

# Speaking to Read: Meta-Analysis of Peer-Mediated Learning for English Language Learners



Journal of Literacy Research  
2014, Vol. 46(3) 358–382  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1086296X14552179  
jlr.sagepub.com



**Mikel W. Cole<sup>1</sup>**

## Abstract

This meta-analysis examines the effectiveness of a group of instructional approaches (i.e., cooperative, collaborative, and peer tutoring) at improving literacy outcomes for English language learners. Main effects analyses of a sample of 28 experimental and quasi-experimental studies reveal that peer-mediation is more effective for ELLs than individualized or teacher-centered comparison conditions ( $g=.486$ ,  $SE=.121$ ,  $p<.001$ ). A number of potential moderators were examined, and two study quality variables proved significant. Also, grade level was a significant moderator, with middle school students demonstrating much smaller gains than elementary or high school students. Finally, descriptive analysis of moderators provides tentative evidence that ELLs showed greater gains on word-level outcomes than text-level outcomes and that interventions for which peer-mediation was one of several tightly-woven components were twice as effective as interventions utilizing peer-mediation alone.

## Keywords

English language learners, peer mediated learning, meta-analysis

## Introduction

English language learners (ELLs) now comprise approximately 20% of the total U.S. school population (National Center for Education Statistics [NCES], 2011), and as the fastest growing group of students in U.S. schools (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2008; McKeon, 2005), this proportion is likely to increase. Moreover, these culturally and linguistically diverse students are dispersing into regions of the country that have historically enrolled few ELLs (Capps et al., 2005). Already, 60% of all

---

<sup>1</sup>Clemson University, SC, USA

### Corresponding Author:

Mikel W. Cole, Clemson University, Clemson, SC, USA.  
Email: mikel.w.cole@gmail.com

teachers have at least one ELL in their classroom (American Federation of Teachers, 2006).

However, most teachers have little preparation to provide the specialized learning these students require (Ballantyne, Sanderman, & Levy, 2008; Harper & de Jong, 2009; Menken & Antunez, 2001). For example, few teacher-preparation programs require that mainstream teachers enroll in courses specifically addressing the needs of ELLs, and those who do usually require only one course. Despite the many ELLs in U.S. schools, practicing teachers receive almost no professional development aimed at helping them work with ELLs. Consequently, most teachers report feeling unprepared to teach ELLs (Ballantyne et al., 2008).

The rapid growth and dispersion of the ELL population and the general lack of teacher preparation for working with them may contribute to the poor performance of ELLs on national and state exams. Unlike the narrowing of the performance gap between ELLs and their English-proficient peers on the Math portion of the National Assessment of Educational Progress (NAEP) during the previous 10 years, the performance gap on the Reading portion remains unchanged during that period (Wilde, 2010).

Previous syntheses of research about effective literacy instruction for ELLs (e.g., August & Shanahan, 2006) have concluded that too little experimental research exists to make strong claims about the effectiveness of many specific instructional practices. The authors of the National Literacy Report on Language Minority Children and Youth claimed that, in general, the characteristics of effective ELL instruction “overlap with those of effective instruction for nonlanguage-minority students” and “these factors need to either be bundled and tested experimentally as an intervention package or examined as separate components to determine whether they actually lead to improved student performance” (p. 520).

The study reported here addresses the need for evidence that might guide efforts to enhance literacy among a growing ELL population. Specifically, this article presents the results of a meta-analysis investigating the overall effectiveness of peer-mediated learning, an approach that has been widely investigated as a means for improving literacy outcomes for ELLs. The results of this synthesis are intended to respond directly to the National Literacy Panel’s recommendation to test the effectiveness of particular instructional approaches by exploring the evidence available from experimental and quasi-experimental studies. Meta-analysis addresses the two research questions guiding this study. First, main effects analyses provide a weighted estimate of the mean effect across studies to address the question, “Does peer-mediated learning improve literacy outcomes for English language learners?” Second, moderator analyses explore the extent to which measured variables explain heterogeneity of effects addressing the question, “Under what circumstances is peer-mediated learning effective for ELLs?”

## Description of Peer-Mediated Learning

Peer-mediated learning refers broadly to an instructional approach that emphasizes student–student interaction, and it is intended to provide an alternative to teacher-centered or individualistic approaches to learning. In practice, peer-mediated learning includes a variety of approaches, each with distinct supporting literatures. The

meta-analysis reported here synthesizes three variations of peer-mediated learning: cooperative, collaborative, and peer tutoring, a distinction used in previous syntheses (e.g., Cohen, 1994; Hertz-Lazorowitz, Kirkus, & Miller, 1992). Nonetheless, there are numerous precedents for treating these theoretically and practically different approaches as similar, though not synonymous, terms (Cohen, 1994; Johnson, Johnson, & Stanne, 2000; Slavin, 1996; Swain, Brooks, & Tocalli-Beller, 2002).<sup>1</sup>

### *Cooperative Learning*

Cooperative learning represents what Slavin (1996) calls “one of the greatest success stories in the history of educational research” (p. 43), and he claims that hundreds of control group evaluations have been conducted since the 1970s, with the most common outcome being an increase in academic achievement. Johnson et al. (2000) conducted a widely cited meta-analysis of the effects of cooperative learning on various measures of academic achievement, and the authors note that “cooperative learning is a generic term referring to numerous methods for organizing and conducting classroom learning” (p. 3).

A definitive characteristic of cooperative learning is the degree of structure in tasks and students’ roles (Oxford, 1997; Slavin, 1996). In this article, degree of structure is the defining criterion that distinguishes cooperative and the collaborative approaches described in the next section. In general, cooperative methods emphasize carefully structured groups, and students typically have well-defined roles. For example, in a jigsaw activity (Aronson & Patnoe, 2011), students in a small group are each responsible for mastering one aspect of a topic and for reporting to their group as the designated expert on that aspect. For the group to demonstrate mastery of the material, each person must adequately learn and then convey an aspect of the overall topic.

### *Collaborative Learning*

Some reviews treat cooperative and collaborative methods as if they are similar, essentially identical, methods (e.g., Cohen, 1994). However, the present meta-analysis follows those researchers who see these two approaches as similar, but distinct, methods for engendering active, student-centered learning (e.g., Hertz-Lazorowitz et al., 1992; Mathews, Cooper, Davidson, & Hawkes, 1995; Oxford, 1997). Essentially, collaborative learning represents a less structured set of approaches to small group learning. Cooperative methods, however, tend to emphasize highly structured student roles and maintain more traditional teacher-centered distributions of power. In collaborative methods, completion of a complex task tends to be the central objective, and students are often left to their own devices to divide the labor, develop relations of power and authority, and to navigate task demands.

### *Peer Tutoring*

Peer-tutoring approaches also vary widely (see Goodlad, 1998, for a more detailed discussion), though in general they utilize older, or more capable and more

academically successful, peers to provide one-to-one instruction for struggling learners. Although peer tutoring can occur *within* grade levels, it is frequently used *between* grade levels, with older students tutoring younger students. Thus, by utilizing well-defined roles and structured relationships of power, peer-tutoring approaches contain many elements of more structured cooperative learning approaches. Of course, as with cooperative and collaborative approaches, peer-tutoring methods emphasize peer-to-peer interaction and seek to foster active, rich discussion from all participants.

## **Theoretical Rationale for Peer-Mediated Literacy Learning**

Teacher-dominated instruction is the norm in most classrooms, and ELLs typically struggle in these monologic classrooms (Gutierrez et al., 1995). Even in programs specifically designed for language learners, ELLs are rarely afforded opportunities for active participation. For example, a nationally representative, longitudinal study in the United States of three widely used models of ELL instruction (i.e., Structured-English Immersion, Early-Exit Transitional Bilingual, and Late-Exit Transitional Bilingual) found that in all three models, teachers dominated classroom discourse and students were rarely provided opportunities for active learning. Instead, in more than half of observed instances, students provided no verbal responses at all (Ramirez et al., 1991).

Peer-mediated learning stands in contrast to the teacher-centered instruction ELLs typically encounter in classrooms. It is based on theories of learning that inform student-centered, dialogic instruction, specifically Vygotskian, sociocultural theories of language development, and psycholinguistic research inspired by Long's Interaction hypothesis (Long, 1996).

### ***Sociocultural Theories of Language Development***

The work of Vygotsky, and sociocultural researchers who have utilized his ideas, provides the primary theoretical rationale for using peer-mediated approaches to improve literacy outcomes for ELLs (Donato, 1994; Lantolf & Thorne, 2006; Lee & Smagorinsky, 2000, Vygotsky, 1978). Vygotsky claims that thought is fundamentally social, and he argues that every cognitive development arises first in social interaction. This intrinsically social view of cognition shapes his description of learning and teaching and gives rise to one of the most well-known but widely misunderstood Vygotskian constructs: the zone of proximal development (ZPD). In its original sense, the ZPD refers primarily to child development mediated by an adult or more proficient other. The ZPD defines the difference between what a child can accomplish unassisted and what a child can accomplish with assistance from a more proficient other. From a Vygotskian perspective, unassisted performance reveals what has already been mastered in past development, and assisted performance indicates what is in the process of being mastered in current development.

The assistance offered to a learner by a more proficient other is typically called *scaffolding*, a term popularized by Wood, Bruner, and Ross (1976). Initial articulations

of scaffolding focused on ways that adults or experts supported children through utterances, gestures, and facial expressions; more recently, second language (L2) scholars have argued that scaffolding can also be provided by peers (Donato, 1994; Lantolf, 2000; Swain & Lapkin, 1995), that is, working together in collaborative groups, students can scaffold one another to co-construct a performance that exceeds what any of the students would have produced individually. Thus, by suggesting that social processes inherent in spoken interaction during collaborative activity might improve individual performance, Vygotskian constructs like mediation and scaffolding provide the basic theoretical framework for the term *peer-mediated learning* used in this study.

### ***Psycholinguistics and the Interaction Hypothesis***

Work conducted in the field of second language acquisition (SLA) represents another theoretical rationale for the use of peer-mediated methods with language learners. Drawing on the work of Long (1981, 1996; see also Gass & Mackey, 2006; Pica, 1994) and his *Interaction hypothesis*, a line of psycholinguistic research suggests that language learning is especially effective during interaction. These researchers claim that interactionally modified language input coupled with purposeful language output improves language learning outcomes in ways that extensive language input alone cannot. Moreover, these researchers argue that interaction focuses learners' attention on language and increases motivation for language learning. This contrasts with a view of SLA that suggests that the process is largely subconscious and driven almost entirely by the amount and availability of comprehensible input and that emphasizes reading and listening comprehension during instruction (e.g., Krashen & Terrell, 1983).

Taken together, these two theoretical perspectives suggest that students engaged in interaction can scaffold each other toward better language and literacy performance than typical, teacher-dominated or individualistic approaches allow. In particular, second language learners struggling to master the academic demands of schooling in the United States at the same time that they are learning the language may especially benefit from instruction that targets oral language and literacy development simultaneously.

### **Empirical Rationale for Peer-Mediated Learning**

This study addresses two primary research questions. The first main effects question asks if peer-mediated learning is effective at improving literacy outcomes for English learners, and the second moderator question asks what variables explain variability in effects. Both questions highlight key questions raised by previous syntheses.

### ***Does Peer-Mediated Learning Improve Literacy Outcomes for ELLs?***

Previous syntheses of qualitative, descriptive (Genesee, Lindholm-Leary, Saunders, & Christian, 2005; Gersten & Baker, 2000), and quantitative, experimental (August &

Shanahan, 2007; Cheung & Slavin, 2005) research of promising instructional practices for ELLs suggested that cooperative and collaborative learning were important components of classroom instruction that improves language, literacy, and academic learning for culturally and linguistically diverse students. Nonetheless, no previous meta-analysis of the effectiveness of peer-mediated learning at improving literacy outcomes for ELLs exists, and the National Literacy Panel's search yielded too few high-quality experimental studies to even compute a mean effect size.

Previous meta-analyses of peer-mediated methods report consistently positive effects when compared with individualistic and teacher-driven approaches (Johnson et al., 2000; Johnson, Maruyoma, Johnson, & Nelson, 1981; Rohrbeck, Fantuzzo, Ginsberg-Block, & Miller, 2003; Roseth, Johnson, & Johnson, 2008). However, these meta-analyses do not focus specifically on ELLs or literacy outcomes. Instead, they provide empirical evidence of the effectiveness of peer-mediated approaches with language majority students.

Related meta-analytical research also suggests that peer-mediated learning might be effective at promoting both spoken and written language outcomes for second language learners (Keck, Iberri-Shea, Tracy-Ventura, & Wa-Mbaleka, 2006; Mackey & Goo, 2007). These two meta-analyses synthesized SLA studies of interaction, and both reported positive effect sizes when interaction was compared with more traditional language pedagogy (approximately 0.46 *SD* and 0.61 *SD*, respectively<sup>2</sup>). However, these studies examined the verbal interactions of students and did not specifically examine the instructional approaches used.

Thus, previous syntheses suggest that peer mediation is generally more effective than teacher-centered or individualistic instruction, that interaction and negotiation of meaning improve oral and written language outcomes for second language learners, and that effective classroom instruction for ELLs typically includes a peer-mediated component. Nonetheless, none of these syntheses directly examine the ability of peer-mediated learning to improve literacy outcomes for ELLs. Thus, the present meta-analysis was conducted to address this more focused question.

### *Under What Circumstances Does Peer-Mediated Learning Improve Literacy Learning for ELLs?*

Methodological, instructional, and learner variables potentially influence the effectiveness of any intervention designed to improve learning, and previous syntheses reported findings that provided the rationale for specific moderator analyses reported in this meta-analysis.

Several meta-analyses of peer-mediated learning compared the effectiveness of various types of peer-mediated learning, but results were inconsistent across syntheses (Johnson et al., 2000; Johnson et al., 1981; Roseth et al., 2008). Thus, it is unclear whether cooperative, collaborative, or peer tutoring is likely to be most effective. Degree of structure is a distinctive characteristic between these approaches and largely defines the difference between studies coded as cooperative and studies coded as collaborative in this meta-analysis. Some previous studies reported that larger academic

gains were associated with more structured approaches (Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Johnson et al., 2000; Johnson et al., 1981; Slavin, 1996), whereas others found that less structured, more conceptually driven and flexible methods were more effective (Cohen, 1994; Johnson et al., 2000; Rohrbeck et al., 2003). Similarly, some previous syntheses questioned whether peer-mediated learning was effective alone or only when included as one component of a complex intervention (August & Shanahan, 2007; Cheung & Slavin, 2005; Johnson et al., 2000; Johnson et al., 1981).

Second language research typically distinguishes between second language (SL) settings (i.e., learners learning a language spoken in the local community) and foreign language (FL) settings (i.e., learners learning a language not spoken in the local community) because of the importance of differences in access and exposure to the target language among settings (e.g., Lightbown, 2000; Turnbull & Arnett, 2002). Mackey and Goo (2007) found no differences in the effectiveness of interaction between SL and FL settings. Importantly, the National Literacy Panel only included studies of English learners conducted in the United States (i.e., SL settings); thus, the results of the National Literacy Panel purposefully left the question of language setting unanswered.

Several syntheses analyzed whether the type of measure used moderated effectiveness. Previous syntheses of literacy outcomes examined whether word-level outcomes differ from text-level outcomes. The National Literacy Panel found that word-level outcomes were associated with larger effects than text-level outcomes (August & Shanahan, 2007), but Keck et al. (2006) reported that lexical outcomes ( $d = .90$ ) were essentially of the same magnitude as grammatical outcomes ( $d = .94$ ). Keck and colleagues reported that effect sizes associated with standardized, school-based, and researcher-created assessments did not differ significantly. This finding is surprising given prior research that found researcher-created assessments generally produced larger effect sizes than standardized measures (e.g., Bloom, Hill, Black, & Lipsey, 2008; Slavin & Madden, 2011).

Finally, a number of other variables were examined in previous syntheses and were coded for analysis as potential moderators in this meta-analysis, including study design and quality, whether or not peer mediation was more effective alone or as one component of a more complex intervention package, teacher experience, age of students, and ethnicity of students.<sup>3</sup>

This review of the empirical literature highlights the importance of both main effects and moderator analyses. No extant synthesis directly examines the effectiveness of peer-mediated learning at improving literacy learning for language learners, despite a clear call from the National Literacy Panel for experimental or quasi-experimental evidence of the effectiveness of particular instructional approaches; consequently, the first research question addressed by this meta-analysis asks whether or not peer-mediated learning is effective at improving literacy outcomes for ELLs. These main effects analyses provide a weighted mean effect size estimate of the average difference between peer-mediated learning and teacher-centered or individualistic learning across studies. Second, the inconclusive evidence in previous studies about the role that methodological, instructional, and learner variables play in the effectiveness



of peer-mediated learning for ELLs suggests that moderator analyses are critical in determining where, how, and with whom peer-mediated learning is most likely to be effective. Consequently, the second research question addressed by this meta-analysis examines the extent to which variations in these variables (e.g., type of peer-mediated learning, English as a second language [ESL] or English as a foreign language [EFL] setting, word- or text-level outcome) across studies moderate the magnitude of the reported effect sizes.

## Method

### *Criteria for Inclusion and Exclusion of Studies*

A number of researchers have argued that not enough experimental evaluations of intervention effectiveness exist in the ELL literature (e.g., August & Shanahan, 2006; Slavin & Cheung, 2005). Therefore, this meta-analysis includes both experimental and quasi-experimental studies, and post hoc analyses provide empirical estimates of differences in effect sizes associated with study design and quality.

*Types of studies.* Experimental and quasi-experimental studies were included in the review. For studies in which non-random assignment was used, studies must have included pre-test data or must have statistically controlled for pre-test differences (e.g., ANCOVA). Similarly, studies that tested more than one treatment against a control group were included as long as one treatment could readily be identified as the focal treatment. If a study did not include a control group, it was excluded. Single-group designs were excluded as single-group effect sizes are not directly comparable, nor able to be synthesized, with two-group designs.

Similarly, published and unpublished studies were included to tap the elusive “gray literature” (Cooper, Hedges, & Valentine, 2009; Lipsey & Wilson, 2001). Published manuscripts included peer-reviewed manuscripts in journals and books, and unpublished manuscripts included dissertations and technical reports. Technical reports are often rigorous reports with methodological detail not presented in the abbreviated format favored by journals; dissertations also contain more detail than a published journal article but quality varies widely; however, neither technical reports nor dissertations undergo the peer-review process. Consequently, post hoc analyses explored differences in effect sizes associated with published versus unpublished studies.

For practical purposes studies must have been published in English, though the research may have occurred in any country with participants of any nationality. In addition, the target language must have been English to facilitate direct comparisons with ELLs in U.S. schools; however, participants may represent any language background, and instruction may have occurred in any language.

*Types of participants and interventions.* Studies must have tested the effects of peer-mediated learning involving students between the ages of 3 and 18, again to facilitate comparisons with U.S. students in K-12 educational settings. For example, in studies



of peer tutoring, both students for whom outcomes were measured and students who acted as tutors must have been within this age range to preserve the focus on peer interactions. Also, participants must have included students identified as ELLs (though methods of identification and definitions of ELLs vary across states and districts in the United States), and results must have been exclusively, or disaggregated, for ELLs.

Interventions may utilize a number of instructional activities, but peer–peer interaction must have been a focal aspect of the intervention. Furthermore, comparison groups must not have received instruction for which peer-mediated learning was widely used, and studies that only provided a cooperative intervention were coded separately from those that involved more complex interventions in which peer-mediated methods were just one component (e.g., Success for All/Éxito para Todos). Studies for which peer–peer interaction could not be identified as a focal feature of the intervention were excluded, as were studies for which comparison groups used extensive peer assistance. When published descriptions were insufficient to determine a study's eligibility, study authors were contacted. If sufficient information was still unavailable, the study was excluded.

*Types of outcomes and instruments.* Diverse instruments were used to assess effectiveness, including norm-referenced tests, researcher and teacher-created measures, and psychological and sociological instruments. These characteristics were coded to enable both inferential moderator and descriptive analyses.

### *Search Strategy for Identifying Relevant Studies*

Multiple databases were searched using consistent combinations of keywords, though specific formats varied according to individual database preferences (e.g., AND used between terms for the PsychINFO search). Several databases were combined into simultaneous searches. For instance, the ProQuest search included the following individually selected databases: Dissertation Abstracts International, Ethnic News Watch, and several subsets of the Research Library collection—core, education, humanities, international, multicultural, psychology, and social sciences. Similarly, PsychINFO included the following databases, which were manually selected: ERIC, IBSS, CSA Linguistics, Language, and Behavior, PsychArticles, PsychINFO, and Sociological Abstracts. Furthermore, potentially relevant studies were cross-cited using the bibliographies of previous syntheses and identified studies.

All studies were identified through the following process: titles and abstracts were first skimmed to identify potentially relevant studies; if a study appeared to be a possible candidate, the full study was retrieved to the extent possible. If the study was not immediately available, Interlibrary Loan requests and librarian searches were pursued. If this did not succeed, attempts were made to contact the author of the study. Studies not retrieved at that point were deemed unavailable.

Studies were excluded at this point if closer examination revealed that they violated inclusion criteria or if an effect size could not be extracted from the information provided. As noted previously, attempts were made to retrieve necessary information

from the authors, though in many cases data were no longer available or the authors could not be reached. To facilitate future syntheses that might use different inclusion criteria, these “near miss” studies are included in the References, but no further analyses were conducted with these studies.

The researcher functioned as the primary coder, and all of the studies were coded by the researcher. Reliability of inclusion and exclusion criteria, as well as coding of key substantive and methodological variables, was assessed by comparing the primary coding with the coding of two independent coders. The additional coders were doctoral students with relevant experimental and statistical training. After some discussion of the inclusion and exclusion criteria and multiple rounds of practice with examples, the other coders made inclusion/exclusion decisions for a subsample of 30 abstracts.

### *Description of Methods Used in Primary Studies*

Previous syntheses suggested that high-quality experimental studies of effective literacy instruction for ELLs are scarce. Consequently, it seemed appropriate to cast a wide net, a long-standing approach to social science meta-analyses (e.g., Smith, Glass, & Miller, 1980). As a result, many small-sample studies utilizing quasi-experimental designs, with and without cluster randomization, were included and few large-sample studies with rigorous randomization were found. Furthermore, the broad conceptualization of peer-mediated learning resulted in a variety of interventions and approaches to data collection. The quality of included studies has a potential effect on the final synthesis; consequently, an analysis was conducted to determine the extent to which study quality was related to reported effects. Studies were coded to reflect the extent to which they used randomization, and the level at which randomization occurred. Similarly, studies were coded to assess the degree to which baseline equivalence between the control and treatment groups was measured in the original studies. For the sake of moderator analysis, study quality was assessed on a three-level scale as follows: (a) high-quality studies assessed pre-test equivalence *and* used a covariate to control for pre-test differences, (b) medium-quality studies assessed pre-test equivalence *or* used a covariate to control pre-test differences, and (c) low-quality studies did neither.

In several studies, pre-test data were available, but the original researchers did not use pre-test data in their post-test data analyses. That is, pre-test differences were left unadjusted in final analyses. In these situations, post hoc adjustments were made to control for pre-test differences. Specifically, pre-test means were subtracted from post-test means for both the treatment and the control groups, and these differences were used as the mean gain scores from which effect sizes were computed.

### *Analyses Used in This Meta-Analysis*

Main effects analyses involved computation of a standard mean difference effect size, which is a weighted estimate of the difference between treatment and comparison

groups (Cooper et al., 2009; Lipsey & Wilson, 2001). All effect sizes were calculated as Hedge's  $g$ , which is an inverse-variance weighted estimate to control for sample size bias. Random effects models were assumed, primarily because the assumptions of a fixed model were untenable (e.g., the assumption that there is only a single effect size throughout the observed sample and all possible populations from which it could have been conceptually drawn). Heterogeneity was assessed using the  $Q$  statistic, which describes the degree to which effect sizes vary beyond the degree of expected sampling error, and  $I^2$ , which indicates the amount of heterogeneity that exists between studies (Higgins, Thompson, Deeks, & Altman, 2003). Finally, participants in educational research are often assigned using nested groups (e.g., intact classrooms), and the estimates provided by standard analyses are unadjusted for these cluster effects. Although the effect size estimates are not usually too distorted by cluster effects, the standard errors on which the inverse-variance weights are computed are often dramatically incorrect (Hedges, 2007). Adjustments to standard errors were made using McHugh calculations (McHugh & Lipsey, 2007).

Empirical examination of publication bias involved three distinct analyses. First, a simple difference in means between published and unpublished studies provided a preliminary estimate of the magnitude of differences in average effect sizes. Second, the visual inspection of a funnel plot with missing studies imputed provided an analysis of the probability that small studies reporting null or negative results were under-reported in the literature. Third, computation of Egger's regression provided null hypothesis estimates that tested whether smaller sample sizes were associated with larger gains. None of these tests can directly confirm the presence or absence of publication bias; rather, taken together, they provide information about the likelihood of publication bias in observed results (Cooper et al., 2009; Lipsey & Wilson, 2001).

Moderating variables are those that may affect overall effect size estimates through covariation with the independent variables of interest. A number of study, treatment, and participant variables were analyzed as moderators in Comprehensive Meta Analysis (CMA) analysis and as correlates in SPSS. Separate analyses were conducted for each of these variables, and the results for these moderator analyses are presented separately for each moderator of interest. The included sample of studies was too small to allow for meta-regression analyses with sufficient statistical power to prevent capitalizing on chance. Thus, independent moderator analyses were conducted and simple bivariate correlations were provided to empirically examine the possibility of confounding variables. Finally, the relatively small sample of studies suggests that small effects might have gone undetected because of power limitations; consequently, descriptive analyses of differences in the means of moderator variables were also included to identify patterns of potential interest and future study.

### **Coding Reliability**

Coding reliability was assessed through the measurement of inter-rater reliability. Following exclusion/inclusion reliability assessment, the researcher met with the additional coders to discuss and practice using the coding manual on three examples.

Following this initial training, the coders coded five studies independently. The researcher then met again with the coders to discuss the initial coding and to practice together on two additional examples. Following the second training session, the two additional coders coded 10 more studies independently. Thus, the coders independently coded 15 studies each, with a total subsample of 25 studies included for the assessment of reliability. The studies were drawn evenly from published and unpublished studies. Cohen's kappa was calculated for categorical variables, while Pearson's  $r$  was calculated for continuous variables. Inter-rater reliability varied considerably across variables; mean Cohen's kappa for categorical variables was .787 with a range of .318 to 1.0. Pearson's  $r$  was calculated for continuous variables, and mean agreement among raters was .927 for continuous variables, though inter-rater reliability for continuous variables ranged between .85 and 1.0. Problematic variables were discussed and revised, and ultimately, all differences were resolved to consensus.

## Results

### *Included Sample*

The sample of studies included for this study consisted of 28 independent samples that tested the effectiveness of peer-mediated learning at improving reading and writing outcomes. In some cases, a single study report provided data for multiple independent samples, such that the following 22 reports provided data for 28 independent samples. Table 1 provides a brief summary of the included studies.

As indicated in Table 1, more than half of the studies were published since 2000, indicating that the field of peer-mediated learning remains quite active. Only 9 of the included studies were published in peer-reviewed journals, whereas the other 13 were dissertations or technical reports. This suggests that the search method tapped the unpublished literature deeply enough to facilitate analysis of publication bias. Similarly, 9 studies were conducted in the United States and the other 13 were conducted in other countries. Only 2 studies conducted in the United States were published in peer-reviewed journals, supporting the claims of previous syntheses that found few published studies (e.g., August & Shanahan; Genesee et al., 2006). Notably, only one study used a true experiment, indicating that insufficient variability in research design existed in the included sample to facilitate moderator analysis of this variable.

### *Main Effects Analysis*

A random effects model of the un-corrected and un-Winsorized data generated a mean effect size estimate for the 28 literacy outcomes of .551 ( $SE = .111$ ,  $p < .001$ ); however, after adjustments for outliers, pre-test differences, and cluster randomization, the mean effect size estimate decreased and the variance increased slightly (.486,  $SE = .121$ ,  $p < .001$ ), suggesting that outliers and cluster randomization exerted a measurable effect on the original estimates. The adjusted distribution of literacy outcomes is illustrated by the forest plot in Figure 1.

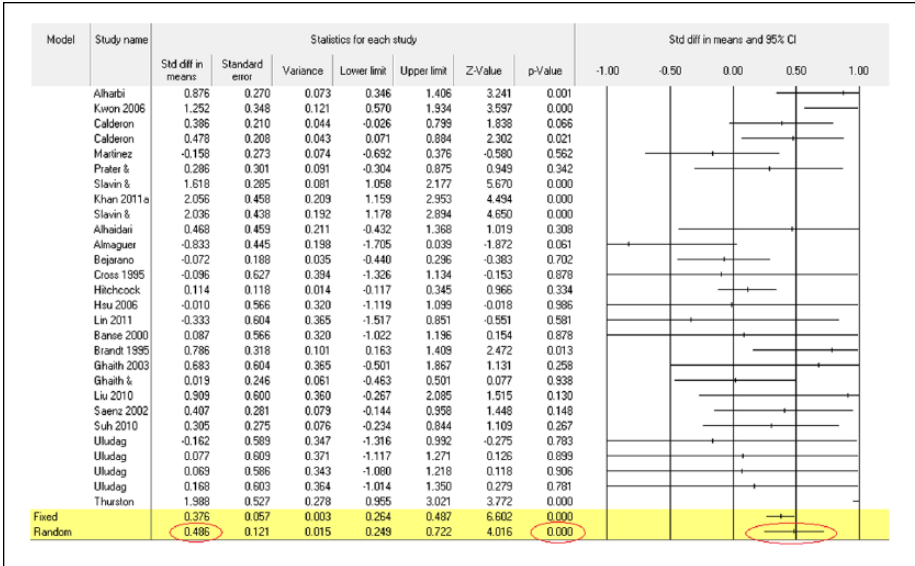
**Table 1.** Summary of Included Studies.

Lead author	Year	Publication type	Country	Construct	Design	Grade level
Alhaidari	2006	Dissertation	Saudi Arabia	Cooperative	Quasi-experiment	Elementary
Alharbi	2008	Dissertation	Saudi Arabia	Cooperative	Experiment	High school
Almaguer	2005	Journal	USA	Peer tutoring	Quasi-experiment	Elementary
Banse	2000	Dissertation	Burkina Faso	Collaborative	Quasi-experiment	High school
Bejarano	1987	Journal	Israel	Cooperative	Quasi-experiment	Middle school
Brandt	1995	Dissertation	USA	Cooperative	Quasi-experiment	High school
Calderón	1997	Technical report	USA	Cooperative	Quasi-experiment	Elementary
Cross	1995	Technical report	USA	Collaborative	Quasi-experiment	High school
Ghaith	2003	Journal	Lebanon	Cooperative	Quasi-experiment	High school
Ghaith	1998	Journal	Lebanon	Cooperative	Quasi-experiment	Middle school
Hitchcock	2011	Technical report	USA	Cooperative	Quasi-experiment	Elementary
Hsu	2006	Dissertation	Taiwan	Collaborative	Quasi-experiment	High school
Kwon	2006	Dissertation	South Korea	Collaborative	Quasi-experiment	High school
Lin	2011	Journal	Taiwan	Cooperative	Quasi-experiment	Middle school
Liu	2010	Journal	Taiwan	Collaborative	Quasi-experiment	Middle school
Martinez	1990	Dissertation	USA	Cooperative	Quasi-experiment	Elementary
Prater	1993	Journal	USA	Cooperative	Experiment	Elementary
Saenz	2002	Dissertation	USA	Peer tutoring	Quasi-experiment	Elementary
Slavin	1998	Technical report	USA	Cooperative	Quasi-experiment	Elementary
Suh	2010	Journal	South Korea	Collaborative	Quasi-experiment	Elementary
Thurston	2009	Journal	Catalonia	Peer tutoring	Quasi-experiment	Elementary
Uludag	2010	Dissertation	Jordan	Collaborative	Quasi-experiment	Middle/high school

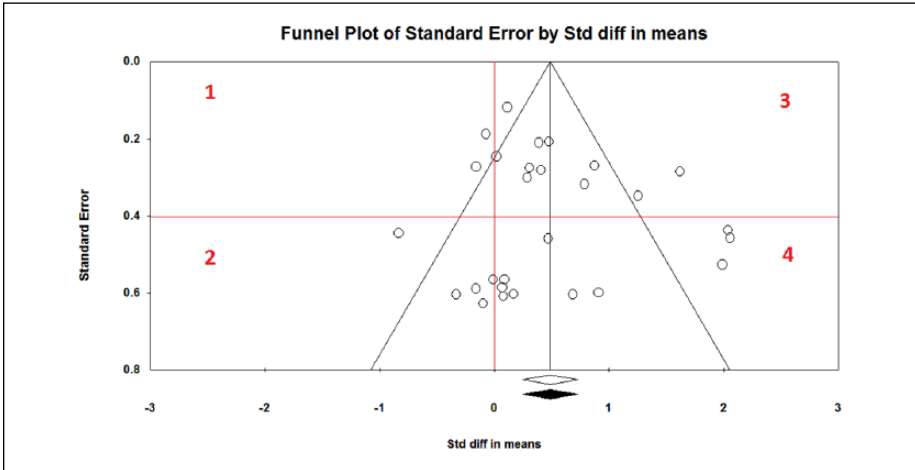
### *Publication Bias*

The mean effect size for published studies (.442,  $SE = .24$ ) is not much smaller than the mean effect size for unpublished studies (.524,  $SE = .142$ ), and this difference between the mean effect sizes of  $-.082$  provides a crude estimate of the upper bounds of potential publication bias.

Examining a funnel plot with effect sizes plotted against standard errors is one meta-analytically appropriate method of visually examining the distribution for the presence of publication bias. In this case, the standard error serves as a proxy for sample size, and because smaller samples are much more likely to lack the statistical power required to attain statistical significance, we look at the small-sample studies to detect publication bias (Lipsey & Wilson, 2001). If there is no such bias, we expect small studies with negative and null results (i.e., Quadrant 2 in Figure 2) to be as frequent as small studies with positive results. The funnel plot in Figure 2 includes black circles for studies imputed to achieve a symmetric distribution, the “trim and fill” technique (Lipsey & Wilson, 2001). In this case, there were no studies imputed to achieve a symmetric distribution, which is inconsistent with the possibility of publication bias. Similarly, the black diamond indicates that the anticipated mean did not change at all under publication bias conditions.



**Figure 1.** Forest plot of literacy outcomes.  
*Note.* See Table 1 for list of included studies with full author names and dates.



**Figure 2.** Funnel plot with missing studies imputed.

Egger’s regression intercept is a computational alternative to the visual inspection of the distribution (Sterne & Eggers, 2005). Generally, it is assumed that publication bias will be positive, that is, in the direction of statistically significantly positive effects, and because it provides a more conservative estimate of significance, the *p*

**Egger's regression intercept**

Intercept	1.02288
Standard error	0.76464
95% lower limit (2-tailed)	-0.54887
95% upper limit (2-tailed)	2.59462
t-value	1.33771
df	26.00000
P-value (1-tailed)	0.09629
P-value (2-tailed)	0.19257

**Figure 3.** Egger's regression estimates.

value of the single-tailed test at  $\alpha = .05$  is typically reported. The null hypothesis tests whether the ratio of the *effect size (ES)/standard error (SE)* is  $>0$ . The Egger's regression test presented in Figure 3 provided additional evidence of the improbability of publication bias in the literacy outcome distribution. The intercept was not significantly greater than zero for the one-tailed test (1.02,  $t = 1.338$ ,  $p = .096$ ) or the two-tailed test ( $p = .193$ ).

In conclusion, these analyses provided no evidence that publication bias was likely for the distribution of studies. In addition, several studies in the sample had null or negative effect size estimates; thus, it seems unlikely that the literature search failed to uncover those studies that for one reason or another simply were not published. Overall, the difference in means between published and unpublished studies, the funnel plot, and Egger's regression estimates all suggested that publication bias was not a factor in interpreting the data from the present meta-analysis.

### **Moderator Analysis**

The distribution of literacy effect sizes was heterogeneous, as indicated by the  $Q$  (97.135,  $p = .000$ ) and  $I^2$  (72.204) statistics; consequently, post hoc examination of moderator variables was conducted to explain some of this variability in effect sizes. Nonetheless, analysis of moderators was primarily motivated by a priori questions of interest, and findings remain qualified by the recognition that small differences may have been difficult to detect with the size of the sample used and confounding and lurking variables may have tempered any observed differences between sub-groups. Table 2 summarizes the results for measured variables reported in the 28 included studies, and the presence of significant bivariate correlations, analyzed as chi-square statistics, with other measured variables is indicated in the last column.

As indicated in the  $Q$ -between column, the distribution of literacy outcomes demonstrated few statistically significant moderators; however, three moderators were



**Table 2.** Summary of Moderator Analyses.

Moderator (sub-group)	No. in sub-group	Effect size point estimate	SE of Estimate	p value of estimate	Q-within of sub-group	I <sup>2</sup> of the sub-group	Q-between in the random effects model	Observed inter-correlation
Study quality							10.635 (p = .005)	Yes
High	17	.637	.144	.000	56.534 (p = .000)	71.7		
Medium	8	.328	.311	.291	31.991 (p = .000)	78.119		
Low	3	-.095	.173	.582	.170 (p = .981)	0.000		
Instrument type							1.107 (p = .575)	Yes
Researcher created	17	.411	.147	.005	35.743 (p = .003)	55.236		
Standard narrow	7	.338	.168	.033	50.012 (p = .000)	88.003		
Standard broad	4	.746	.420	.045	5.677 (p = .128)	47.156		
Post hoc researcher adjusted							9.058 (p = .003)	Yes
Yes	3	-.095	.173	.583	.170 (p = .918)	0.000		
No	25	.554	.129	.000	88.612 (p = .000)	72.916		
Construct							1.391 (p = .499)	Yes
Cooperative	14	.632	.168	.000	64.105 (p = .000)	79.721		
Collaborative	10	.376	.162	.02	9.94 (p = .355)	9.460		
Peer tutoring	4	.310	.414	.454	19.234 (p = .000)	84.403		
Setting							.023 (p = .879)	Yes
EFL	17	.504	.170	.003	45.017 (p = .000)	64.458		
ESL	11	.465	.184	.012	51.969 (p = .000)	80.758		
Segregated							.504 (p = .478)	Yes
Yes	5	.373	.135	.006	5.755 (p = .218)	30.942		
Other (not and unknown)	23	.518	.155	.001	91.38 (p = .000)	75.952		
Language of instruction							.274 (p = .872)	Yes
L1 (L1 only and bilingual)	9	.457	.168	.007	20.971 (p = .007)	61.853		
L2 only	8	.402	.247	.104	36.976 (p = .000)	80.976		
Unknown	11	.583	.258	.024	38.447 (p = .000)	73.99		
Culturally relevant							.101 (p = .951)	Yes
Yes	2	.433	.148	.003	.095 (p = .758)	0.000		
No	9	.474	.246	.053	51.54 (p = .000)	84.478		
Not U.S.	17	.504	.17	.003	45.017 (p = .000)	64.458		
Grade level							10.863 (p = .004)	Yes
Elementary	12	.539	.182	.003	59.259 (p = .000)	81.437		
Middle	6	-.007	.134	.961	2.841 (p = .724)	0.000		
High	10	.7	.204	.001	17.633 (p = .039)	49.047		
SES							.052 (p = .820)	Yes
Low	11	.516	.214	.016	45.141 (p = .000)	77.847		
Other (includes high and unknown)	17	.456	.147	.002	48.222 (p = .000)	66.820		
Student Hispanic							.005 (p = .945)	
Hispanic	10	.471	.18	.009	41.128 (p = .000)	78.117		
Other (Asian, Arabic, African, Pakistani, Lebanese)	18	.488	.172	.005	54.233 (p = .000)	68.654		
Student Asian							.697 (p = .404)	
Asian	6	.705	.32	.028	18.652 (p = .002)	73.193		
Other	22	.418	.125	.001	67.671 (p = .000)	68.967		

Note. EFL = English as a foreign language; ESL = English as a second language; SES = socioeconomic status.

statistically significant at the  $p = .05$  level: study quality, post hoc researcher adjusted, and grade level. Post hoc adjustments of literacy outcomes resulted in much smaller effect sizes on average ( $g = -.095$ ) than unadjusted ( $g = .554$ ), with the direction of the effect actually switching to support the comparison groups. Similarly, as study quality increased, so did the magnitude of the mean effect size, a finding that is somewhat counterintuitive. One might expect that high-quality designs would mitigate the influence of bias and accident resulting in lower effects on average; however, this is similar to the findings in other meta-analyses of peer-mediated instruction that reported that low-quality studies tended to report lower effect sizes (e.g., Keck et al., 2006). Finally, grade level proved to be a statistically significant moderator, mostly because middle school students showed much smaller gains ( $g = -.007$ ) than high school ( $g = .7$ ) or elementary ( $g = .539$ ).

Descriptive analysis of the moderators revealed interesting patterns among some of the non-significant moderators; nonetheless, for these variables that did not achieve statistical significance as moderators, either for lack of adequate power or because no moderating effect truly existed, tentative analysis indicates that future research might fruitfully examine the extent to which these variables influence the effectiveness of peer-mediated learning at improving literacy outcomes for ELLs. For example, standardized broadband measures produced larger gains (.746,  $p = .045$ ) than either researcher-created instruments (.411,  $p = .005$ ) or narrow-band standardized tests (.338,  $p = .033$ ). This finding is different from other research that has typically found that broadband standardized instruments generate the smallest effect sizes (e.g., Lipsey et al., 2012). Moreover, word-level outcomes (.702,  $p = .052$ ) were nearly twice the magnitude of text-level outcomes (.473,  $p = .004$ ), echoing the findings of the National Literacy Panel (August & Shanahan, 2007) that found that text-level outcomes tended to be smaller than word-level outcomes. Complex interventions for which peer-mediated learning was one of many tightly interwoven components (.633,  $p = .001$ ) reported outcomes nearly twice the magnitude of interventions that were purely peer mediated (.385,  $p = .012$ ), giving some support to claims that peer-mediated instruction works better when coupled with direct instruction (e.g., Cheung & Slavin, 2005; Genesee et al., 2005).

Finally, setting was not a significant moderator and the point estimates for ESL (.465,  $p = .012$ ) and EFL (.504,  $p = .003$ ) were quite similar.

## Discussion

The first question addressed in this study was whether peer-mediated learning promotes literacy outcomes for ELLs. Main effects analyses indicate that on average, peer-mediated learning approaches improved literacy outcomes by nearly half a standard deviation (.486,  $SE = .121$ ,  $p < .001$ ) when compared with teacher-centered or individualistic instruction. Moreover, there is no evidence that publication bias is present, suggesting that this estimated effect is robust to potential sources of bias (i.e., effects of outliers, cluster randomization, and publication bias).

To provide a sense of practical significance, one can imagine the impact an effect size of this magnitude would have on the performance gap between ELLs and non-ELLs on the fourth-grade NAEP Reading exam in 2008 (Lipsey et al., 2012). Non-ELLs had a mean score of 223 and ELLs had a mean score of 193 (Wilde, 2010), with a mean difference of 30 points. The standard deviation for all students on the 2008 NAEP reading exam was 34 points (Digest of Education Statistics, 2011); thus, the performance gap expressed as an effect size is approximately .882 SDs.<sup>4</sup> The mean effect size for peer-mediated learning of .486 SDs is more than half the magnitude of the achievement gap. This is not meant to say that using peer-mediated learning more consistently in the instruction of ELLs *will* cut the achievement gap in half as there are a number of factors that contribute to this performance disparity, including availability of appropriate program models for ELLs, validity concerns associated with using norm-referenced and standardized assessments with ELLs, and broad sociocultural influences that shape ELLs' identities and opportunities to learn. Rather, this comparison is intended to provide an intuitively interpretable estimate of the practical significance of an effect size of this magnitude.

The second research question asked what variables moderate the effectiveness of peer-mediated learning for ELLs, and moderator analyses revealed only three statistically significant moderators: study quality, post hoc researcher adjustments, and grade level. Two of these variables highlight the importance of carefully considering and measuring the quality of the research being synthesized. Pre-test differences, when uncontrolled, can exert noticeable effects on post-test scores, and other variables of study quality also exert a measurable impact on the results. This meta-analysis supports the findings of Keck et al. (2006), which reported that higher effect sizes were associated with higher quality studies. To clarify, study quality was not closely associated with the inclusion of dissertations and technical reports. Of the 17 studies coded as high quality, most were unpublished dissertations ( $n = 9$ ) or technical reports ( $n = 4$ ); conversely, most of the three studies coded as low quality were published in peer-reviewed journals ( $n = 2$ ).

The finding that the effectiveness of peer mediation is moderated by grade level was largely because studies of middle school ELLs reported much lower mean effect sizes ( $g = .007$ ) than studies with high school ELLs ( $g = .7$ ) or elementary ELLs ( $g = .539$ ). There are a number of possible explanations for this finding, including developmental differences that make socialization particularly difficult for middle school students. Middle school is a particularly difficult time for ELLs, in general. ELL enrollment growth is considerably larger in middle schools than in elementary schools (Capps et al., 2005); heterogeneity among ELLs in terms of socioeconomic status, years of U.S. residency, and native language and literacy proficiency is greater for middle school students than for younger students (Rubinstein-Ávila, 2003); middle school students learning English are confronted with more conceptually dense texts in subject areas (Cummins, 2007; ELL Working Group, 2009), and adolescent ELLs have fewer years to master these additional academic and language demands of U.S. schools than their elementary counterparts (Short & Fitzsimmons, 2007). Consequently, it is little surprise that ELLs drop out of school at higher rates than the general

education population beginning in middle school (Rubinstein-Ávila, 2003). It is unclear whether lower mean effect sizes for ELLs are part of the cause for or the result of these other issues, but the consequences are significant for both the learners and the schools who struggle to “retain the growing number of ELLs who are enrolled and to ensure their literacy development and content-rich education” (Rubinstein-Ávila, 2003, p. 123).

Descriptive analyses offer tentative information relevant to questions uncovered in the literature review. Standardized, broadband measures yielded larger effect sizes than researcher-created measures; although three of the four studies utilizing standardized, broadband measures were also coded as high-quality studies, it seems unlikely that this finding is closely associated with study quality. Seventeen studies were coded high quality, and nine of those studies used researcher-created measures, suggesting that high-quality studies were not more likely to select standardized measures. Word-level outcomes were larger on average than text-level outcomes, offering tentative support of the findings reported by the National Literacy Panel (August & Shanahan, 2006) that found consistently larger effects for word-level outcomes. This finding suggests that though peer-mediated learning appears to offer a direct effect at improving comprehension directly (.473,  $p = .004$ ), it may exert an even larger indirect effect on reading comprehension by promoting word-level outcomes that support long-term improvements in reading comprehension (.702,  $p = .052$ ). This finding might also indicate that text-level outcomes remain resistant to the kinds of short-term indicators of learning typically used in experimental studies. Future research should continue to explore the relationship between oral language and reading comprehension, and longitudinal research should examine the long-term effects of peer-mediated learning on word- and text-level outcomes to determine whether the pattern observed in this meta-analysis remains consistent over time. Complex interventions that utilized a peer-mediated learning component reported larger effects on average than interventions that utilized only peer mediation, providing tentative evidence in support of claims that peer mediation is most effective when coupled with direct instruction (Cheung & Slavin, 2005).

Overall, this study provides an estimate of the effectiveness of peer-mediated learning at improving literacy outcomes for ELLs, but questions about important moderating variables remain largely unanswered due to limitations in sample size. Study-level variables like quality and type of measure used appear important, but the effects of many relevant teacher-level, student-level, and instructional variables remain to be clarified by future research.

## **Next Steps**

The number of included studies in this meta-analysis was modest and limited the power of moderator analyses to detect potentially meaningful relationships among variables of interest. Future research should capitalize on the ongoing interest in peer-mediated learning for ELLs, and the increased number and quality of available studies would enable stronger conclusions about possible moderating variables (e.g., the

finding here of no difference between FL and SL settings). This study might inform future research aimed not just at determining the effectiveness of approaches to ELL instruction but also at efforts to determine the processes and pedagogical factors that make peer-mediated learning effective for ELLs. For instance, too few studies reported information about teachers (e.g., years of experience and certification to work with ELLs) to enable moderator analyses of teacher-level variables. Likely, this line of research will require researchers to examine academic and linguistic mechanisms, and both qualitative and quantitative designs will be needed to fully understand the nature and limitations of these mechanisms. In addition, variables not included in this meta-analysis (e.g., relationships of power and student motivation) might be examined in the future to understand their influence on the effectiveness of peer mediation for ELLs. Finally, digital and 21st-century literacies are increasingly essential for ELLs' participation in society and the workforce, and future research should explore the ways that peer-mediated learning facilitates ELL interaction during technologically mediated literacy practices.

This study addresses the long-standing call of the National Literacy Panel to provide empirical support for the effectiveness of particular instructional practices at improving literacy outcomes for ELLs, and the reported moderator analyses provide a useful starting point for more sophisticated analyses of the circumstances under which peer mediation is most effective with ELLs. The evidence documented in this study supports the rich qualitative evidence of previous syntheses that suggest that peer-mediated learning is an important component of quality classroom instruction for culturally and linguistically diverse students.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **Notes**

1. "Peer-mediated learning" is sometimes used to refer to a more specific subset of these approaches, especially when used with learning disabled students (e.g., Dion, Fuchs, & Fuchs, 2007).
2. Both interaction meta-analyses reported separate effect sizes for treatment and control groups, rather than the single effect size estimate provided by the standard mean difference effect size reported in this study. Consequently, the two effect sizes presented here are simply the difference between the average treatment and average control effect sizes actually reported. This is not directly comparable with the weighted estimates provided in the "Results" section of this study; instead, they provide a measure of the magnitude of the average difference between groups.
3. Table 2 provides a summary of primary moderator analyses.

4. See Lipsey et al. (2012) for a discussion of this and other methods for establishing the practical significance of effect sizes.

## References

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

- \*Alhaidari, M. S. (2006). *The effectiveness of using cooperative learning to promote reading comprehension, vocabulary, and fluency achievement scores of male fourth- and fifth grade students in a Saudi Arabian school* (Doctoral dissertation, The Pennsylvania State University, University Park). Available from Dissertations & Theses: A&I. (Publication No. AAT 3229376)
- \*Alharbi, L. (2008). *The effectiveness of using cooperative learning method on ESL reading comprehension performance, students' attitudes toward CL, and students' motivation toward reading of secondary stage in Saudi public girls' schools* (Doctoral dissertation, West Virginia University, Morgantown). Available from Dissertations & Theses: A&I. (Publication No. AAT 3328514)
- \*Almaguer, I. (2005). Effects of dyad reading instruction on the reading achievement of Hispanic third-grade English language learners. *Bilingual Research Journal*, 29, 509-526.
- American Federation of Teachers. (2006). *Where We Stand: English language learners*. Retrieved from <http://www.aft.org/pdfs/teachers/wwwsell1106.pdf>
- Aronson, E., & Patnoe, S. (2011). *Cooperation in the classroom: The jigsaw method* (3rd ed.). London, England: Pinter & Martin.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ballantyne, K. G., Sanderman, A. R., & Levy, J. (2008). *Educating English language learners: Building teacher capacity*. Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved from <http://www.ncela.us/files/uploads/3/EducatingELLsBuildingTeacherCapacityVol1.pdf>
- \*Banse, N. (2000). *The effect of cooperative task-oriented group activities on student achievement in English as a foreign language in large classes in Burkina Faso (West Africa)* (Doctoral dissertation, State University of New York at Buffalo). Available from Dissertations & Theses: A&I. (Publication No. AAT 9958242)
- \*Bejarano, Y. (1987). A cooperative small-group methodology in the second language classroom. *TESOL Quarterly*, 21, 483-501.
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. (2008). Performance trajectories and performance gaps as achievement effect size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1, 289-328.
- \*Brandt, F. J. (1995). *The effects of cooperative learning on achievement and self-esteem of high school students with learning disabilities* (Doctoral dissertation, Fordham University, New York City). Available from Dissertations & Theses: A&I. (Publication No. AAT 9530943)
- \*Calderón, M., Hertz-Lazarowitz, R., Ivory, G., & Slavin, R. E. (1997). *Effects of bilingual cooperative integrated reading and composition on students transitioning from Spanish to English* (Tech. Rep.). Baltimore, MD: Center for Research on the Education of Students Placed at Risk at Johns Hopkins University.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwanto, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute.

- Cheung, A., & Slavin, R. E. (2005). Effective reading programs for English language learners and other language-minority students. *Bilingual Research Journal*, 29, 241-270.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64(1), 1-35.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Cummins, J. (2007). Pedagogies of the poor? Realigning reading instruction for low-income students with scientifically based reading. *Educational Researcher*, 36(9), 564-572.
- \*Cross, J. B. (1995). *Effects of whole language immersion (WLI) on at-risk secondary students*. Retrieved from ERIC Document Reproduction Services. (ED392027)
- Digest of Education Statistics. (2011). *Table 126: Average reading scale score and standard deviation, by sex, race/ethnicity, percentile, and grade: Selected years, 1992 through 2011*. Retrieved from [http://nces.ed.gov/programs/digest/d11/tables/dt11\\_126.asp](http://nces.ed.gov/programs/digest/d11/tables/dt11_126.asp)
- Dion, E., Fuchs, D., & Fuchs, L. S. (2007). Peer mediated programs to strengthen classroom instruction: Cooperative learning, reciprocal teaching, classwide peer tutoring, and peer assisted learning strategies. *The Sage Handbook of Special Education* (pp. 450-459).
- Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33-56). Norwood, NJ: Ablex.
- ELL Working Group. (2009). The American Recovery and Reinvestment Act: Recommendations for addressing the needs of English language learners: Executive summary. Retrieved from: <http://www.stanford.edu/~hakuta/ARRA/ELL%20Stimulus%20Executive%20Summary.pdf>
- Fantuzzo, J. W., Riggio, R. E., Connely, S., & Dimeff, L. A. (1989). Effects of reciprocal peer teaching on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81(2), 173-177.
- Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2008). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions* (Report). Portsmouth, NH: Center on Instruction.
- Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19(1), 3-17.
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, 10, 363-386.
- Gersten, R. M., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children*, 66, 454-470.
- \*Ghaith, G. M. (2003). Effects of the learning together model of cooperative learning on English as a foreign language reading achievement, academic self-esteem, and feelings of school alienation. *Bilingual Research Journal*, 27, 451-474.
- \*Ghaith, G. M., & Yaghi, H. M. (1998). Effect of cooperative learning on the acquisition of second language rules and mechanics. *System*, 26, 223-234.
- Goodlad, S. (1998) Students as tutors and mentors, In S. Goodlad (Ed.), *Mentoring and tutoring by students* (pp. 1-17). London, England: Kogan Page.
- Gutiérrez, K. D., Larson, J., & Kreuter, B. (1995). Cultural tensions in the scripted classroom: The value of the subjugated perspective. *Urban Education*, 29(4), 410-442.
- Harper, C. A., & de Jong, E. J. (2009). English language teacher expertise: The elephant in the room. *Language and Education*, 23, 137-151.
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341-370.



- Hertz-Lazarowitz, R., Kirkus, V. B., & Miller, N. (1992). An overview of the theoretical anatomy of cooperation in the classroom. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: the theoretical anatomy of group learning* (pp. 1-16). New York: Cambridge University Press.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557-560. Retrieved from <http://www.bmj.com/cgi/reprint/327/7414/557?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=measuring+inconsistency+in+meta+analysis&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT>
- \*Hitchcock, J., Dimino, J., Kurki, A., Wilkins, C., & Gersten, R. (2011). *The impact of Collaborative Strategic Reading on the reading comprehension of grade 5 students in linguistically diverse schools* (Tech. Rep.). Washington, DC: National Center for Educational Evaluation and Regional Assistance.
- \*Hsu, M. H. (2006). *The effects of dramatic activities on reading comprehension of senior high school EFL students in Taiwan* (Doctoral dissertation, Texas A&M University, College Station). Available from Dissertations & Theses: A&I. (Publication No. AAT 3229360)
- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Retrieved from <http://www.co-operation.org/pages/cl-methods.html>
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89, 47-62.
- Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. M. Norris, & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91-131). Amsterdam, The Netherlands: John Benjamins.
- Khan, S. A. (2011). The effect of cooperative learning on academic achievement of low achievers in English. *Language in India*, 11(3), 232-243.
- Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: language acquisition in the classroom*. Hayward, CA: Oxford Pergamon.
- \*Kwon, S. (2006). *Roles of output and task design on second language vocabulary acquisition* (Doctoral dissertation, University of Florida, Gainesville). Available from Dissertations & Theses: A&I. (Publication No. AAT 3228763)
- Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching*, 33, 79-96.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford Univ Press.
- Lee, C. D., & Smagorinsky, P. (Eds.). (2000). *Vygotskian Perspectives on Literacy Research: Constructing Meaning through Collaborative Inquiry*. Cambridge, UK: Cambridge University Press.
- Lightbown, P. M. (2000). Anniversary article: Classroom SLA research and second language teaching. *Applied Linguistics*, 21, 431-462.
- \*Lin, C., Chan, H., & Hsiao, H. (2011). EFL students' perceptions of learning vocabulary in a computer-supported collaborative environment. *Turkish Online Journal of Educational Technology*, 10(2), 91-99.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, US Department of Education.

- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- \*Liu, T., & Chu, Y. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55, 630-643.
- Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259-278.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). New York: Academic Press.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In *Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies* (pp.407-453). New York: Oxford University Press.
- Martinez, L. J. (1990). *The Effect of Cooperative Learning on Academic Achievement and Self-concept with Bilingual Third Grade Students* (Doctoral dissertation). San Diego, CA: United States International University, School of Human Behavior, San Diego Campus.
- Mathews, R. S., Cooper, J. L., Davidson, N., & Hawkes, P. (1995). Building bridges between cooperative and collaborative learning. *Change*, 27(4), 34-40.
- McHugh, C. M., & Lipsey, M. W. (2007). *Calculations for correcting test statistics and standard errors for clustering*. Retrieved in personal communication February, 2009.
- McKeon, D. (2005). *Research talking points: English language learners*. Washington, DC: National Education Agency. Retrieved from <http://www.nea.org/home/13598.htm>
- Menken, K., & Antunez, B. (2001). *An overview of the preparation and certification of teachers working with limited English proficient (LEP) students*. Washington, DC: National Clearinghouse for Bilingual Education.
- National Center for Education Statistics. (2011). *The condition of education 2011* (NCES 2011-033). Retrieved from <http://nces.ed.gov/pubs2011/2011033.pdf>
- Oxford, R. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal*, 81, 443-456.
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes?. *Language Learning*, 44(3), 493-527.
- \*Prater, D. L., & Bermudez, A. B. (1993). Using peer response groups with Limited English Proficient writers. *Bilingual Research Journal*, 17, 99-116.
- Ramírez, J. D., Yuen, S. D., & Ramey, D. R. (1991). *Final report: Longitudinal study of Structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children* (Technical report). Washington, DC: United States Department of Education.
- Rohrbeck, C. A., Ginsberg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95(2), 240-257.
- Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin*, 134(2), 223-246.
- Rubenstein-Ávila, E. (2003). Facing reality: English language learners in middle school English. *English Education*, 35, 122-136.
- \*Saenz, L. M. (2002). *Peer-assisted learning strategies for Limited English Proficient students with learning disabilities* (Doctoral dissertation). Vanderbilt University, Nashville, TN.
- Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners – A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

- Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21, 43-69.
- Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247-284.
- Slavin, R. E., & Madden, N. A. (1998). Success for All/Esito para todos: Effects on the reading achievement of students acquiring English (CRESPAR Report No. 19). Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk. (ERIC Document Reproduction Service No. ED 423327.)
- \*Slavin, R. E., & Madden, N. (2001). Effects of bilingual and English-as-a-second-language adaptations of success for all on the reading achievement of students acquiring English. In R. E. Slavin & M. Calderón (Eds.), *Effective programs for Latino students* (pp. 207-230). Mahwah, NJ: Lawrence Erlbaum.
- Slavin, R. E., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4(4), 370-380.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *Benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University.
- Sterne, J. A. C., & Eggers, M. (2005). Regression methods to detect publication bias and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 99-110). Hoboken, NJ: John Wiley.
- \*Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning*, 26, 370-378.
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 22, 171-185.
- Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, 37(3), 285-304.
- \*Thurston, A., Duran, D., Cunningham, E., Blanch, S., & Topping, K. (2009). International online reciprocal peer tutoring to promote modern language development in primary schools. *Computers & Education*, 53, 462-472.
- Turnbull, M., & Arnett, K. (2002). Teachers' uses of the target and first languages in second and foreign language classrooms. *Annual Review of Applied Linguistics*, 22, 204-218.
- \*Uludag, A. (2010). *English interactive online (EIO) language policy and educational, cultural, and political concerns* (Doctoral dissertation, Arizona State University, Phoenix). Available from Dissertations & Theses: A&I. (Publication No. AAT 3410577)
- Vygotsky, L. S. (1978). *Mind and society: the development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Wilde, J. (2010, May). *Comparing results of the NAEP long-term trend assessment: ELLs, former ELLs, and English-proficient students*. Paper presented at the National Conference of the American Educational Research Association in Denver, CO.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving\*. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

## Author Biography

**Mikel W. Cole** is an Assistant professor at Clemson University. His research interests focus on English language learners, including: effective pedagogy for emergent bilinguals, preparation of teachers to work with culturally and linguistically diverse students, equitable policies for ELLs, and ways to bridge community and out-of-school language and literacy practices with school-based learning.