) Self-Efficacy Measurement Index; (f) whether or not researchers addressed fidelity of implementation of intervention; (g) whether or not subjects were randomly assigned to intervention and control groups; and (h) publication bias. In sections below, the Self-Efficacy Measurement Index and the Duration of Treatment Index are described in greater detail. For continuously scaled moderators, we conducted meta-regression models. To understand the amount of variation explained by the moderators, we also calculated the meta-regression R^2 following formulas provided by Aloe, Becker, and Pigott (2010).

Duration of Treatment Index. Two items on the coding form that addressed duration of exposure to a treatment were combined to generate a “Duration of Treatment Index.” To generate that index, two items were multiplied: (a) total number of sessions focused on intervention only and (b) an indicator representing the length of each individual session. Indicators for session length ranged from 1 (for 0–30 minutes) to 2 (for 31–60 minutes) to 3 (more than 1 hour). Thus, if five 35-minute treatment sessions were given, the Duration of Treatment Index would be 10.

Reading Self-Efficacy Measurement Index. We decided to consider as most comprehensive and precise those measures of reading self-efficacy that used or reflected Bandura’s procedures for the development of self-efficacy measurement. The rationale for this choice arose from the opinion that Bandura’s work on self-efficacy measurement was theoretically, conceptually, empirically, and methodologically grounded (Bandura, 1986, 1997, 2006). These measures typically included specific reading tasks based on texts that were included in the measurement. Each item in the instrument focused on a specific task related to the text provided. After readers looked at the text and the task they would be asked to address, they indicated on a scale the degree to which they believed they could address the task successfully. Bandura (2006) recommended scales from 0 to 100 ranging in 10-unit intervals from “Cannot do” at 0 through “Moderately certain can do” at 50 to “Highly certain can do” at 100. Some instruments have used alternative scaling procedures that captured Bandura’s intention (Schunk & Rice, 1989). Although researchers (Gambrell et al., 1996; Wigfield, Guthrie, & McGough, 1996) have developed some questionnaires or subscales within questionnaires that measure reading self-efficacy or closely related aspects of reading self-perception, few have included task-specific and text-based items such as those recommended by Bandura.

A Self-Efficacy Measurement Index was developed based on Bandura’s recommendations for measuring self-efficacy to differentiate among the range of approaches manifested in the measurement of self-efficacy and to detect possible differences in their sensitivities. The index consisted of three items from the

coding forms: (a) Did the self-construct characteristic in the measurement address reading self-efficacy only or reading along with other literacy areas? (b) Did the self-construct characteristic in the measurement address only self-efficacy, other self-constructs only, or a mixture of self-efficacy and other self-constructs? (c) Was the level of specificity for reading self-efficacy reading task-specific without text, reading task-specific with text, or oriented to reading but not task-specific? The highest level of quality would be *reading only* for the first item, *self-efficacy* for the second item, and *reading task-specific with text* for level of specificity or a score of 3. The lowest quality rating for a reading self-efficacy measure was 0 because those measures did not focus only on reading, addressed self-constructs other than self-efficacy, and were not reading task specific with text provided.

Quality of study indicators. Although the development of a Quality of Study Index was explored, we decided to use discreet items related to quality on the coding sheet, namely, whether or not researchers addressed fidelity of implementation of intervention in their studies and whether or not researchers randomly assigned subjects to intervention and control or comparison groups.

Type of intervention. The type of intervention implemented in these studies was of much importance to us because information about interventions of the kind we collected has, to the best of our knowledge, not been collected or analyzed in the past. We drew on Bandura's description of four major sources shaping self-efficacy beliefs described earlier (Bandura, 1986, 1997).

In order to gather data to evaluate the moderating effect of these major sources shaping self-efficacy, the first three authors independently read and reviewed descriptions of interventions provided in each included study and, using a matrix that included the four sources shaping self-efficacy beliefs, indicated on the matrix if he or she believed that one or more of the sources manifested in the intervention description. We then tabulated the total number of sources evident in a study, a tabulation that could range from 0 to 4. If two or more of us confirmed that a particular source, such as enactive mastery experience, was manifested in the intervention as described in a study, those sources were tabulated. If only one of us acting as coder indicated that a source was manifested, that source was not included in the count. No study included any mention of physiological reactions. Therefore, the largest number of major sources shaping self-efficacy was three in the present set of included studies. The results of our tabulation are provided in Table 1. A brief summary of each study's intervention appears in Supplementary Table S2 (available in the online version of the journal).

Finally, we conducted two analyses to assess for potential publication bias. First, we conducted a moderator analysis that tested the difference in effect sizes between the published and unpublished studies (Polanin, Tanner-Smith, & Hennessy, 2016). Second, we assessed funnel plot asymmetry using Duval and Tweedie's (2001) trim and fill analysis. The results of this analysis indicate whether and how many effect sizes are potentially missing due to asymmetry. The procedure also estimates an average effect size based on the inclusion of the missing effect sizes. We treat this analysis as purely exploratory.

TABLE 1
Summary of intervention studies examining impact on reading self-efficacy

Author (date of publication)	Publication type	Grade	Students' reading proficiency	Sample size	Reading self-efficacy <i>g</i>	Reading comprehension <i>g</i>	Sources shaping self-efficacy identified by type
<i>Research design: Treatment and control group</i>							
Antoniou and Souvignier (2007)	J	Mixed	Struggling readers	T = 45; C = 28	0.04	0.60	M, V, P/F
Cantrell et al. (2014)	J	Middle	Struggling readers	T = 330; C = 281	0.16	0.04	M, V, P/F
Chirchick (2009)	D	Elementary (4th grade)	English learners	T = 14; C = 14	-0.36	-0.28	M
Chirchick (2009)	D	Elementary (5th grade)	English learners	T = 12; C = 10	0.20	-0.28	M
Guthrie, Klauda, and Ho (2014)	J	Middle	Non-struggling	T = 588; C = 587	0.06	-0.09	M, P/F
Hendry et al. (2005)	J	College	Non-struggling	T = 60; C = 66	0.04	N/A	0
Khajavi and Ketabi (2012)	J	College	Non-struggling	T = 30; C = 30	0.54	0.44	M, V, P/F
Lancaster (2011)	D	College	Struggling	T = 68; C = 103	0.15	N/A	M, P/F
Mihandoost, Elias, Nor, and Mahmud (2011)	J	Elementary	Struggling	T = 32; C = 32	0.55	N/A	M
Schunk and Rice (1987)	J	Elementary	Struggling	T = 10; C = 10	2.20	1.21	M, V, P/F
Schunk and Rice (1989)	J	Elementary	Struggling	T = 11; C = 11	1.20	1.26	M, V, P/F
Swan (1998)	D	Elementary	Mixed	T = 42; C = 44	0.12	0.35	0
Taboada Barber et al. (2014)—Grade 6	J	Middle	Mixed	T = 327; C = 96	0.26	N/A	M, V
Taboada Barber et al. (2014)—Grade 7	J	Middle	Mixed	T = 89; C = 99	0.35	N/A	M, V

(continued)

TABLE 1 (continued)

Author (date of publication)	Publication type	Grade	Students' reading proficiency	Sample size	Reading self-efficacy <i>g</i>	Reading comprehension <i>g</i>	Sources shaping self-efficacy identified by type
<i>Research design: Treatment and comparison group</i>							
Burrows (2012)	D	College	Non-struggling	T = 79; Comp = 79	0.49	N/A	M, V, P/F
Chang and Ho (2009)	J	College	Non-struggling	T = 34; Comp = 28	2.06	0.44	0
Mason (2004)	J	Elementary	Struggling	T = 16; Comp = 16	0.30	1.14	M, V
Nelson and Manset-Williamson (2006)	J	Mixed	Struggling	T = 9; Comp = 11	-0.46	N/A	M, V
Nolan (2012)	D	Elementary	Struggling	T = 39; Comp = 78	-0.62	-1.87	M, V, P/F
Schunemann, Sporer, and Brunstein (2013)	J	Elementary	Non-struggling	T = 117 Comp = 62	0.28	0.34	M, V, P/F
Schunk and Rice (1991)	J	Elementary	Struggling	T = 10; Comp = 10	1.26	1.28	M, V, P/F
Schunk and Rice (1993)	J	Elementary	Struggling	T = 11; Comp = 11	2.73	2.56	M, V, P/F
Stekel (1983)	D	Elementary	Struggling	T = 15; Comp = 15	0.67	N/A	M, V, P/F
Walters-Parker (2006)	D	High school	Struggling	T = 43; Comp = 27	-0.19	-0.12	M, V
Wigfield et al. (2004)	J	Elem.	Non-struggling	T = 150; Comp = 200	0.00	N/A	M
Zayyad (2009)	D	Middle	Struggling	T = 9; Comp = 9	-0.24	0.41	M, V, P/F
<i>Research design: Within-group pre-post</i>							
Dohrman-Swain (1998)—Goal setting	D	Middle	Struggling	T = 9	0.16	0.28	P/F
Dohrman-Swain (1998)—Teacher feedback	D	Middle	Struggling	T = 11	0.38	0.45	M
Garfield (2000)	D	Elementary	Struggling	T = 14	0.82	N/A	M, V
Gavigan (2010)	D	Middle	Struggling	T = 4	0.22	N/A	0
McCrudden, Perkins, and Putney (2005)	J	Elementary	Mixed	T = 23	0.73	-0.08	M, V
Nelson (2008)	D	Middle	Struggling	T = 35	0.05	0.75	M, P/F
Wozniak (2010)	D	Elementary	Struggling	T = 14	0.47	N/A	M, V

Note. J = journal; D = dissertation; T = treatment; C = control; Comp = comparison; *g* = average weighted effect size; Mixed = elementary + middle schools; M = mastery experience; V = vicarious experience; P/F = persuasion and/or feedback.

Relationships Between Reading Self-Efficacy and Reading Comprehension

A total of 21 studies out of 33 included in the meta-analysis of reading self-efficacy also provided data on reading comprehension. To better understand the relationship that arose between reading self-efficacy and reading comprehension, we undertook the statistical exploration of that relationship. We conducted a meta-analysis of those 21 studies to discover the impact of interventions on reading comprehension. We also applied a meta-regression model (Borenstein et al., 2009) to this group of studies to test the moderating influence of reading comprehension on the reading self-efficacy effect size. We then constructed three categories of *g* scores reflecting the impact of interventions on reading comprehension: (a) negative effect size ($g < 0.000$), (b) low to modest effect size (g between 0.001 and 0.600), and (c) robust effect size ($g > 0.601$). We then used reading comprehension as reflected in these three categories as a moderator of reading self-efficacy. With consideration of membership in a particular study design category (i.e., treatment–control, treatment–comparison, or pretest–posttest), we also conducted a moderator analysis to determine the impact of membership in study design category on the reading comprehension effect size.

Results

Descriptive Analysis

The initial and extended search procedures generated 253 citations. Of these, 131 were journal articles and 122 were dissertations. Applying our inclusion criteria while reviewing these studies, the first three authors identified 30 articles and dissertations that met our criteria. Of the studies included, 18 were journal articles and 12 were dissertations (see Table 1). One journal article (Taboada Barber et al., 2015) and two dissertations (Chirchick, 2009; Dohrman-Swain, 1998) were resourced twice because each of those studies included data on two independent treatment groups. We maintained a list of all excluded studies and the rationale for their exclusion. Many studies were excluded because they were qualitative rather than quantitative in design, single subject design, not focused on reading self-efficacy but on broader self-constructs not related to reading, did not include a sufficient measure of reading self-efficacy, or lacked sufficient information to calculate an effect size.

Table 1 identifies the studies included, the type of publication, the grade level of subjects in the study (Elementary, Middle, Elementary and Middle, High School, or College), students' reading proficiency (struggling, non-struggling, or mixed), the number of subjects in the treatment and control groups, and the treatment's effect sizes for reading self-efficacy and for reading comprehension. The total number of students who received treatment was 2,300. An additional 1,411 and 546 students participated in the control or comparison groups, respectively. Most of the studies were conducted with elementary school level students ($n = 15$), the second most frequent level being middle school ($n = 11$). Information about gender was infrequently included in the study report, and therefore was not used in the analysis.

The 30 studies included in our analysis represent different research designs. The review included 12 two-group, treatment–control studies (40%); 12 two-group, treatment–comparison studies (40%); and six within-group, pre–post

studies (20%). Some treatment–control studies (Chirchick, 2009; Taboada Barber, et al., 2015) were designed so that the same treatment was administered to two separate, independent groups that enabled us to include effect sizes for each of the two distinct groups. One within-group, pre–post study (Dohrman-Swain, 1998) was designed so that two different treatments were administered to two independent groups, so that enabled us to include effect sizes for each of the two treatments. In all, the meta-analysis included 33 effect sizes.

We calculated several indices using study-reported information. The Duration of Treatment Index ranged from 6 (3 sessions \times 2 [between 30–60 minutes]) (Stekel, 1983) to 450 (150 sessions \times 3 [more than 1 hour]) (Garfield, 2000). The average Duration of Treatment Index was 79. A Self-Efficacy Measurement Index for the quality of reading self-efficacy measurement implemented in each study was also generated. That Index included three items. Of the studies included in the analysis, five studies received the highest index (3), nine studies received the middle index score (2), and 16 studies received the lowest index (1).

The experimental treatments provided for subjects in these studies varied widely. Some treatments were those well known in the field, such as Sustained Silent Reading (Walters-Parker, 2006), READ 180 (Nelson, 2008), Reciprocal Teaching (Schunemann, Sporer, & Brunstein, 2013), Concept Mapping (Khajavi & Ketabi, 2012), and CORI (Guthrie, Klauda, & Ho, 2014). Other treatments included a potpourri of reading strategies (Antoniou & Souvignier, 2007; McCrudden, Perkins, & Putney, 2005; Schunk & Rice, 1987, 1989, 1993). Several treatments utilized feedback (Antoniou & Souvignier, 2007; Cantrell et al., 2014; Schunk & Rice, 1991) and/or self-regulation (Dohrman-Swain, 1998; Lancaster, 2011; Mason, 2004; Nolan, 2012).

As described earlier, we categorized each study’s intervention according to the number of major sources that shape self-efficacy beliefs, according to Bandura (1986, 1997). Of the total set of studies, 12 study interventions included three major sources shaping self-efficacy beliefs. Twelve study interventions included two sources; six included one; and three had none according to our understanding of the intervention described in those studies.

Treatment–Control Design Results

Twelve studies were included in the treatment–control study design analysis with two studies offering two independent treatments, thereby providing 14 treatments that enabled us to estimate a treatment–control effect size (see Table 2). The results of the weighted average applying a random model (Borenstein et al., 2009) indicated a small but statistically significant positive increase in students’ reading self-efficacy ($g = 0.24$, 95% confidence interval [CI] [0.10, 0.39], $p = .001$) following treatment. A small amount of variability remained that was statistically significant ($Q = 33.44$, $p = .001$, $I^2 = 61.12$). The significance of this heterogeneity suggests that variance among the studies could be attributable to factors other than random error. It should be noted that, to test robustness of these results, we also calculated effect sizes whereby we included pretest data. When not available, we imputed the pretest–posttest correlation using a value of 0.70. The results indicated very little difference from the results using posttest data only, and therefore, we used the posttest-only results throughout the remainder of the analyses.

TABLE 2*Overall synthesis results for three types of effect sizes and combined*

Type of design	<i>k</i>	<i>g</i>	95% CI	<i>Q</i> -between (<i>p</i>)	<i>I</i> ² , <i>τ</i> ²
Treatment–control	14	0.24	[0.10, 0.39]	33.44 (.001)	61.12, .04
Treatment–comparison	12	0.44	[0.04, 0.84]	90.57 (.001)	87.85, .39
Pretest–posttest	7	0.36	[0.16, 0.57]	14.20 (.027)	57.74, .04
All designs combined	33	0.33	[0.19, 0.46]	141.47 (.001)	77.38, .10

Note. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval. *I*² and *τ*² represent measures of effect size variability.

Given a significant amount of between-study heterogeneity, we conducted moderator analyses to evaluate variables that could affect effect size outcomes (see Table 3). Eight categorical moderator analyses were tested. The Reading Self-efficacy Measurement Index, when comparing the three levels of measurement quality, significantly moderated effect size ($Q_B = 11.06, p = .004$). The lowest index quality ($n = 8$) generated the lowest effect size ($g = 0.10, 95\% \text{ CI } [0.06, 0.18]$); the middle index quality ($n = 5$) produced a larger effect size ($g = 0.44, 95\% \text{ CI } [0.10, 0.78]$); and the highest index quality ($n = 1$) yielded the highest effect size ($g = 1.20, 95\% \text{ CI } [0.43, 1.96]$).

We gathered data on whether or not fidelity of implementation and randomization were integrated into the study's design and execution. In some studies, no mention of either was manifested. In those instances, we assumed they were not part of the study design or were not executed. Fidelity of implementation, considered as either addressed or not addressed in these studies, approached significance but did not significantly moderate the effect size ($Q_B = 3.42, p = .065$). Those studies that did not address fidelity of implementation yielded the higher effect size ($g = 0.50, 95\% \text{ CI } [0.13, 0.86]$), whereas those studies addressing fidelity of implementation produced a lower effect size ($g = 0.14, 95\% \text{ CI } [0.05, 0.24]$). Random assignment of subjects to intervention and control groups did not have a significant moderating impact on effect size ($Q_B = 0.01, p = .962$).

Several additional variables tested did not significantly moderate the effect size: (a) grade level, when comparing elementary to all others (middle, high school, college, and mixed) ($Q_B = 2.08, p = .150$); (b) prior reading performance level of the sample (struggling, non-struggling, or mixed) ($Q_B = 1.96, p = .375$); (c) whether a study was peer reviewed or not ($Q_B = 2.13, p = .144$); and (d) number of major sources shaping self-efficacy beliefs (Bandura, 1986, 1997) that were manifested in the study's intervention ($Q_B = 3.85, p = .279$). However, while the total number of sources did not significantly moderate effect size, vicarious experience, as a major source shaping self-efficacy, was a significant moderator ($Q_B = 5.86, p = .015$).

We applied a meta-regression model to test the moderating effects of the Duration of Treatment on effect size. For the studies included in the treatment–control category, length of treatment had no statistically significant relationship with the effect size ($k = 14, \beta = -0.0001, 95\% \text{ CI } [-0.001, 0.001], p = .911$).

TABLE 3*Treatment–control effect size moderator analysis*

Measure	Level	<i>Q</i> -between (<i>p</i>)	<i>k</i>	<i>g</i>	95% CI	
Grade level	Elementary	2.08 (.150)	6	0.56	[0.01, 1.12]	
	All others ^a		8	0.15	[0.06, 0.24]	
Prior reading performance of sample	Struggling	1.96 (.375)	8	0.37	[0.05, 0.68]	
	Non-struggling		3	0.12	[-0.09, 0.33]	
	Mixed		3	0.27	[0.10, 0.43]	
Number of major sources in intervention shaping self-efficacy		3.85 (.279)	3	0.64	[0.13, 1.15]	
			2	0.17	[0.03, 0.31]	
			1	0.18	[-0.38, 0.73]	
			0	2	0.07	[-0.20, 0.34]
Major source in intervention shaping self-efficacy	Mastery	1.72 (.189)				
	Yes		12	0.28	[0.11, 0.45]	
	No	2	0.07	[-0.20, 0.34]		
	Vicarious	5.86 (.015)				
	Yes		7	0.42	[0.16, 0.69]	
	No	7	0.08	[-0.02, 0.18]		
	Persuasion/feedback	0.469 (.493)				
	Yes		7	0.32	[0.08, 0.57]	
	No	7	0.23	[0.08, 0.37]		
	Reading Self-Efficacy Measurement Index		11.06 (.004)			
Lowest index quality		8		0.10	[0.06, 0.18]	
Middle index quality		5		0.44	[0.10, 0.78]	
Highest index quality		1		1.20	[0.43, 1.96]	
Quality of study: Fidelity of implementation addressed		3.42 (.065)				
	Yes		7	0.14	[0.05, 0.24]	
	No	7	0.50	[0.13, 0.86]		
Quality of study: Random assignment to intervention and control		0.01 (.962)				
	Yes		11	0.26	[0.06, 0.46]	
	No	3	0.26	[0.10, 0.41]		
Published		2.13 (.144)				
	Peer reviewed		10	0.31	[0.13, 0.49]	
	Not peer reviewed	4	0.09	[-0.13, 0.32]		

Note. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval.

a. All others: combination of middle, high school, college, and mixed grade levels.

Treatment–Comparison Design Results

Twelve studies were included in the treatment–comparison study design analysis that provided data to estimate a treatment–comparison effect size (see Table 2). The results of the weighted average applying a random-effects model indicated a moderate, statistically significant positive increase in students’ reading self-efficacy ($g = 0.44$, 95% CI [0.04, 0.84], $p = .03$) following treatment. A moderate amount of variability remained that was statistically significant ($Q = 90.57$, $p = .001$, $I^2 = 87.85$). A sensitivity test using pretest and posttest treatment and control data entry that required pre–post correlations ($r = .70$) generated a very small difference in effect size outcome ($g = 0.45$).

Moderator analyses were conducted to evaluate the stability of the effect size estimates and to identify variables that had a significant impact on effect size outcomes. As shown in Table 4, eight categorical moderator analyses were tested for the treatment–comparison studies. The number of sources shaping self-efficacy beliefs that were present in a study’s intervention did have a significant moderating effect ($Q_B = 42.47$, $p = .001$). For those interventions having three major sources shaping self-efficacy ($n = 7$), the effect size was moderately high ($g = 0.53$, 95% CI [–0.03, 1.08]). However, for this moderator only, we reran the analysis after removing what appeared to be an outlier with an intervention that had no major sources shaping self-efficacy (Chang & Ho, 2009) but which produced a large effect size ($g = 2.06$, 95% CI [1.44, 2.67]). The results of the sensitivity test indicated that, with the removal of the outlier, the number of sources shaping self-efficacy beliefs that were present in a study’s intervention did not have a significant impact on moderating effect size ($Q_B = 3.62$, $p = .164$). As for the type of major source shaping self-efficacy, no major source was found to moderate estimated effect sizes.

As was the case for peer-reviewed treatment–control studies, peer-reviewed studies in this category also had an impact on effect size that approached significance ($Q_B = 3.50$, $p = .061$, $g = 0.80$, 95% CI [0.17, 1.42]). Fidelity of implementation did not significantly moderate the effect size ($Q_B = 2.53$, $p = .112$). As with treatment–control studies, treatment–comparison studies that did not address fidelity of implementation yielded the higher effect size ($g = 0.84$, 95% CI [0.18, 1.51]), whereas those studies that addressed fidelity of implementation produced a lower effect size ($g = 0.15$, 95% CI [–0.39, 0.69]). Random assignment to intervention and comparison groups did not significantly moderate effect size ($Q_B = 0.83$, $p = .362$).

Several other variables in this category of study design did not significantly moderate effect size: (a) grade level, when comparing elementary to all other grade levels ($Q_B = 0.05$, $p = .817$); (b) prior reading performance level of the sample ($Q_B = 0.46$, $p = .498$); and (c) the Reading Self-efficacy Measurement Index, when comparing the three levels of measurement quality ($Q_B = 0.80$, $p = .671$). As was done with the treatment–control category, we applied a meta-regression model to this group of treatment–comparison studies to test the moderating effects of the Duration of Treatment and found that it did not have a statistically significant relationship with the effect size ($k = 12$, $\beta = -0.0066$, 95% CI [–0.0162, 0.0030], $p = .176$).

TABLE 4*Treatment–comparison effect size moderator analysis*

Measure	Level	<i>Q</i> -between (<i>p</i>)	<i>k</i>	<i>g</i>	95% CI
Grade level	Elementary	0.05 (.817)	7	0.47	[−0.02, 0.95]
	All others ^a		5	0.36	[−0.44, 1.15]
Reading performance	Struggling	0.46 (.498)	8	0.35	[−0.23, 0.96]
	Non-struggling		4	0.64	[0.05, 1.23]
Number of major sources in intervention shaping self-efficacy		42.47 (.001)			
	3		7	0.53	[−0.03, 1.08]
	2		3	−0.10	[−0.48, 0.28]
	1		1	0.00	[−0.21, 0.21]
Major source in intervention shaping self-efficacy	0		1	2.06	[1.44, 2.67]
	Mastery	25.34 (.001)			
	Yes		11	0.26	[−0.78, 0.60]
	No	1	2.06	[1.44, 2.67]	
	Vicarious	0.43 (.513)			
	Yes		10	0.32	[−0.10, 0.74]
	No	2	1.01	[−1.01, 3.02]	
	Persuasion/feed-back	0.15 (.701)			
Yes	7		0.53	[−0.390, 1.08]	
No	5	0.35	[−0.21, 0.21]		
Reading Self-Efficacy Measurement Index		0.80 (.671)			
	Lowest index quality		5	0.28	[−0.42, 0.98]
	Middle index quality		4	0.37	[0.15, 0.59]
	Highest index quality		3	1.15	[−0.63, 2.92]
Quality of study: Fidelity of implementation addressed		2.53 (.112)			
	Yes		7	0.15	[−0.39, 0.69]
No	5	0.84	[0.18, 1.51]		
Quality of study: Random assignment to intervention and control		0.83 (.362)			
	Yes		5	0.73	[−0.05, 1.50]
No	7	0.29	[−0.23, 0.81]		
Published		3.50 (.061)			
	Peer reviewed		7	0.80	[0.17, 1.42]
Not peer reviewed	5	0.02	[−0.51, 0.55]		

Note. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval.

a. All others: combination of middle, high school, college, and mixed grade levels.

Within-Group Pretest–Posttest Design Results

Six studies were included in this subset analysis with one study offering two independent treatments, thereby providing seven treatments that enabled us to estimate a pretest–posttest effect size (Table 2). The results of the weighted average indicated a statistically significant, small-to-moderate positive increase in students' reading self-efficacy ($g = 0.36$, 95% CI [0.16, 0.57], $p = .001$) following treatment. A small amount of variability remained that reached statistical significance ($Q = 14.20$, $p = .027$, $I^2 = 57.74$). An implication of this finding is that an unexplained variable or system of variables had an impact on the dispersion. Sensitivity analyses that entailed the modification of the pretest–posttest correlations showed slight impact on the average effect size ($g = 0.34$, 95% CI [0.16, 0.60], $p = .001$). We used the original results throughout this analysis because the difference was small.

Moderator analyses were again conducted to evaluate the stability of the effect size estimates and to identify variables that had a significant impact on effect size outcomes. As shown in Table 5, seven categorical moderator analyses were tested. Grade level, when comparing elementary to all other grade levels, significantly moderated the effect size ($Q_B = 10.76$, $p = .001$). Interventions in studies whose subjects were in elementary school ($n = 3$) yielded the higher effect size ($g = 0.66$, 95% CI [0.41, 0.91]), whereas those interventions in studies whose subjects were in higher grade levels ($n = 4$) produced a lower effect size ($g = 0.17$, 95% CI [0.03, 0.32]). Prior reading performance level of the sample did not significantly moderate effect size ($Q_B = 2.82$, $p = .093$) but approached significance. Interventions in studies whose subjects were a mixture of struggling and non-struggling readers ($n = 1$) generated a higher effect size ($g = 0.73$, 95% CI [0.28, 1.19]). Those interventions in studies whose subjects were struggling ($n = 6$) produced a lower effect size ($g = 0.31$, 95% CI [0.11, 0.51]). As major sources shaping reading self-efficacy, both vicarious experience ($Q_B = 10.76$, $p = .001$) and persuasion/feedback ($Q_B = 5.79$, $p = .016$) had a moderating influence on effect sizes. Whether a study was peer-reviewed or not also had an impact on effect size that approached significance ($Q_B = 2.82$, $p = .093$). However, only one study of the seven in this pre–post category was peer-reviewed ($g = 0.73$, 95% CI [0.28, 1.19]).

Additional variables in this category of study design did not significantly moderate effect size: (a) number of major sources shaping self-efficacy beliefs that appeared in a study's intervention ($Q_B = 1.44$, $p = .488$); (b) reading Self-efficacy Measurement Index ($Q_B = 3.171$, $p = .205$); and (c) fidelity of implementation ($Q_B = 1.84$, $p = .175$). For the studies included in this category, length of treatment had no statistically significant relationship with the effect size ($k = 7$, $\beta = 0.0001$, 95% CI [–0.0011, 0.0012], $p = .926$).

All Study Design Types Combined

Thirty-three studies with distinct samples were described in 30 of our included sources. The total number of studies differs from the number of sources because three sources from this group each described two separate studies with independent samples. These 33 studies provided data to estimate an all-study-designs-combined effect size as shown in Table 2. We present these results here simply as an exploratory analysis for the readers' benefit, and as we explained above, we

TABLE 5*Pretest–posttest effect size moderator analysis*

Measure	Level	<i>Q</i> -between (<i>p</i>)	<i>k</i>	<i>g</i>	95% CI
Grade level	Elementary	10.76 (.001)	3	0.66	[0.41, 0.91]
	All others ^a		4	0.17	[0.03, 0.32]
Reading performance	Struggling	2.82 (.093)	6	0.31	[0.11, 0.51]
	Mixed		1	0.73	[0.28, 1.19]
Number of major sources in intervention shaping self-efficacy		1.44 (.488)			
	2		4	0.49	[0.10, 0.88]
	1		2	0.28	[−0.04, 0.60]
	0		1	0.22	[0.10, 0.88]
Major source in intervention shaping self-efficacy	Mastery	1.85 (.173)			
	Yes		5	0.46	[0.15, 0.40]
	No	2	0.21	[0.02, 0.88]	
	Vicarious	10.76 (.001)			
	Yes		3	0.66	[0.41, 0.91]
	No	4	0.17	[0.03, 0.32]	
	Persuasion/feedback	5.79 (.016)			
	Yes		2	0.07	[−0.15, 0.30]
No	5	0.48	[0.23, 0.72]		
Reading Self-Efficacy Measurement Index		3.17 (.205)			
	Lowest index quality		5	0.28	[0.06, 0.51]
	Middle index quality		1	0.47	[0.06, 0.87]
	Highest index quality		1	0.73	[0.28, 1.19]
Quality of study: Fidelity of implementation addressed		1.84 (.175)			
	Yes		4	0.22	[0.01, 0.43]
	No		3	0.55	[0.12, 0.98]
Published		2.82 (.093)			
	Peer reviewed		1	0.73	[0.28, 1.19]
	Not peer reviewed	6	0.31	[0.11, 0.51]	

Note. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval.

a. All others: combination of middle, high school, college, and mixed grade levels.

caution readers about drawing conclusions based on these findings. The results of the weighted average applying a random-effects model indicated a moderate statistically significant positive increase in students' reading self-efficacy ($g = 0.33$, 95% CI [0.19, 0.46], $p = .001$) following treatment. A moderate amount

of variability remained that was also statistically significant ($Q = 141.47, p = .001, I^2 = 77.38$). This suggests that variance among the studies could be attributable to factors other than random error.

Eight categorical moderator analyses were conducted, as shown in Table 6. Study design (treatment–control, treatment–comparison, or within-group pretest–posttest) had no statistically significant impact on effect size ($Q_B = 1.39, p = .499$). However, the Reading Self-Efficacy Measurement Index, when comparing the three levels of measurement quality, significantly moderated effect size ($Q_B = 5.90, p = .052$). The lowest index quality ($n = 18$) generated the lowest effect size ($g = 0.19, 95\% \text{ CI } [0.03, 0.34]$); the middle index quality ($n = 10$) produced a larger effect size ($g = 0.39, 95\% \text{ CI } [0.20, 0.57]$); and the highest index quality ($n = 5$) yielded the highest effect size ($g = 1.03, 95\% \text{ CI } [0.23, 1.84]$). This finding, which will be explored further in our Discussion section, may carry a message to researchers about the importance of instruments that measure reading self-efficacy and the care that ought to be given to their construction and/or selection. Whether or not studies had been peer-reviewed also significantly moderated the effect size ($Q_B = 5.44, p = .020$). Those studies that were peer-reviewed produced the higher effect size ($g = 0.49, 95\% \text{ CI } [0.29, 0.69]$), whereas those not peer-reviewed produced the lowest ($g = 0.17, 95\% \text{ CI } [-0.01, 0.35]$). This result may reflect publication bias, a common phenomenon in educational research (Cheung & Slavin, 2016). Grade level, when comparing elementary to all others, had an impact on the effect size that approached significance ($Q_B = 3.20, p = .074$).

For the combination of all studies, the number of major sources shaping self-efficacy beliefs that appeared in a study's intervention ($Q_B = 4.51, p = .212$) did not significantly moderate effect size. However, as was done with the treatment–comparison analysis, we conducted a sensitivity analysis for this moderator only. We removed one study (Chang & Ho, 2009) that had no major source shaping self-efficacy according to Bandura's criteria and that appeared to be an outlier producing a large effect size ($g = 2.06, 95\% \text{ CI } [1.44, 2.67]$). The sensitivity test results showed that, with the removal of the outlier, the number of sources shaping self-efficacy beliefs that were present in a study's intervention had a significant impact on moderating effect size ($Q_B = 38.09, p = .001$). As shown in Table 6, those studies including an intervention with three major sources shaping self-efficacy ($n = 12$) generated the highest effect size ($g = 0.55, 95\% \text{ CI } [0.21, 0.88]$), those studies with two sources ($n = 11$) produced a more modest effect size ($g = 0.24, 95\% \text{ CI } [0.08, 0.40]$), and those studies ($n = 6$) having only one major source shaping self-efficacy yielded a smaller effect size ($g = 0.16, 95\% \text{ CI } [-0.06, 0.38]$). However, no single type of major source shaping self-efficacy reached significance in moderating estimated effect sizes.

Fidelity of implementation did significantly moderate the effect size ($Q_B = 9.81, p = .002$). Those studies that did address fidelity of implementation ($n = 18$) yielded the lower effect size ($g = 0.14, 95\% \text{ CI } [-0.01, 0.28]$), whereas those studies not addressing fidelity of implementation produced a higher effect size ($g = 0.60, 95\% \text{ CI } [0.35, 0.86]$). However, random assignment to intervention and comparison groups did not have a significant moderating impact on effect size ($Q_B = 0.30, p = .861$).

TABLE 6*All study design types combined*

Measure	Level	<i>Q</i> -between (<i>p</i>)	<i>k</i>	<i>g</i>	95% CI
Grade level	Elementary	3.20 (.074)	16	0.53	[0.23, 0.82]
	All others ^a		17	0.23	[0.09, 0.37]
Prior reading performance of sample	Struggling	0.26 (.878)	22	0.31	[0.12, 0.50]
	Non-struggling		7	0.40	[0.10, 0.72]
	Mixed		4	0.33	[0.14, 0.53]
Study design	Treatment–control	1.39 (.499)	14	0.24	[0.10, 0.39]
	Treatment–comparison		12	0.44	[0.04, 0.84]
	Within-group pre–post		7	0.36	[0.16, 0.57]
Number of major sources in intervention shaping self-efficacy		4.51 (.212)			
	3		12	0.55	[0.21, 0.88]
	2		11	0.24	[0.08, 0.40]
	1		6	0.16	[–0.06, 0.38]
Major source in intervention shaping self-efficacy	0		4	0.55	[–0.07, 1.17]
	Mastery	0.38 (.540)			
	Yes		28	0.30	[0.16, 0.45]
	No	5	0.47	[–0.03, 0.96]	
	Vicarious	1.91 (.167)			
	Yes		20	0.42	[0.21, 0.62]
	No	13	0.23	[0.05, 0.40]	
	Persuasion/feedback	0.01 (.940)			
	Yes		16	0.34	[0.13, 0.54]
	No	17	0.33	[0.14, 0.52]	
Reading Self-Efficacy Measurement Index		5.90 (.052)			
	Lowest index quality		18	0.19	[0.03, 0.34]
	Middle index quality		10	0.39	[0.20, 0.57]
Quality of study: Fidelity of implementation addressed	Highest index quality	9.81 (.002)	5	1.03	[0.23, 1.84]
	Yes		18	0.14	[–0.01, 0.28]
	No		15	0.60	[0.35, 0.86]
Quality of study: Random assignment to intervention and control		0.30 (.861)			
	Yes		16	0.36	[0.15, 0.57]
No	10	0.27	[–0.04, 0.57]		
Published		5.44 (.020)			
	Peer reviewed		18	0.49	[0.29, 0.69]
Not peer reviewed			15	0.17	[–0.01, 0.35]

Note. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval.

a. All others: combination of middle, high school, college, and mixed grade levels.

For the combination of all studies, prior reading performance level of the sample ($Q_B = 0.26, p = .878$) did not significantly moderate effect size. As was the case in the prior three subset analyses, length of treatment had no statistically significant relationship with the effect size ($k = 33, \beta = -0.0005, 95\% \text{ CI } [-0.0023, 0.0013], p = .588$).

Publication Bias Analysis

We conducted Duval and Tweedie's (2001) funnel plot analysis on all three groups of effect sizes. For the treatment-control and treatment-comparison groups of studies, the results indicated that zero effect sizes were missing. For the within-group pretest-posttest studies, the results indicated that three effect sizes were missing, and inclusion of the effect sizes in the random-effects model would decrease the effect ($d = 0.18, 95\% \text{ CI } [-0.04, 0.41]$). Finally, we combined all effect sizes and conducted an additional trim and fill analysis revealing that zero studies were missing. Taken together, we concluded that this is a robust set of studies, and publication bias was only a minor concern (see Supplementary Figures S1-S4, available in the online version of the journal).

The Impact of Interventions on Reading Comprehension and Its Moderating Influence on Self-Efficacy

Twenty-one of the 33 studies in this meta-analysis of interventions' impact on reading self-efficacy also included data on the impact those interventions had on reading performance. Understanding aspects of the relationship between reading self-efficacy and comprehension can deepen our grasp of effective reading processes and what drives them. As was done when running the meta-analysis of reading self-efficacy, we used Hedges' g as an indicator of effect size because of the number of studies using small samples. We also used a random model, assuming that these studies drew on different populations rather than a single population, as was done with the self-efficacy analysis. The results of applying a random model showed a small to moderate and statistically significant positive increase in students' reading performance following treatment ($g = 0.33, 95\% \text{ CI } [0.08, 0.58], p = .001$). A small amount of variability remained that was statistically significant ($Q = 162.80, p = .001, I^2 = 87.69$) and suggested that causes other than random error contributed to outcome. Study design (treatment-control, treatment-comparison, or within-group pretest-posttest) had no statistically significant impact on effect size ($Q_B = 0.50, p = .779$).

To explore the relationship between the reading self-efficacy and reading comprehension outcomes, we conducted a moderator analysis of reading self-efficacy using three categories of Hedges' g scores generated from a meta-analysis of the 21 studies that had reading comprehension data. As shown in Table 7, the three categories of g scores reflecting the impact of interventions on reading comprehension were as follows: (a) negative effect size ($g < 0.00$), (b) low to medium effect size (g between 0.01 and 0.60), and (c) medium to large effect size ($g > 0.60$). Reading comprehension, when applied to the three categories, significantly moderated reading self-efficacy effect size ($Q_B = 8.14, p = .017$). The negative effect size category of reading comprehension included the lowest self-efficacy effect size ($g = -0.03, 95\% \text{ CI } [-0.38, 0.32]$); the low to medium effect

TABLE 7*Reading comprehension as moderator affecting reading self-efficacy effect sizes*

Level ^a	<i>Q</i> -between (<i>p</i>)	<i>k</i>	<i>g</i>	95% CI
	8.14 (.017)			
Negative		6	-0.03	[-0.38, 0.32]
Low to modest		8	0.42	[0.09, 0.74]
Large		7	0.97	[0.34, 1.61]

Note. Twenty-one studies have been included in this analysis of combined designs because 12 studies did not provide data on reading comprehension outcomes. *k* = number of studies; *g* = average weighted effect size; CI = confidence interval.

a. Categories of Hedges' *g* scores based on impact of study interventions on reading comprehension: Negative = $g < 0.00$; Low to modest = $g = 0.01$ to 0.60 ; Large = $g > 0.60$.

size category produced a larger self-efficacy effect size ($g = 0.42$, 95% CI [0.09, 0.74]); and the medium to large effect size category included the highest self-efficacy effect size ($g = 0.97$, 95% CI [0.34, 1.61]).

We also applied a meta-regression model to this group of studies to test the moderating influence of reading comprehension on the reading self-efficacy effect size. For the studies included, reading comprehension had a statistically significant relationship with the reading self-efficacy effect size ($\beta = 0.65$, 95% CI [0.36, 0.94], $p = .001$). Furthermore, reading comprehension in this meta-regression model explains 54% of the variance in reading self-efficacy according to an index (R^2_{Meta}) and formula proposed by Aloe et al. (2010).

Together these analyses indicate not only a strong relationship between reading comprehension and reading self-efficacy but also suggest that higher levels of comprehension support higher levels of reading self-efficacy.

Discussion

This investigation was conducted to discover the effect of interventions on reading self-efficacy, the effect of various moderators on estimated effect sizes, and the relationships between reading self-efficacy and reading comprehension. We analyzed 18 published and 12 unpublished papers of interventions that included 33 measures of reading self-efficacy and accounted for a total of 4,257 participants. With respect to our first research question regarding the impact of interventions on reading self-efficacy, our investigation revealed that the magnitude of that impact was statistically significant regardless of study design. Those interventions that were included in treatment–comparison study designs generated the largest effect size, whereas those in the treatment–control category of study design generated the smallest effect size. The interventions of all study design types combined produced a moderate effect size.

As pointed out in our introduction, other researchers (Guthrie et al., 2007; Lazowski & Hulleman, 2015) who have conducted meta-analyses of educational interventions that included the measurement of self-efficacy have found that interventions had a moderate impact on students' motivation or performance. In

conducting a meta-analysis of the impact of CORI on motivation, Guthrie et al. (2007) found five studies in which CORI, when compared with alternative instruction, generated a mean self-efficacy effect size of 0.49. Although Lazowski and Hulleman (2015) were not able to find interventions that reflected only self-efficacy theory and so could only include self-efficacy theory with interventions grounded in other motivational theories, they did find that interventions focused on motivation, collectively, had a moderate impact on students' performance, behavior, or motivation with an average effect size of $d = 0.49$ (95% CI = [0.43, 0.56]). These findings resonate with our results.

Implications for Theory of Change in Reading Self-Efficacy

This meta-analysis enabled us to close several gaps in our understanding of the impact of interventions on reading self-efficacy. Our investigation demonstrated that the magnitude of the impact of interventions that included measures of reading self-efficacy as outcomes was statistically significant irrespective of study design. Beyond the finding that reading self-efficacy is malleable and can be altered through interventions, we also found that, when evaluating all study designs combined, intervention effects become larger as the number of major sources of self-efficacy included in the intervention increases. When comparing interventions that utilized one, two, or three major sources shaping self-efficacy, those having only one source produced the smallest effect size while those with three sources produced the largest effect size.

These findings address the theory of self-efficacy development proposed by Bandura (1986, 1997) that was presented as a general model in our introduction. Prior to this meta-analysis, we lacked a synthesis of studies using a range of interventions that would enable us to understand if and how the four major sources of self-efficacy (i.e., mastery experiences, vicarious experiences, verbal and social persuasion, and emotional and physiological states) in Bandura's theory of development, especially as it applied to reading self-efficacy, affected outcome. Based on our analysis, we now know that a significant correlation between the number of major sources shaping reading self-efficacy in an intervention and levels of growth in reading self-efficacy exists. The reciprocal interaction of the three sources combined may contribute to an explanation of our findings regarding the cause of reading self-efficacy's growth (Bandura, 1986). In addition, we found that for the treatment-control and pretest-posttest studies, vicarious experience as a major source of self-efficacy had a moderating influence on effect size, but this finding did not hold for treatment-comparison or all study types combined. Furthermore, we found a positive relationship between successful reading and reading self-efficacy, suggesting that as comprehension increased its impact on self-efficacy's effect size also increased.

Many of the studies included in this analysis used reading strategy instruction of various kinds, especially blends of specific reading strategies into programs, such as CORI (Guthrie, Klauda, & Ho, 2014; Wigfield et al., 2004) or READ 180 (Nelson, 2008). Some included a focus on enhancing reading self-efficacy through motivational intervention components, such as choice of interesting texts (Wozniak, 2010), goal setting (Nelson & Manset-Williamson, 2006), or a process goal orientation (Schunk & Rice, 1989). We did find a correlation between the

number of major sources of self-efficacy beliefs appearing in an intervention and the intervention's impact on measures of reading self-efficacy for treatment-comparison studies and for all study types combined when an outlier (Chang & Ho, 2009) was removed. Of the eight interventions that generated a Hedges' g score above 0.60, five of them included three of Bandura's major sources shaping self-efficacy (Schunk & Rice, 1987, 1989, 1991, 1993; Stekel, 1983). Two of the eight interventions included two major sources shaping self-efficacy (Garfield, 2000; McCrudden, Perkins, & Putney, 2005), and one intervention did not include any identified major source (Chang & Ho, 2009). None of the studies assessed students' physiological reactions.

Subsequent research could be conducted to explore the Bandura-based theory of self-efficacy development further because, as we have pointed out, we did not identify any interventions designed by researchers who explicitly articulated a self-efficacy theory that included the catalytic role of major sources shaping self-efficacy. Furthermore, the manipulation and measurement of emotional and physiological states as sources shaping self-efficacy might deepen our knowledge of the sensitivity of reading self-efficacy response to more sources shaping it and our understanding of Bandura's model.

Findings from this synthesis that support Bandura's theory of self-efficacy development also have implications applicable to practice. For educators seeking to enhance students' reading self-efficacy through interventions, developing and implementing interventions with more major sources shaping reading self-efficacy appear to be a reasonably grounded decision for the delivery of reading instruction. Enhancing the quality of those major sources might also contribute to self-efficacy's improvements. For example, attending to the quality with which strategies are modeled for students, providing vicarious experiences for them, offering persuasive feedback, and encouraging attention to readers' emotional and physiological states while reading could contribute to more potent motivational and performance outcomes.

The Effect of Other Moderators on Reading Self-Efficacy

Several others moderators were analyzed to discover their impact on the average effect size of reading self-efficacy, including grade level, prior reading performance of subjects, and duration of treatment.

Grade Level

Although we combined all grades into only two categories in our analysis, elementary (Grades 1–5) and all others, we found that in the instance of our pre-post effect size moderator analysis grade level had a statistically significant impact on average effect size. However, when all studies were combined for the analysis, 16 of the studies were conducted with elementary school subjects generating an average effect size of $g = 0.53$.

Elementary school level students ordinarily make greater gains in terms of achievement effect sizes in comparison to secondary school students (Bloom, Hill, Black, & Lipsey, 2008). Cheung and Slavin (2016), in their study of how methodological features in educational studies affect effect sizes, found that elementary school children showed slightly greater gains in effect sizes ($ES = +0.20$)

than secondary school level students ($ES = +0.17$), although the difference did not reach statistical significance. While the meta-analysis conducted by Lazowski and Hulleman (2015) included interventions to promote many kinds of theoretically grounded motivation, they found that average effect sizes by grades ranged from $d = 0.57$ for Grades 6 to 8 to $d = 0.42$ for high school with students in the elementary school grades having an average effect size of $d = 0.52$. However, when comparing various grade levels, their Q statistic was not significant and revealed that effect sizes of interventions included in their analysis did not differ based on grade level. While these analyses are far from conclusive, they indicate that further research is warranted to discover whether or not interventions to promote reading self-efficacy in the elementary grades, as distinguished from other theoretically grounded motivation theories, have more effect on students than interventions in higher grades. The reading self-efficacy of children in the elementary grades may be more susceptible to enhancement than in higher grades.

Prior Reading Performance of Subjects

Early in our work, we suspected that in studies with interventions that had struggling readers as their subjects less growth in reading self-efficacy would occur, especially compared with studies using non-struggling readers. Twenty-two of these 33 effect sizes arose from interventions with subjects who were coded as “struggling” readers. In our investigation, no significant differences were found in reading self-efficacy average effect sizes between students considered struggling versus non-struggling or mixed. However, we know that adolescents who were struggling with reading at the beginning of a 3-year, 50-minute daily intervention that included supplemental intervention made minimal improvements, suggesting that struggling readers at the middle and high school levels require longer and more intense interventions (Vaughn et al., 2012).

Duration of Treatment

When beginning this meta-analysis, we suspected that students receiving more exposure to an intervention over a longer period of time and with longer sessions would manifest higher levels of reading self-efficacy. Thus, we had anticipated that the Duration of Treatment Index would have a significant positive relationship with average effect sizes, perhaps because of our belief that more instruction would result in reading self-efficacy’s growth. That the outcome was not significant suggests other interpretations, one being that improvements in reading self-efficacy arising from these interventions came early and did not require sustained treatment. On the other hand, some of the treatments yielding little or no improvement in reading self-efficacy may have demonstrated improvement with more extensive treatment. Unfortunately, none of the included studies examined duration of treatment or session length empirically.

Fidelity of Implementation

Attention to fidelity of implementation, which we viewed as an indicator of study quality, did not reach statistical significance as a moderator in any of the three categories of study design. However, when all study designs were combined, fidelity of implementation did have a statistically significant impact on

effect size. Perhaps ironically, those studies in which fidelity of implementation was not addressed or mentioned ($n = 15$) generated a higher effect size ($g = 0.60$) than those that addressed it ($g = 0.14$). Another indicator of study quality, randomization, did not have a statistically significant impact on effect size in any category of study including all studies combined.

Publication Bias

When all study designs were combined for analysis, we found that published studies, and thus those peer reviewed, generated a larger effect size than studies not published. However, this finding of “publication bias” did not arise in any separate study design category. Our funnel plot analysis on all three groups and for all studies combined indicated that our set of studies was robust and that publication bias was not a significant concern.

Meaning and Measurement of Reading Self-Efficacy

Challenges entailed in operationally defining and measuring reading self-efficacy appeared in the studies included in this meta-analysis. In reviewing studies for inclusion, we encountered conceptualizations and definitions of self-efficacy along with its measurement that would have benefited from a more thoroughly developed and empirically grounded theoretical foundation. Schunk and Pajares (2009) have lamented the chronic mismeasurement of self-efficacy. Researchers, they reported, do not assess self-efficacy at appropriate levels of specificity, use items more closely related to self-concept or self-esteem than to self-efficacy, and overlook the suggestions that Bandura (2006) provided with respect to the measurement of self-efficacy. Specifically, Schunk and Pajares (2009) state, “Decontextualized or atheoretical self-efficacy assessments that lack consistency with the criterion task distort the influence of self-efficacy” (p. 50).

To account for the possible influence of the means of measuring reading self-efficacy on average effect size outcomes, we designed a Reading Self-efficacy Measurement Index that used as our benchmark Bandura’s (2006) procedures for the development of self-efficacy measurement. The index consisted of three factors, described earlier, that were reflected in specific items included in our coding sheets. Those reading self-efficacy instruments that were *reading only*, that focused on *self-efficacy* specifically, and that were *reading task-specific with a text* got the highest ranking.

Bandura-like measurements of self-efficacy are commonly researcher-developed measures that reflect the kind of intervention designed by the researcher(s). They are not standardized measures. Some educators (Cheung & Slavin, 2016, p. 286) opine that studies using researcher-developed measures that are inherent to treatments given, and thereby considered “overaligned,” should not be included in reviews of program evaluations that are intended to inform practice or policy because they overstate effect sizes. However, expecting researchers focused on reading self-efficacy to use only standardized measures of reading self-efficacy places a serious hurdle in the path to further research that includes measures sensitive to treatments given to improve reading self-efficacy because reading self-efficacy is a task-specific construct.

Adopting the policy position that researcher-developed measures inherent to treatment should be excluded from reviews of program evaluations, as Cheung and Slavin (2016) suggest, could have far-reaching ramifications for researchers in the field of self-efficacy, especially reading self-efficacy. Alternative solutions to understandable concerns about the overall alignment expressed by policy-shaping agencies, such as What Works Clearinghouse and the Department of Education's Institute of Education Sciences, warrant reflection. For example, consideration might be given to the development of a standardized measure for reading self-efficacy, such as the creation, validation, and inclusion of reading self-efficacy items in reading assessment instruments like the *Gates-MacGinitie Reading Tests* (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000) or an independent measure of the reading self-efficacy construct. An additional approach to the resolution of this debate would be to implement standardized *procedures* rather than standardized *measures*. Invariant procedures and materials characterize standardized measures, and this standardization is viewed as an attribute that provides rigor to the measurement process. In contrast, standardized procedures are characterized by invariant processes but may include variable materials or protocols tailored to a specific task or domain of interest. Standardized procedures that are theoretically based, well-described, and carried out in a replicable manner, similar to those recommended by Bandura (2006), may thus merit consideration. This is especially important when accounting for the domain-specific and task-specific nature of reading self-efficacy.

On the Relationship Between Interventions That Impact Reading Self-Efficacy and Reading Comprehension

Across a spectrum of developmental levels, researchers (Chapman & Turner, 1995; Wigfield & Guthrie, 1997) have found positive relationships between students' self-efficacy beliefs and reading performance. In our ancillary analysis of a subset of studies that included reading comprehension as well as self-efficacy measures, we found a strong positive relationship between reading self-efficacy and reading comprehension. However, the untangling of that relationship to clarify its underlying dynamics was limited because the correlations or coefficients between the two constructs did not allow predictive or causal claims. It is possible that, as students gain in reading comprehension, they gain in reading self-efficacy as their perceptions of themselves as capable readers rise. It is also plausible that, as readers gain in reading self-efficacy, they make advances in reading comprehension because of their enhanced beliefs in themselves as capable readers. A third explanation warrants consideration. Perhaps a bidirectional reciprocity exists between reading self-efficacy and reading comprehension such that their interaction enables both to grow, consistent with Bandura's (1986) triadic reciprocal causation model. While unable to make definitive claims about the relationship between these two constructs, our inability to disentangle them should not detract from the confirmation of their close connections. Further research is needed to clarify their complex and important relationship.

Limitations

Although we identified 30 studies that met our criteria for inclusion in this meta-analysis, these studies were divided into the three categories by study design

that led to 12 treatment–control studies, 12 treatment–comparison studies, and 6 within-group, pretest–posttest studies.

Because two treatment–control studies were designed so that the same treatment was presented to two independent groups and one pretest–posttest study was similarly configured, we were able to add three additional effect sizes to our pool. Had more studies been included in each of these study design categories, somewhat different outcomes might have been generated. As we analyzed moderators within each study design category, membership size in a moderating category was quite small, and one or two studies with exceptionally strong outcomes could have a large impact on effect size outcomes.

Several studies included in our analysis also had relatively small sample sizes. Therefore, differences in effect sizes within moderator analyses might have been stronger in several instances with larger sample sizes, but small samples led to insufficient power to detect those differences at significant levels. We also combined all studies across study design and recognize that the results of that combination had its own limitations even though we justified the combination on the basis of all studies addressing a common purpose: to discover the impact of an intervention of reading self-efficacy.

Few studies among those in our analysis used what could be deemed measures sensitive to fluctuations in reading self-efficacy. Mismeasurement has been a chronic issue in research on self-efficacy (Schunk & Pajares, 2009) and has often been manifested in studies of reading self-efficacy (Piercey, 2013). The range of shortcomings in reading self-efficacy measurements observed in this analysis spanned items that failed to focus on what individuals believe they can accomplish successfully, failed to focus on specific reading skills and tasks, and failed to provide a context for self-judgments because no specific text was provided (Bandura, 2006). Because of these shortcomings, we believe that data on reading self-efficacy extracted from included studies may distort the accuracy of subjects' self-efficacy for reading and their degree of change.

Last, our search for possible studies that could have been included in this analysis was limited. It did not include all possible conference abstracts and book chapters. Thus, some studies may have been overlooked. Furthermore, we did not systematically send out emails to leaders in the field who might have knowledge of “file drawer” studies they themselves completed but did not publish or studies by colleagues that were not published or disseminated (Rosenthal, 1979).

Conclusion

Self-efficacy has a rich and well-established theoretical foundation as a motivational construct and engine for engagement (Bandura, 1997; Schunk & Pajares, 2004). The importance of the construct and the potency of its impact have been acknowledged for decades. As Schunk and Pajares (2009) have pointed out, self-efficacy has a “powerful influence on individuals' motivation, achievement, and self-regulation” (p. 35). Students with relatively high self-efficacy are more prepared to engage, engage longer, demonstrate more interest, and manifest more achievement than students doubtful of their ability to learn successfully (Bandura, 1997). With a deeper understanding of reading self-efficacy, of how it might be enhanced through interventions, and of how improvements in its measurements

might better detect the impact of interventions, we have better chances of providing students with richer opportunities to enhance not only their motivation and engagement while reading but also their comprehension of texts.

References

- Supplementary Table S1 (available in the online version of the journal) lists studies included in the reading self-efficacy meta-analysis.
- Aloe, A. M., Becker, B. J., & Pigott, T. D. (2010). An alternative to R2 for assessing linear models of effect size. *Research Synthesis Methods, 1*, 272–283. doi:10.1002/jrsm.23
- Ballard, T. Y. (2007). *The effect of audio self-modeling on the reading fluency, comprehension, vocabulary and reading level of first and second grade students who are at-risk in reading* (Doctoral dissertation, Touro University International). Available from ProQuest Dissertations and Theses database. (UMI No. 3256153)
- Bandura, A. (1986). *Social foundations of thought & action: A social cognitive theory*. Upper Saddle River, NJ: Pearson Education.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age.
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 172–177. doi:10.1080/19345740802400072
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2006). *Comprehensive meta-analysis* (Version 2.2.064). [Computer software]. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley.
- Chandler, W. G. (2012). *Video self-modeling and improving oral reading fluency* (Doctoral dissertation, University of Kentucky). Available from ProQuest Dissertations and Theses database. (UMI No. 3579265)
- Chapman, J. W., & Turner, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology, 87*, 154–167.
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*, 283–292. doi:10.3102/0013189X16656615
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170–177. doi:10.1037/1082-989X.1.2.170
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463. doi:10.1111/j.0006-341X.2000.00455.x

- Eccles, J. S., Lord, S., & Buchanan, C. M. (1996). School transitions in early adolescence: What are we doing to our young people? In J. A. Graber, J. Brooks-Gunn, & A. C. Petersen (Eds.), *Transitions through adolescence: Interpersonal domains and context* (pp. 251–284). Hillsdale, NJ: Lawrence Erlbaum.
- Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher*, 49, 518–533.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13. doi:10.1037/0022-0663.93.1.3
- Guthrie, J. T., & Coddington, C. S. (2009). Reading motivation. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 503–525). New York, NY: Routledge.
- Guthrie, J. T., McRae, A., & Klauda, S. L. (2007). Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist*, 42, 237–250. doi:10.1080/00461520701621087
- Guthrie, J. T., Van Meter, P., McCann, A. D., & Wigfield, A. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly*, 31, 306–332. doi:10.1598/RRQ.31.3.5
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107–128.
- Holden, G. W., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990) Self-efficacy of children and adolescents: A meta-analysis. *Psychological Reports*, 66, 1044–1046. doi:10.2466/pr0.1990.66.3.1044
- Horowitz, J. S. (2004). *Mothers instilling literacy in kids (M.I.L.K.™): A program evaluation* (Doctoral dissertation, University of California, Irvine, and University of California, Los Angeles). Available from ProQuest Dissertations and Theses Database. (UMI No. 3131712)
- Lazowski, R. A., & Hulleman, C. S. (2015). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, 86, 602–640. doi:10.3102/0034654315617832
- Linnenbrink-Garcia, L., & Patall, E. A. (2016). Motivation. In L. Corno & E. Anderman (Eds.), *Handbook of educational psychology* (pp. 91–103). Washington, DC: American Psychological Association.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.
- Martin, A. K. (2010). *Collaborating for convergence: Instructional interventions for children's reading of expository text*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning*, 57, 417–442. doi:10.1111/j.1467-9922.2007.00421.x
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 38, 30–38. doi:10.1037/0022-0167.38.1.30

- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578. doi:10.3102/00346543066004543
- Phan, H. P., & Ngu, B. H. (2016). Sources of self-efficacy in academic contexts: A longitudinal perspective. *School Psychology Quarterly, 31*, 548–564. doi:10.1037/spq0000151
- Piercey, R. R. (2013). *Reading self-efficacy in early adolescence: Which measure works best?* (Unpublished doctoral dissertation). University of Kentucky, Lexington.
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*, 207–236. doi:10.3102/0034654315582067
- Ren, W. (2000). Library instruction and college student self-efficacy in electronic information searching. *Journal of Academic Librarianship, 26*, 323–328. doi:10.1016/S0099-1333(00)00138-5
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. doi:10.1037/a0026838
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*, 54–61. doi:10.1037/0022-0663.79.1.54
- Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (pp. 115–138). Greenwich, CT: Information Age.
- Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35–54). New York, NY: Routledge.
- Solheim, O. J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology, 32*, 1–27. doi:10.1080/02702710903256601
- Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction, 16*(1), 57–71. doi:10.1016/j.learninstruc.2005.12.006
- Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65–94. doi:10.1037/0033-2909.133.1.65
- Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *Journal of Educational Research, 100*, 81–101. doi:10.3200/JOER.100.2.81-101
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature. *Review of Educational Research, 78*, 751–796. doi:10.3102/0034654308321456
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the qualities of academic motivation. *Educational Psychologist, 41*, 19–32. doi:10.1207/s15326985ep4101_4
- Vaughn, S., Wexler, J., Leroux, A. J., Roberts, G., Denton, C. A., Barth, A. E., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities, 45*, 515–525. doi:10.1177/0022219411402692

- Voorhees, S. (2011). Why the dog eats Nikki's homework: Making informed assignment decisions. *Reading Teacher*, 64, 363–367. doi:10.1598/RT.64.5.8
- Waleff, M. L. (2010). *The relationship between mastery orientation goals, student self-efficacy for reading and reading achievement in intermediate level learners in a rural school district*. Ann Arbor, MI: ProQuest LLC.
- Wang, J. H., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. *Reading Research Quarterly*, 39, 162–186.
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth on their reading. *Journal of Educational Psychology*, 89, 420–432. doi:10.1037/0022-0663.89.3.420
- Wigfield, A., Guthrie, J. T., & McGough, K. (1996). *A questionnaire measure of children's motivations for reading* (Instructional Resource No. 22). Athens, GA: National Reading Research Center, Universities of Georgia and Maryland College Park.

Authors

NORMAN J. UNRAU is professor emeritus at California State University, Los Angeles, 1604 Manzanita Lane, Manhattan Beach, CA, 90266, USA; email: nunrau@calstatela.edu. He has coedited *Theoretical Models and Processes of Reading* (5th and 6th editions), written *Content Area Reading and Writing: Fostering Literacies in Middle and High School Cultures*, and published research on reading and motivation in professional journals. He serves with California State University colleagues and secondary teachers in the development and evaluation of expository reading and writing courses to reduce college remediation.

ROBERT RUEDA is professor emeritus of educational psychology at the Rossier School of Education, University of Southern California, 1440 El Travesia Drive, La Habra Heights, CA 90631, USA; email: rueda@usc.edu. His research has centered on the sociocultural basis of motivation, learning, and instruction with a focus on reading and literacy in English learners and students in at-risk conditions. He is a member of the National Academy of Education and is a fellow of the American Psychological Association and the American Educational Research Association. He recently served as the associate editor of the *American Educational Research Journal* (*Teaching, Learning, and Human Development*) and, in addition, serves on the editorial boards of several other educational journals.

ELENA SON is a lecturer at Hankuk University of Foreign Studies in South Korea, 107 Imun-ro, Dongdaemum-gu, Seoul, 02450, Republic of Korea; email: eson@usc.edu. She was a research professor at Brain and Motivation Research Institute at Korea University. She obtained her doctoral degree from University of Southern California. Her research interests center on academic motivation, particularly of culturally diverse learners, motivation interventions, bridge between research and practice, and second language learning. Son has presented her research at the annual meetings of the American Educational Research Association (AERA) and the American Psychological Association (APA). She has served as a reviewer for educational journals, including *AERA open* and *Journal of Experimental Education*.

Unrau et al.

JOSHUA R. POLANIN, PhD, is a principal researcher at American Institutes for Research, 1000 Thomas Jefferson St. NW, Washington, DC 20007, USA; email: jpolanin@air.org. His research focuses on improving the methods of meta-analysis. He is also the principal investigator of a National Institute of Justice–funded systematic review and meta-analysis on the longitudinal relationship between school violence and delinquency.

REBECCA J. LUNDEEN, EdD, is an educational psychologist, nationally certified school psychologist, and educator with experience in psychoeducational assessment, program evaluation, intervention, and instructional design; University of Southern California Rossier School of Education, Waite Phillips Hall, 3470 Trousdale Parkway, Los Angeles, CA 90089, USA; email: rebecca.lundeen@gmail.com. She teaches graduate courses that apply evidence-based learning, motivation, and instructional design principles in practice to improve performance. She also provides consultation to local education agencies and families.

ALISON K. MURASZEWSKI, EdD, is an adjunct faculty member at Rossier School of Education at the University of Southern California, Waite Phillips Hall, 3470 Trousdale Parkway, Los Angeles, CA 90089, USA; email: alisonkeller4@yahoo.com. Her expertise is in learning and motivational theories and principles to advance learning, self-regulation, metacognition, and instructional design. Her research has focused on strategies for improving instruction for students with emotional and behavioral challenges. She also supervises student teachers working on their Clear California Teaching Credential.