Research Ideas and Outcomes 4: e29375
doi: 10.3897/rio.4.e29375

PhD Thesis

Effectiveness of peer-mediated learning for English language learners: A meta-analysis

Mikel W Cole ‡

‡ Clemson University, Clemson, United States of America

Corresponding author: Mikel W Cole (mikel.w.cole@gmail.com)

Reviewed v1

Received: 28 Aug 2018 | Published: 29 Aug 2018

Citation: Cole MW (2018) Effectiveness of peer-mediated learning for English language learners: A meta-analysis. Research Ideas and Outcomes 4: e29375. https://doi.org/10.3897/rio.4.e29375

Abstract

Background

This manuscript reports the findings from a series of inter-related meta-analyses of the effectiveness of peer-mediated learning for English language learners (ELLs). Peer-mediated learning is a broad term that as operationalized in this study includes cooperative learning, collaborative learning, and peer tutoring. Drawing from research on interaction in second language acquisition, as well as from work informed by Vygotskian perspectives on socially-mediated second language learning, these meta-analyses synthesize the results of experimental and quasi-experimental studies.

New information

Included studies were conducted with language learners between the ages of 3 and 18 in order to facilitate comparisons to US students in K-12 educational settings. All participants were identified as ELLs, though learners in both English as a Second Language (ESL) and English as a Foreign Language (EFL) settings were included. Similarly, learners from a variety of language backgrounds were included in order to facilitate generalizations to the linguistic diversity present in US schools, and abroad. Main effects analyses indicate that peer-mediated learning is effective at improving a number of outcome types, including: language outcomes, academic outcomes, and social outcomes. Funnel plots and Egger's

regression analyses were conducted to examine the probability of publication bias, and it appears unlikely in most analyses. Moderator analyses were also conducted, where sample sizes were sufficient, to examine what measured variables were capable of explaining heterogeneity in effect sizes between studies.

Introduction

This dissertation presents the results of a meta-analysis of the effectiveness of peer-mediated learning for English language learners (ELLs)*1. Chapter One provides the background for and significance of the study. Chapter Two reviews the relevant first language (L1) and second language (L2) literatures for peer-mediation. Chapter Three details the methodology. Chapter Four presents the results of the various analyses, and Chapter Five discusses how the results address the research questions, as well as the limitations of and future research suggested by this meta-analysis.

Background

Currently, more than eleven million students in K-12 schools in the United States speak a language other than English at home, meaning that linguistically-diverse students now comprise more than 20% of the total school age population (National Center for Education Statistics 2011). Moreover, ELLs are the fastest growing population of students in U.S. schools (McKeon 2005), and their performance on high-stakes tests continues to lag behind the performance of their mainstream peers (Digest of Education Statistics 2009). As the population of linguistically-diverse students grows, ELLs are dispersing into states and schools historically unprepared to meet the unique needs of this group of students. Consequently, linguistically diverse students present an increasingly salient concern for schools across the country.

Not only is the population of ELLs rapidly growing and dispersing throughout US schools, ELLs are a remarkably heterogeneous group of students (Genesee et al. 2005, Rumbaut and Portes 2001, Solano-Flores 2008), and this heterogeneity has pervasive relevance for educators and researchers, alike. For example, report that more than three fourths of ELLs are born in the United States, but the foreign-born quadrant of the population comes from all over the world. The immigrant status of students, as well as related variables like length of residence in the United States, is important because of recent moves to require documentation of residency status in states like Alabama (e.g., Hispanic Interest Coalition of Alabama et al. v. Governor Robert Bentley et al. 2011), in order to qualify for Specially Designed Academic Instruction in English programs like Newcomer Centers (e.g., US Department of Education, 2016), to analyze country of origin differences among subgroups (e.g., Hispanics) for variables like parental education, socioeconomic status, and language proficiency, in order to determine the linguistic appropriateness of translated assessments for speakers of regional dialects (Solano-Flores 2008), and for classroom teachers to design culturally relevant instruction (e.g., Fradd and Lee 2003).

Statement of the Problem

School-level Silence: Sociopolitical Context and Program Models

ELLs are a linguistically diverse group of students, collectively speaking more than 400 languages (Francis et al. 2008); ironically, ELLs face pervasive messages of silence that deny access to and discourage use of their native languages, cultural practices, and cultural ways of knowing as learning resources. Historically, schooling in the United States has been actively structured to silence the linguistic capital of culturally and linguistically diverse students. As examples, the brutal assimilation of Native Americans in Boarding Schools during the nineteenth and twentieth centuries and the pervasive segregation of Mexican-Americans in the Mexican schools of the Southwest remain testimony to a doctrine of subtractive cultural assimilation and the persistence of a deficit perspective that views English language proficiency as the most important indicator of intelligence or academic potential (Gifford and Valdés 2006, Macedo 1994, Ruiz 2001, Valenzuela 1999, Wiese and García 1998). Eradicating students' languages was intentional and rationalized as a national security concern; in fact, this English-as-American argument found voice even amongst a few of the founding fathers, who were generally unrestrictive in early language laws and who resisted the establishment of a national language or language academy (Ovando 2003, Schmid 2001).

This historical legacy of silence persists in contemporary examples of lost opportunities to learn and instances of the ongoing denial of students' access to their own language and literacy practices (Gándara 2000, Gutiérrez et al. 2000, Valdés 2001, Valenzuela 1999). States are increasingly moving towards inclusion, or mainstreaming, for all students in response to legislation initially written largely for students with special education status. In practice, this "push to mainstream" means that ELLs find themselves in classrooms with teachers unprepared to teach them and increasingly in political environments that actively and explicitly discourage the use or study of their language and culture (Harper and Jong 2009). Arguably done for reasons of equity and to minimize the linguistic segregation of ELLs, some researchers counter that contemporary conceptions of equity underlying inclusion arguments represent conservative values that actually work to maintain the status quo and inequitable relations of power (Platt et al. 2003), effectively silencing students by placing them in classrooms where they will be positioned as deficient and where their linguistic and cultural capital will be structurally unavailable as learning resources.

Empirical evidence indicates that context influences student learning, and both the sociopolitical environment and the model of education provided to students contribute to ELLs' academic success (Gitlin et al. 2003, Gutiérrez et al. 1995, Ogbu and Simons 1998, Portes and Rumbaut 2001, Ramírez et al. 1991, Valenzuela 1999). Portes and Rumbaut (2001) report large-scale sociological data that demonstrates how "the context of reception" shapes a number of outcomes for immigrants, including academic success. Interestingly, the context of reception, which is partly a measure of attitudes in the receiving community toward particular immigrant groups, varies across immigrant groups (e.g., Asians versus Mexicans), within immigrant groups (e.g., Mexican versus Cuban versus Puerto Rican) and across time for the same immigrant group (e.g., Cubans in Florida). Similarly, Gándara et

al. (2003) indicate that state and local policy implementation often structurally positions ELLs inequitably; in the case of California, they argue that deficiencies in teacher training, facilities, curriculum and materials, and assessments contribute to lower ELL academic performance state-wide. Additionally, schools tend to operate under an epistemology that favors middle-class and White values, values that are often at odds with indigenous and cultural ways of knowing (Gutiérrez et al. 1995, Moll et al. 1992, Sleeter 2001). These studies indicate that students tend to learn better when they have access to their cultural knowledge and linguistic proficiencies and when linguistic, cultural, and racial differences are understood and respected; that is, students learn best when their human and cultural capital are given voice, not silenced.

Perhaps the most widely-researched aspect of linguistic capital present in the effectiveness literature for ELLs is language of instruction (Baker and Kanter 1981, Greene 1998, Ramírez et al. 1991, Rolstad et al. 2005, Rossell and Baker 1996, Slavin and Cheung 2005, Thomas and Collier 2004, Willig 1985). In this case, language of instruction refers to the language in which instructional services are provided, and it typically does not directly measure students' use of their native languages. Nonetheless, despite some notable disagreements in definitions of program models, methodologies, and interpretations of results (see for example the debate between Rossell and Baker (1996) and Greene (1998)), the clear consensus among these syntheses is that bilingual approaches that utilize students' native languages are at least as effective as monolingual approaches that utilize only English. Specifically, students acquire English proficiency and attain grade-level parity with non-ELLs in content areas faster when instructed at least part of the time in their first languages. However, there are typically corollary differences associated with each of the program models. For example, parents are often more involved in bilingual programs where they understand the language of instruction (Ramírez et al. 1991), thereby promoting secondary sources of academic success for linguistically-diverse students (e.g., assistance with homework).

Teacher-level Silence: Pedagogy, Preparation, and Dispositions

Current schooling practices continue to manifest messages of silence for linguistically-diverse students and teachers often reinforce these messages, creating classroom atmospheres like the following example where the teacher invokes a traditional "Initiate-Respond-Evaluate" discourse pattern that effectively stifles students: "I was struck by the silence when I entered the classroom. The teacher, positioned at the front of the traditionally organized room, began to speak. 'Where's the adjective in this sentence?'"(Gutiérrez et al. 2000, p.14). To clarify, this example is not exceptional; rather, this teacher-directed model of instruction is quite common, even in programs specifically designed for ELLs. A nationally-representative, longitudinal study of the effectiveness of three ELL program models (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 (Ramírez et al. 1991). Elsewhere, researchers argue that these

"monologic" spaces magnify cultural dissonance between students and teachers and work to reify inequitable power relations (Gutiérrez et al. 1995).

Unfortunately, most teachers of ELLs remain largely unprepared to provide the specialized learning this growing and heterogeneous group of students requires (Ballantyne et al. 2008, Harper and Jong 2009, Menken and Antunez 2001). In fact, most ELLs sit in classrooms taught by teachers that report feeling woefully unprepared to teach them (Ballantyne et al. 2008). Despite a well-established, affirmative obligation to ensure that students receive instruction capable of providing equitable access to the language of instruction (i.e., Lau v. Nichols 1974 and Castaneda v. Pickard 1981), most ELLs receive no specialized instruction at all (Ballantyne et al. 2008, Menken and Antunez 2001). Given a long history of state and local control of education and a move by some states to mandate English-only models of instruction, the kinds of language support services available to ELLs vary widely, ranging from full immersion in dual languages to just a couple of hours of pull-out support in English. Thus, the relatively few ELLs who receive services receive very different kinds of instruction, often with no indication that the variations of instruction are designed to match variations amongst types of ELLs (e.g., age, language proficiency, length of residence). Denying ELLs access to adequately trained teachers and accessible curricula ensures their silence and disempowerment throughout schooling and beyond.

Even in classrooms where talking and rich discussion are the norm, English learners are often silenced during class discussions because of inequitable distributions of power between students and teachers (Valenzuela 1999, Yoon 2008). Moreover, these power inequities often indicate the presence of beliefs and attitudes that inhibit students' academic success. What teachers believe about linguistically and culturally diverse students has a tremendous impact on student engagement and academic success, and it also shapes the nature of the instruction that teachers provide (Gándara 2000, Gutiérrez et al. 2000, Maxwell-Jolly 2000, Stritikus and Garcia 2000, Tijerino and Asato 2002). Teachers acting as "street-level bureaucrats" (Lipsky 2010) have tremendous power to shape the nature of the instructional services they provide, for worse or for better, by exploiting what Jim Cummins calls "cracks in the structure" (Cummins 2001). Not surprisingly, Echevarria et al. (2006) indicated that consistency of training and degree of implementation proved more influential to the effectiveness of ELL pedagogy than did regional differences. Baca et al. (1994), agree with Echevarria and associates that achieving high levels of implementation fidelity is crucial to program success; however, they report that changes of attitudes and practice amongst the teacher education faculty is difficult to accomplish. Taken together, this suggests that teacher preparation and certification to work with ELLs, familiarity and facility with the intervention, and beliefs and attitudes are important variables to consider in the effectiveness of any intervention intended for ELLs. Moreover, it suggests that teachers support or interrupt inequitable power relations through their internal orientations to students, and to linguistic diversity more broadly, so that silencing of ELLs occurs in ways that are not always readily observable.

Student-level Silence: Positioning, Identity, and Resistance

ELLs are also positioned towards silence by distributions of power at the student level, distributions at once informed by sociopolitical factors in the local context and driven by the reorganization of social strata and identity formation at the student level (Cummins et al. 2005, Duff 2001, Harklau 2000, Leki 2001, Morita 2004, Norton 1997, Norton and Toohey 2001, Oortwijn et al. 2008, Rollinson 2003, Valenzuela 1999). First, individual differences in language proficiency, culture, Length of residency, official language status (e.g., ELL, Former English Learner, Native Speaker), length of residency, and socioeconomic status all contribute to learners' identities and the way they are positioned in school and during classroom interactions. For example, Davies (2003) provides a sociolinguistic analysis of the pragmatic demands of joking for ELLs interacting with native speakers of ELLs, and the author describes differences in approaches for initiating interactions, as well as ELL self-reports of not initiating or participating in interactions because of perceived powerlessness when interacting with native speakers of English in English-speaking contexts. Similarly, Bonny Norton's construct of "investment' posits that individual learner characteristics are not immutable, and learners exercise agency as they position themselves in response to social ascriptions of place and power. Moreover, investment theory argues that individuals have multiple desires that interact with changes in context and relations of power that mediate individual motivation to participate in and ability to navigate social interactions. At every level, power mediates interactions for English language learners, especially when interacting with native speakers; and although language learners exercise autonomy, they are nonetheless constrained to some extent by the social positions made available in specific contexts.

Consequently, learners' identities and motivations affect academic success in dynamic and complex ways; sometimes peer influences and individual aspirations drive learners to pursue school success, and sometimes peer networks and individual responses to power inequities lead learners to resist schooling (Deyhle 1995, Iddings and McCafferty 2007, Kamberelis 1986, Kamberelis 2001, Lensmire 1998, Pavlenko and Norton 2007, Prior 2001, Talmy 2004, Talmy 2008, Valenzuela 1999, Voloshinov 1973). For example, Valenzuela (1999) reveals that even in a schooling context structured to systematically subtract the cultural and linguistic capital of students, social capital (i.e., the networks of relationships and resources contained within those network) varies considerably from student to student; some students had access to community and friendship support for schooling and tended to display a pro-schooling orientation, while other students participated in social networks that failed to support or actively rejected pro-school behavior. She argues that student identity and their access to caring, supportive individuals largely mediated their school success or failure. Importantly, student resistance to schooling is a key example of student autonomy, and like other identity and attitudinal positions, resistance can both promote or detract from positive orientations to schooling. Valenzuela recounts a school-wide, student-led walk-out of the high school she studied, and she documented the ways that perceptions of students' language and culture and deficiencies in teachers' preparation and school functioning contributed to the students' decision to stage the protest. Similarly, Deyhle (1995) describes Navajo students resistance

to the racism of their Anglo educators and the cultural and linguistic assimilation orientation of their schools. Interestingly, Deyhle claims that students most secure and supported in their indigenous identities were most likely to succeed in the Anglo-oriented culture of the schools, providing insight into the particular ways these students manage to resist the silencing of their cultural and linguistic capital while successfully navigating the challenges and demands of schooling.

In conclusion, it is worth reiterating the primary focus of the proposed study—to investigate the effectiveness of peer-mediated learning for improving language, academic, and social outcomes for ELLs. This framing of "the problem" is intended to show the multi-faceted ways that issues of power and inequity interact with learning for ELLs. However, it is not intended to advance a claim that interactive learning methods will solve all of the inequities that ELLs face. Cooperative learning alone is no panacea. Rather, it is the thesis of this statement of the problem that questions of educational effectiveness for ELLs demand attention to the ways that power and inequity interact with learning.

General Research Questions

Specifically, the meta-analysis reported in this dissertation seeks to answer the following two primary research questions. More specific questions and hypotheses are presented in Chapter 3, following the literature review in Chapter 2 that presents the case for examining specific variables of interest.

1. Is peer-mediated instruction effective for promoting academic or language learning for English language learners in K-12 settings?
2. What variables in instructional design, content area, setting, learners, or research design moderate the effectiveness of peer-mediated learning for English language learners?

Significance of the Study

The results of the proposed meta-analysis are intended to contribute to a growing literature on the effectiveness of specific instructional approaches for the fastest growing group of students in US schools, which contributes to an on-going discussion of equitable, high-quality instruction for ELLs. The results of the meta-analysis will offer a concise synthesis of multiple evaluation studies; specifically, standardized mean effect size estimates for language, academic, and attitudinal outcomes will provide systematic evidence of the effectiveness of peer-mediated instruction in key sets of learning outcomes for ELLs. Additionally, meta-analysis enables a systematic analysis of moderating factors that are important to consider when interpreting current and future evidence and when considering instructional decisions that might arise during implementation of peer-mediated learning in actual classroom contexts. As discussed in the Methods section, inclusion of studies conducted within the US and in other countries enables results to be broadly generalizable while allowing for analysis of the contribution of context as a moderator of effectiveness

(i.e., are results produced in English-as-a-Foreign-Language and English-as-a-Second-Language settings significantly different?).

Literature Review

Peer-mediated Learning

As indicated, the purpose of this meta-analysis is to synthesize the empirical literature on the effectiveness of peer-mediated learning for English language learners in K-12 settings; specifically, the meta-analysis computes main effects and identifies important mediators of effectiveness using experimental and quasi-experimental studies. Thus, the most relevant literature to review consists of previous meta-analyses and quantitative syntheses of peer-mediation; however, important qualitative studies, especially highly-cited reviews and syntheses are included to ensure that relevant theoretical, instructional, and empirical variables are not overlooked by focusing exclusively on experimental designs in the literature review.

What is Peer-mediated Learning?

In this paper, “peer-mediated learning” refers to an instructional approach that emphasizes student-student peer interaction, and it is intended to provide a contrast to teacher-centered or individualistic approaches to learning. In practice, peer-mediated learning includes a variety of approaches, each with supporting literatures that are typically distinct from one another. Specifically, this meta-analysis synthesizes three distinct varieties of peer-mediated learning: cooperative, collaborative, and peer tutoring, a distinction employed in previous syntheses (e.g., Cohen 1994, Hertz-Lazarowitz et al. 1992). As illustrated below, there are numerous precedents for treating these theoretically and practically different approaches as similar, if not synonymous terms (Cohen 1994, Johnson et al. 2000, Slavin 1996, Swain et al. 2002)*2.

The use of peer-mediated as a term to include multiple varieties of instruction not only emphasizes the similarities amongst these methods, it also reflects an underlying bias in this paper. The author currently sees a sociocognitive reading of Vygotskian theory as a conceptual common grounds between traditional second language acquisition models of L2 learner interaction and sociocultural models of L2 learner interaction, and Vygotskian perspectives on learning and cognitive development would describe all three approaches (i.e., cooperative, collaborative, and peer tutoring) as peer-mediated learning (see for example, Lantolf 2000)*3. Nonetheless, this paper does not assert that Vygotskian theory is explicitly or implicitly invoked by all of the authors or analyses included in this synthesis. Rather, it is posited that Vygotskian theory provides a heuristic lens that enables a coherent synthesis of varied literatures.

Thus, the treatment of several varieties of peer-mediated learning as similar does not imply that they are identical; rather, the intention is to focus on what they have in common, especially when compared to teacher-driven or individualistic approaches. However, for the

sake of clarity and to maintain an awareness of how the varieties do differ in meaningful ways, each of the three focal varieties of peer-mediated learning is briefly reviewed separately below.

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 1970's, with the most common outcome being some kind of 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" (Johnson et al. 2000).

A commonly definitive characteristic of cooperative learning approaches is the degree of structure (Oxford 1997, Slavin 1996); in fact, in this paper, degree of structure is the defining criterion that distinguishes cooperative and collaborative approaches. In general terms, cooperative methods emphasize carefully-structured groups, and students typically have well-defined roles to play. For example, in Jigsaw, students are each responsible for mastering one piece of the target content and typically report back to the team as the designated expert on that piece of the content. In order for the group to demonstrate mastery of the material, each person must adequately learn and then convey that individual piece of the overall content. In other forms of cooperative learning, students are assigned roles like Reporter and Researcher. Nonetheless, cooperative methods vary in their degree of structure, and Johnson et al. (2000) also analyze the eight methods of cooperative learning synthesized in their meta-analysis along a five-point continuum of structure ranging from what they call direct (i.e., structured) to conceptual (i.e., unstructured).

The description of Jigsaw above highlights another important component that defines cooperative methods—interdependence. The concept of interdependence is closely tied to group goals, and is intended to measure the extent to which individual members rely on each other for success. Several versions of cooperative learning suggest that students are motivated to participate in cooperative tasks because the group shares a common goal; however, researchers argue that commonly shared group goals are insufficient alone (e.g., Johnson et al. 2000, Slavin 1996). For instance, "free riders" may simply float along on the work of others under the sole condition of group responsibility for goal completion. Instead, these researchers theorize that there must also be individual accountability, and the combination of individual accountability and group goals contribute to the establishment of group interdependence. Nobody wins unless the group wins, and the group can only win if everyone demonstrates individual learning. Kluge (1999) suggests that there are several types of interdependence that can be established, including: team interdependence, resource interdependence, goal interdependence, reward interdependence, identity interdependence, and outside enemy interdependence; importantly, Kluge argues that these elements of interdependence do not have to all be present, and he suggests that practitioners may want to mix and match elements to suit their context and teaching style.

Collaborative Learning

A number of reviews treat cooperative and collaborative methods as if they are similar, if not identical, methods (e.g., Cohen 1994). However, this meta-analysis follows in the footsteps of researchers that see these two approaches as similar, but distinct, methods for engendering active, student-centered learning (e.g., Hertz-Lazarowitz et al. 1992, Mathews et al. 1995, Oxford 1997). In the most general sense, the two methods are quite similar; for example, they both structure learning by placing learners in small groups, and both approaches place explicit emphasis on encouraging peer-peer interaction and the active construction of meaning. Nonetheless, a more nuanced understanding of the two approaches reveals that the methods operate noticeably different from one another. Mathews et al. (1995) provide a nice distillation of some of the most important differences, including: role and degree of involvement of the teacher, relations of power between teachers and students, the necessity of training of students to work in small groups, and important differences in task construction and group formation. Essentially, collaborative learning represents a less-structured, more “democratic” set of approaches to small group learning. Cooperative methods, on the other hand, tend to emphasize highly-structured student roles and maintain more traditional teacher-student 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

While cooperative and collaborative methods dominate the field of peer-mediated learning approaches, it is important to recognize that there is considerable diversity of approaches within the field. Inclusion of peer tutoring approaches is intended to illustrate this diversity, while acknowledging that other peer-mediated approaches exist. Peer tutoring approaches also vary widely (see Goodlad 1998 for a more detailed discussion), though in general they utilize older, or more capable (i.e., academically successful) peers to provide one-one instruction for struggling learners. Although this can occur within grade levels, it is frequently used between grade levels, with older students being the tutor to younger tutees. 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-peer interaction and seek to foster active, rich discussion from all participants. Fantuzzo et al. (1989) explicitly tested several key components of reciprocal peer tutoring, a particular form of peer tutoring that emphasizes more equitable relations of power between peers and in which both partners are responsible for teaching the other partner, to determine which aspects of peer tutoring are responsible for its effectiveness. In particular, the authors attribute peer tutoring’s effectiveness to the combination of preparing to teach, actually teaching, and individual and joint accountability for learning success. Thus, they see group interdependence as an important part of its success in ways that are similar to cooperative learning, but they emphasize that the requirements of teaching activate particularly important cognitive and social learning processes; consequently, peer

tutoring adds an instructional element typically underemphasized or completely absent in cooperative and collaborative methods.

How Does Peer-mediation Promote Learning?

Slavin (1996) review of the state of the field of cooperative learning research, outlines four theoretical perspectives within cooperative learning alone. These four perspectives (motivation, social cohesion, cognitive development, and cognitive elaboration) are each associated with different interventions, contextual variables, and emphases on tasks and student roles; however, Slavin suggests that these differing theoretical orientations need not be seen as mutually-exclusive frameworks. Rather, they may be seen as interactive aspects of a complex process, and Fig. 1 presents his conceptual model as one way of integrating the objectives and emphases of these four perspectives.

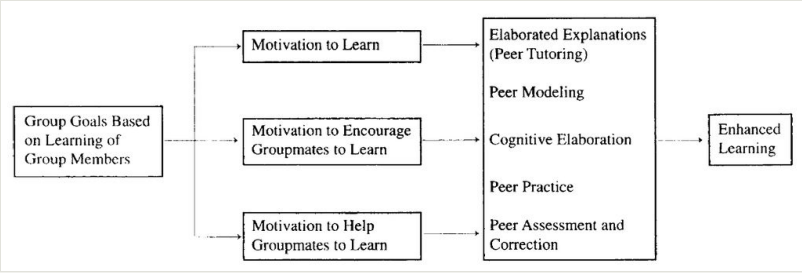


Figure 1. doi

Slavin's conceptual model of effective cooperative learning.

Thus, according to his model, group interdependence is a necessary component of enhanced learning through cooperation. Group interdependence is mediated by a number of motivational factors that contribute to several specific components of peer-mediated learning, including: elaborated explanations, peer modeling, and peer assessment and correction. It seems clear from the literature base of individual studies from which Slavin draws that not all of the individual components in the third box need be present for peer-mediated learning to be effective; rather, group interdependence fosters motivation which enables some of the individual components to occur. Slavin even acknowledges that limited evidence suggests that group interdependence need not always be present, but he argues that it is easiest to make cooperative methods effective when interdependence is present (Slavin 1996).

Cohen (1994) reviews the extant literature on the conditions for making small group instruction effective, and she identifies a number of factors that must be managed when implementing small groups. Unlike Slavin (1996), Cohen's analysis attempts to "move away from the debates about intrinsic and extrinsic rewards and goal and resource interdependence that have characterized research in cooperative learning" in order to focus on the kinds of tasks and kinds of discourses that promote learning. Similarly, as an alternative to the psychological focus of most cooperative learning research, she proposes a sociological heuristic that examines distributions of power between teachers and students

and between the students themselves. For example, in the oldest study of small group interaction that she reviews (i.e., Barnes and Todd, 1977), Cohen reports that some small groups engaged in destructive discourses (e.g., verbally attacking one another), and Cohen argues that students need both cognitive and social skills to participate effectively in small groups.

In addition to including aspects of power and equity, Cohen (1994) introduces the concept of productivity*4, which she distinguishes from related terms like effectiveness. Her key argument for preferring productivity is that the amount and kinds of interaction needed to promote achievement differ according to the kinds of outcomes desired. For instance, she argues that the kinds of interaction needed for successfully completing a worksheet collaboratively with a partner are very different from the kinds of interaction needed to foster higher-order or innovative thinking. Furthermore, she argues that the term productivity also enables analyses of equal-status interactions or the adoption of prosocial behaviors with members of social or ethnic out-groups in ways that effectiveness does not typically include. In particular, the idea that certain kinds of interactions may promote particular outcomes suggests that researchers should carefully analyze the kinds of interaction that occur, in addition to more superficial measures of intervention fidelity, and also that analyses should examine the relationship between the type of discourse and the type of outcome measured.

Empirical Evidence for Peer-mediated Learning

Both quantitative and qualitative evidence support the claim that peer-mediated learning is effective at promoting numerous kinds of outcomes; while the qualitative syntheses, with some exceptions like Slavin's narrative, "best-evidence" reviews (Slavin 1996, Slavin and Cooper 1999), presented below predominantly present theoretically-driven analyses, the quantitative syntheses tended to compare the effectiveness of particular models of peer-mediated learning or to compare the effectiveness of particular peer-mediated methods with different types of students.

Qualitative Evidence

Kluge (1999) offers a brief narrative synthesis of research on cooperative learning, and he reports positive outcomes on a variety of variables, including: use of high-order and strategic thinking, academic achievement tests, relationships with classmates, self-esteem, increased turn-taking when compared with whole-group instruction, and "discrete and integrative" language outcomes.

Cohen (1994) also provides a narrative synthesis, though her search and analytical methods are far more transparent and rigorous, and her included sample is much larger than the sample synthesized in Kluge (1999). Notably, she excludes studies that compare cooperative learning to traditional instruction, which is the precise contrast intended in this meta-analysis, opting instead to focus on studies that compare various forms of cooperative learning. She also favors studies conducted in classrooms, and systematically rejects lab studies if the task "bore no resemblance to classroom instruction". Finally, she

rejects discourse analyses, peer response groups for writing instruction, and peer tutoring; thus, her included sample is quite different from the studies that will be included in this meta-analysis. Nonetheless, she reports a theoretically-driven synthesis of both qualitative and quantitative studies that includes outcomes like induction of general principles of gears in a physics class, sophistication of debugging statements in a computer class, and more traditional measures of academic achievement. Cohen's analysis of the effectiveness of different models of cooperative learning is also more nuanced than Kluge's, and she reports that models that use both goal and reward interdependence tend to be more effective than models that employ either alone. Also, she finds that some models of cooperative learning may be more effective for particular groups of students. For example, Cohen reports that in some studies, White students performed best in the more competitive forms of cooperative learning, and Mexican American students seemed to perform better in traditional forms of instruction than cooperative forms (Cohen 1994).

Slavin (1996) is a narrative review of high-quality, experimental or quasi-experimental studies; thus, effect sizes are reported throughout the review, but the synthesis is conducted narratively and examines evidence for each of the four theoretical perspectives on cooperative learning illustrated in Fig. 1. This is a fairly typical of what Slavin calls "best evidence synthesis" (e.g., Slavin 1986, Slavin 1990), and it is included under the qualitative syntheses because of its heavy reliance upon and analysis of theory. Like Cohen (1994), Slavin reports that considerable evidence supports the motivational perspective's claim that group rewards used together with individual accountability produce the strongest group interdependence, and consequently, the strongest effects on achievement. For example, Slavin reports that studies with both group rewards and individual accountability (n=52) had a median effect size of .32SD compared to a median effect size of .07SD for studies that did not include both components (n=25). Similarly, Slavin presents results from a couple of studies that compared group rewards with individual accountability to individual components of accountability, and they consistently reported effect sizes that were much larger for the combined condition than conditions that just contained one component or another. However, like Cohen (1994), Slavin concedes that under certain conditions the group reward and individual accountability combination may not be necessary, including: complex tasks with more than one right answer, highly-structured peer interaction, or volunteer study groups. Nonetheless, he maintains that group rewards do not harm, and may actually improve, achievement results for those situations that do not require well-structured group interdependence. Finally, Slavin also confirms that some studies have reported stronger effects for certain types of students (e.g., Black over other ethnicities or those that prefer cooperative methods over those that prefer competitive methods); however, his evidentiary base is thin for these claims, and he ultimately argues that the results are mixed and inconclusive.

In another best evidence synthesis of qualitative and quantitative studies, Slavin and Cooper (1999) review the effectiveness of peer-mediated methods at promoting equitable relations amongst ethnic and racial groups in schools. Unlike the limited evidence regarding equity or racial diversity presented in Cohen (1994) and Slavin (1996), Slavin and Cooper provide a theoretical and empirical review of the rationale for using peer-

mediated methods to improve intergroup relations, and they argue that these approaches offer promise for helping schools shift from viewing diversity as a problem to solve to utilizing diversity as a resource for learning and socialization. Based largely on Gordon Allport's Contact Hypothesis*5, Slavin and Cooper argue that under the right conditions of equitable power relations, increased contact with members of racial and ethnic out-groups can promote inter-racial relations. Nonetheless, Slavin and Cooper claim that all too often "cross-ethnic interaction between students is superficial and competitive" (p. 649), and like Allport, they caution that poorly-structured interaction between groups can actually increase stereotypes and racial tensions. Moreover, Slavin and Cooper note that researchers like Cohen indicate the importance of establishing equitable conditions, arguing that not all students will have equal opportunities to contribute without direct teacher engagement in the process. However, with training and practice, teachers can actively and successfully promote the status of low-status students and foster an atmosphere of cooperation and respect.

Like many of the quantitative syntheses discussed below, Slavin and Cooper (1999) reviewed mostly quantitative, evaluation research from around the world for several different methods of peer-mediated learning, including: Student Teams-Achievement Divisions (STAD), Teams Games Tournament (TGT), Team-Assisted Individualization (TAI), Jigsaw, Group Investigation (GI), Learning Together (LT). The most common outcome was number of cross-racial friendship, though a few studies included related outcomes like cross-racial interaction during free time and positive ethnic attitudes. In a narrative, synthetic approach often called "vote counting" (Lipsey and Wilson 2001), Slavin and Cooper report that 16 of 19 studies demonstrated positive impacts on cross-racial friendships "when the conditions of contact theory are fulfilled" (p. 656).

Finally, another synthesis of cooperative learning explores the literature on the effectiveness of cooperative methods with Asian students in preschool to college settings (Than et al. 2008). Using a method much like Slavin's best evidence synthesis, the authors include only experimental and quasi-experimental studies in their formal analysis, but they draw heavily from the theoretical literature in their broader analyses. Unlike the work of other researchers who find ecological validity in the inclusion of multiple forms of peer-mediated learning (e.g., Cohen 1994, Rohrbeck et al. 2003), Than and colleagues explicitly exclude peer tutoring and collaborative approaches in order to maintain a tighter focus on the specific structures associated with cooperative methods, and the authors explicitly limited the range of outcomes to measures of academic achievement. Thus, the results are not informative of much of the literature included in this proposed meta-analysis, but the careful focus on the influence of cultural norms is uniquely informative. Specifically, the authors report that only seven of fourteen included studies demonstrated positive results, and they argue that cultural norms specific to Asian cultures make "Western…student-centered learning" approaches ineffective (p. 82). For example, Than and colleagues point to Asian students' preference for teacher-centered, lecture formats and teacher's frequent unwillingness to alter traditional roles and distributions of power as cultural norms interfering with key tenets of cooperative learning (i.e., active construction of knowledge and equitable distributions of power). Similarly, the authors claim that an Asian principle of

"survive in harmony" that dictates students make individual decisions without creating overt disagreements conflicted with the more argumentative, confrontational nature of "face-face promotive interaction" typical of peer-mediated learning methods (p. 84).

Quantitative Evidence

Unlike the theoretically-oriented syntheses presented above, the following quantitative reports offer more methodologically-focused syntheses that compare various models of cooperative learning to one another (Johnson et al. 2000, Johnson et al. 1981), components of a particular model of peer-mediated learning (Fantuzzo et al. 1989), or the effectiveness of a particular model with different student populations (Rohrbeck et al. 2003, Roseth et al. 2008).

Johnson et al. (1981) report the results of a meta-analysis dating to the time that the technique was first being developed; that is, the use of meta-analysis to study cooperative learning has strong precedent. The authors report the results of 122 studies and 286 independent effect sizes, dividing the effect sizes into the following categories: individualistic, interpersonal competition, cooperative, and cooperative with inter-group competition. Speaking directly to the fundamental question of this meta-analysis, Johnson and colleagues report that cooperative methods had a mean effect size .78SD larger than individualistic methods. In fact, the two forms of cooperative (with or without intergroup competition) performed equally well, on average. Cooperation with competition also produced consistently larger effect sizes than interpersonally-competitive methods, with a mean difference of .37SD. Thus, this early meta-analysis offers consistent support for the claim that peer-mediated approaches outperform individualistic learning approaches. The authors also conducted some tentative moderator analyses and argue that type of task (low versus high cognitive complexity), size of the cooperative group, task interdependence, duration of the study, year of publication, sample size, and journal quality are consistently significant predictors of effect size variation. Notably, subject area was not a significant predictor of effect size variation in any of the comparisons, suggesting peer-mediated approaches are useful across content areas.

More recently, Johnson et al. (2000) synthesize the results of 164 studies with 194 independent effect sizes; the authors selected from over 900 studies identified with the keyword "social interdependence", revealing a psychological orientation to the topic. Unlike the earlier 1981 meta-analysis just discussed, this meta-analysis attempts to provide a comprehensive comparison of the most widely-researched models of cooperative learning, including: Learning Together (LT), Academic Controversy (AC), Student-Team-Achievement-Divisions (STAD), Teams-Games-Tournaments (TGT), Group Investigation (GI), Jigsaw, Teams-Assisted-Individualizations (TAI), and Cooperative Integrated Reading and Comprehension (CIRC). Johnson, Johnson, & Stanne's meta-analysis reports separate effect sizes for each comparison group, as well as confidence intervals, to provide separate estimates of the effectiveness of each cooperative method against competitive or individualistic methods. Notably, of the eight approaches included in these analyses, all eight methods produce mean effect sizes superior to competitive (range g=.18 to g=.85)*6 and individualistic approaches (range g=.13 to g=1.04). Learning Together, developed by

Johnson and Johnson who co-authored the meta-analysis, consistently produces the largest effect sizes against both competitive and individualistic approaches, while the effect sizes for competitive and individualistic approaches are statistically equivalent. Like Johnson et al. (1981), this meta-analysis offers strong and consistent support that a wide-variety of peer-mediated approaches are more effective at producing academic achievement gains for school-aged children than more traditional, competitive or individualistic approaches.

Interestingly, Johnson et al. (2000) rate each cooperative method on a conceptual scale ranging from direct (very specific, well-defined techniques a teacher can learn quickly) to conceptual (conceptual frameworks teachers learn and use as a template to restructure lessons), a continuum similar to the concept of structure previously discussed. The coded score for each method is actually a composite of five different components of instruction, and the composite score is correlated with the effect sizes presented in the primary analysis. Degree of conceptualness correlates positively with effect sizes versus competitive ($r=.32$, $p<.001$) and individualistic approaches ($r=.46$, $p<.001$). This finding indicates that the more difficult to learn, but ultimately more flexible and dynamic forms of peer-mediated learning (i.e., more conceptual), approaches are more effective at promoting academic achievement. While this echoes the claim in Cohen (1994) that more conceptually-complex forms of group work are important for everything but the most rote forms of learning, Johnson and colleagues' use of a single effectiveness variable suggests that the authors operationalized achievement as effectiveness, not productivity as intended by Cohen.

One approach to determining the important components of an intervention is to systematically examine the contribution of important variables over the course of many separate replications (i.e., a meta-analysis); nonetheless, a more fine-grained approach is to design a study that explicitly tests various components individually and/or in multiple combinations, and Fantuzzo et al. (1989) is a well-cited example of just such a study. The study is a "component analysis" of Reciprocal Peer Tutoring, and although the participants were college-aged and are not directly comparable to the intended population for this meta-analysis, the insightful analysis of peer-mediated learning in an equitable form of peer tutoring is informative, nonetheless. One hundred college-age students were randomly divided into one of four conditions: dyadic structured format, dyadic unstructured, independent structured, and independent unstructured. The dyadic conditions consisted of two students that were randomly paired, and partners took turns being both tutor and tutee, which would rank fairly high in Cohen's construct of equitable power relations between students. The structured groups followed a specific test-item creating and sharing procedure, while the unstructured groups were provided topics for discussion that were related to the final exam taken by all participants. Initial examination of a number of variables (e.g., age, GPA, ethnicity) revealed no significant differences between groups. Analyses of covariance detected positive and significant effects for both dyadic ($F(1,95)=8.68$, $p<.005$) and structured conditions ($F(1,95)=7.06$, $p<.01$), providing a rigorous, direct test of two of the key theoretical components of peer-mediated learning: peer interaction and structure. This finding informs the debate within the field between those that

see strong structure as key (e.g., Slavin 1996) and those that argue that complexity and flexibility are more important (e.g., Cohen 1994), adding to the empirical support for the high-structure camp. Interestingly, the results also indicate a positive interaction between the dyadic and structure components for measures of psychological adjustment, course satisfaction, and a "generalizability" version of the assessment (though not the actual assessment) "due to the relative superiority of the DS [dyadic structured] condition" (p.176), a finding that also supports Slavin (1996) more nuanced argument that positive results can be found for various components of cooperative learning in isolation but that positive effects are more likely when multiple components operate in conjunction (e.g., interdependence and individual accountability).

Finally, two meta-analyses examine the impact of peer-mediated methods for particular groups of students. Rohrbeck et al. (2003) assess the effectiveness of peer-assisted learning (PAL) interventions for elementary-aged children. The authors intend PAL, like peer-mediated learning, to be an inclusive term for a variety of specific approaches including cooperative and peer tutoring; in fact, they claim that syntheses that examine only one form of PAL (e.g., Johnson et al. 1981) lack ecological validity since strict adherence to a particular form of peer-mediated learning fails to reflect the reality of classroom instruction. The authors included 81 studies with sufficient information to compute effect sizes, and after Windsorizing outliers and adjusting for small sample size, the mean main effect was ($d=.33$, $p<.0001$). Moderator analyses indicate that groups with more than 50% minority students produce larger effect sizes, on average, and students in urban settings tend to outperform students in rural settings. In this study, grade level and SES are weaker predictors of effectiveness, and content area is insignificant as a moderator variable. The authors also examine several variables of theoretical interest, and find that interventions that allow more student autonomy are more effective than those with less autonomy; however, the degree to which student roles are structured is not a significant moderator. Thus, this meta-analysis provides a nuanced understanding of structure that suggests that student autonomy and the motivation that accompany it exert a different effect than the contribution made by structured roles. Moreover, this finding lends support to Cohen (1994) claim that cognitive complexity and flexibility are superior to tightly-scripted roles for most kinds of learning. Programs that require interdependence are more successful than those that did not, but insufficient data exists to determine whether or not reciprocal peer roles are more effective than fixed roles.

Roseth et al. (2008) meta-analyze the results of 148 studies on the effectiveness of cooperative methods compared to individualistic and competitive approaches for early adolescents, extending the age-specific findings of Rohrbeck et al. (2003) to a slightly older group of students. Given the developmental emphasis placed on social relationships during this age period, the authors investigate the effects of cooperative methods and the social interdependence they foster on both academic outcomes and peer relationships, and the authors also directly test the relationship between peer relationships and academic achievement. As with previous meta-analyses conducted by Johnson and Johnson, the general pattern of results holds true; overall, cooperation is superior to competition (ES=.46SD) and individualistic approaches (ES=.55SD) at improving academic outcomes,

while competitive and individualistic interventions are statistically equivalent. Similarly, with peer relationship outcomes, cooperation is more effective than competitive (ES=.48SD) and individualistic approaches (ES=.42SD). For both sets of outcomes, removing low quality studies produces larger effect sizes, suggesting that low-quality studies may exert *downward* pressure on effectiveness estimates. Treatment fidelity is also a significant moderator in HLM moderator analysis. To examine the relationship between peer relationships and achievement, 17 studies that included both dependent variables and mean achievement are regressed on estimated mean peer relationship. When study quality is controlled, peer relationships account for approximately 40% of the variance in effect sizes. This finding offers unique theoretical insight into the key components of peer-mediated learning for a particular group of students, and the careful methodological attention to both theoretical and methodological variables provides a nuanced interpretation of theoretical questions about the effectiveness of peer-mediated learning for adolescents.

In conclusion, considerable qualitative and quantitative research supports the assertion that peer-mediated methods of instruction are more effective at promoting multiple kinds of outcomes than individualistic or competitive approaches. Despite decades of consistently positive research, a number of variables of instructional structure (e.g., size of group and composition of groups) and social interaction, as well as important learner (e.g., age) and methodological (e.g., design and measurement) variables, remain important foci of current and future research. In particular, few syntheses of the effectiveness of peer-mediation for particular kinds of students exist, and none of the syntheses discussed so far even mention specific issues involving linguistically diverse students. Thus, questions of whether, why, and under what conditions peer-mediation is effective for English language learners are the focus of the remainder of this literature review.

Peer-mediated Learning and ELLs

While much of the research regarding the effectiveness of cooperative learning reviewed so far is relevant for English language learners, it is important to keep in mind that English language learners are a distinct group of learners who, by definition, must master both academic and language objectives. Thus, when considering ELLs, it is essential to consider whether peer-mediated methods are effective for both academic and language outcomes, and as noted, language outcomes are largely ignored in the studies already reviewed. Moreover, it is essential to understand whether there are important linguistic mechanisms engaged during peer-mediated learning that are conceptually distinct from the more psychological and sociological mechanisms of peer-mediated methods just discussed in order to consider the relevant instructional and theoretical foci for L2 research.

Academic Rationale for Peer-mediated Learning with ELLs

Several recent syntheses of effective instruction for English language learners suggest that cooperative and collaborative models of instruction could be effective for promoting language, literacy, and content-area learning for ELLs (Allison and Rehm 2007, August and Shanahan 2006, Cheung and Slavin 2005, Genesee et al. 2005, Gersten and Baker 2000); however, these syntheses provide only tentative support for peer-mediated models of

education. First, these syntheses review multiple forms of instruction, not just peer-mediated methods. Second, the authors frequently report insufficient, or contradictory, evidence to draw strong conclusions.

For example, the National Literacy Panel on Language-minority Youth and Children (August and Shanahan 2006) reviews studies of literacy outcomes from instructional interventions that included complex, whole-school reform models like Success for All and small, researcher-created interventions targeting one aspect of literacy (e.g., fluency). Yet, across these disparate interventions the panel repeatedly favors approaches that emphasize direct, explicit instruction. In fact, the National Literacy Panel reviews only two studies that focus on peer-mediated learning approaches; while some of the complex approaches like SFA include a strong cooperative learning component, the results for these studies do not indicate whether it is cooperative learning that specifically contributed to the effectiveness of these programs. In fact, other work by Robert Slavin, the creator of SFA, explicitly argues that it is precisely the complex interaction of multiple components that makes these whole-school reform models effective (e.g., Cheung and Slavin 2005). Additionally, the National Literacy Panel Report includes a chapter on qualitative reports that consistently suggest cooperative learning is an important part of high-quality instruction for ELLs (e.g., Gersten and Baker 2000), though the conclusions drawn are tentative and carefully constrained. The authors of the National Literacy Report conclude only that “these attributes 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). Thus, the National Literacy Panel claims that mainstream research is largely sufficient to explain the effectiveness of peer-mediated approaches, and they claim more high-quality research is needed before firm claims can be made about peer-mediated methods, specifically.

Two other high-profile reviews (Genesee et al. 2005, Gersten and Baker 2000) synthesize research for a variety of instructional approaches, so much of the research they review is not directly applicable to this meta-analysis; however, like the National Literacy Panel, they represent some of the most authoritative, qualitative reviews of effective instructional approaches for ELLs.

Investigating effective instructional approaches for ELLs in elementary and middle grades, Gersten and Baker (2000) presents a “multivocal research synthesis” that utilizes focus-group interviews with educators, as well as a more traditional narrative review of experimental and descriptive evaluation studies. In a brief section on using “cooperative and peer tutoring approaches”, the authors suggest that both cooperative and peer tutoring approaches are effective, especially for “decontextualized language concepts with high degrees of cognitive challenge” (i.e., similar to the academic claim in Cohen 1994). However, the authors also report that these methods must be carefully tailored to the academic and linguistic needs of ELLs, that teaching ELLs is not simply “good teaching” (p. 461-464). In a larger and more systematic review of all empirical research conducted in the US since 1980 and reporting academic, literacy, or language outcomes, Genesee et al. (2005) provide syntheses for each of the three outcomes separately. In very brief

discussions of “direct” and “interactive” instructional approaches, the authors conclude that interactive approaches (i.e., peer-mediated) that also include carefully-targeted direct instruction are ideal, and they report that interactive approaches boost literacy and academic gains for ELLs.

No synthesis of the effectiveness of peer-mediated methods at improving academic outcomes for ELLs was identified in the review of extant literature for this meta-analysis, which is a strong warrant for the pursuit of this particular study. Consequently, only high-visibility, individual studies exist to document the academic rationale for using peer-mediated methods with ELLs. What Works Clearinghouse (WWC) reports results for only the most methodologically-rigorous studies, and taken as a whole, the inclusion criteria and analyses make the WWC site something like a quantitative synthesis of research; granted, WWC does not employ meta-analysis or any other formally-synthetic method to make claims across the included studies, so the actual reports are not truly syntheses. For ELLs, What Works Clearinghouse reports separately for the following outcomes: reading/writing, mathematics, and English language development. Of the studies included for reading and writing, only three use peer-mediated methods extensively, and all three demonstrate effectiveness at promoting literacy outcomes for ELLs. Two of the peer-mediated literacy interventions are complex models of which peer-mediated learning is one of multiple components (i.e., Success for All and Bilingual Cooperative Integrated Reading and Composition), and only one of the interventions focuses exclusively on the effectiveness of peer-mediation (Peer-assisted Learning Strategies, or PALS). WWC does not report any interventions for ELLs with math/science outcomes that meet its standards for inclusion, and language outcomes are discussed in the following section that presents the linguistic rationale for using peer-mediation with ELLs.

A closer look at the full reports of the three included interventions with literacy outcomes reveals that a number of important instructional variables differ across these interventions. For example, the most effective of the three interventions is BCIRC, and the WWC report is based almost entirely on Calderón et al. (1998). In the original report, the authors indicate that BCIRC combines extensive use of heterogeneous grouping with carefully-structured roles and procedures for small group interaction with direct instruction of academic and language objectives, thus supporting the claim that a combination of direct and interactive approaches is the most effective for ELLs (e.g., Cheung and Slavin 2005, Genesee et al. 2005). Moreover, the authors indicate that teachers were trained to make extensive use of the linguistic and cultural knowledge of the students; in fact, BCIRC is an intentionally bilingual approach that leverages students’ native language as an instructional resource. The authors attribute the effectiveness of peer-mediated learning for ELLs to “the verification of ideas; the planning of strategies for task completion; the discourse of politeness, consensus seeking, compromising; and the symbolic representation of other intellectual acts are enacted through peer communication” (p. 157). Thus, they offer the most nuanced explanation for the academic effectiveness of peer-mediation of any of the syntheses discussed, so far; however, as a single study, the claim lacks the statistical power and ecological validity that a synthetic finding would likely possess. Moreover, the

fact that the intervention contained several components that were not explicitly tested (e.g., Fantuzzo et al. 1989) also limits the explanatory power of the study.

Like BCIRC, Peer-assisted Learning Strategies (PALS) was evaluated for use in upper elementary ELL classrooms, and like BCIRC, only one evaluation study of the intervention meets WWC standards (Sáenz et al. 2005). PALS utilizes carefully-matched dyads that are taught to interact in structured ways with texts and each other. Importantly, both students take turns being the tutor and tutee despite structuring ability difference into the groupings. In the original study, the authors suggest that PALS is likely to be especially effective for ELLs because of increased opportunities for language production, individualized reading instruction, and practice with academic tasks like summarizing and making predictions. Importantly, the report on PALS in the original study also offers a linguistic rationale for the academic effectiveness for ELLs. While the next section will discuss the linguistic rationale for using peer-mediation with ELLs, most of the outcomes discussed in that section will be language outcomes. Thus, Sáenz and colleagues make an important point regarding the effectiveness of peer-mediated methods with ELLs—the linguistic benefits of peer-mediation likely contribute to both linguistic and academic outcomes.

Linguistic Rationale for Peer-mediated Learning with ELLs

While no formal synthesis of the effectiveness of peer-mediation at promoting academic outcomes for ELLs exists, several theoretical, qualitative, and quantitative syntheses of the effectiveness of peer-mediated learning at promoting language outcomes for ELLs exist. Thus, there is a considerably stronger rationale for using peer-mediation to promote language learning for ELLs than for promoting academic outcomes, and this is a key assertion because it is precisely English language proficiency that defines this group of students. Thus, peer-mediated learning offers promise not only as an effective approach for promoting the academic success of ELLs, it may also be an important tool for removing the fundamental barrier to equal access to the mainstream school curriculum the term ELL is intended to identify: English language proficiency*7.

Oxford (1997) provides a narrative synthesis of three strands of “communicative teaching” in the language classroom that closely mirror the key constructs of this meta-analysis: cooperation, collaboration, and interaction; and she suggests that these strands are related but theoretically distinct*8. Like this meta-analysis, Oxford distinguishes cooperative learning from collaborative primarily in the degree of structure embedded in the activity and the extent to which learner roles are prescribed and consistent across groups and events, whereas collaborative learning tends to be less structured. Like Slavin (1996), she also asserts that positive interdependence must be structured into the activities if cooperative methods are to be effective; however, for collaborative research, she draws a new theoretical distinction. Oxford asserts that collaborative methods have their roots in Dewey’s social constructivism and Vygotskian social psychology, and she asserts that constructs like mediation, scaffolding, and cognitive apprenticeship are central for collaborative theorists. Unlike collaborative approaches, the key objective is not to stimulate motivation through the construction of interdependence among learners; rather, the goal is to incorporate students into a community of learners. Interaction, on the other hand, draws

from a predominantly linguistic base, and this strand draws heavily upon constructs like comprehensible input, comprehensible output, and Michael Long's Interaction Hypothesis. The basic idea is that interaction promotes language learning by providing opportunities for students to modify output in ways that maximize the production of the comprehensible input that drives language acquisition.

Whereas, cooperation is high-structure and collaboration is low-structure in her scheme, she finds that interaction studies vary widely on this variable. Importantly, Oxford identifies a number of additional variables that influence the effectiveness of interactive approaches; including learner variables (i.e., willingness to communicate and learning styles) and grouping dynamics (i.e., group cultures and physical arrangement of the classroom).

In a narrative review of both qualitative and quantitative empirical research, Swain et al. (2002) review the effectiveness of "peer-peer" dialog at promoting listening, speaking, reading and writing language outcomes. Swain and colleagues adopt a Vygotskian lens on language learning that suggests peers can support each other's language acquisition by working within the zone of proximal development to enable language production and comprehension beyond what they might be able to accomplish individually, and agreeing with Oxford, the authors characterize these interactions as collaborative. It is worth noting that many of the studies reviewed are of French immersion students (i.e., English-speaking Canadian students learning French as a second language) and Spanish-learners; thus, the results are informative but not directly applicable to this meta-analysis.

In particular, the findings reported in Swain et al. (2002) are based on microgenetic analyses of language learning as it occurs in interaction, and data sources tend to feature transcripts, as well as pre-/post- measures of learning. For example, Swain et al. (2002) report that peer feedback during reading and writing activities is instrumental, and several important mechanisms are discussed, including: reformulations and recasts, collaborative planning/drafting/revising, metalinguistic talk, finding the main idea, vocabulary comprehension, etc. Swain and colleagues report that in an interesting series of studies by Storch, the nature of peer feedback proved particularly important, and the author rated the feedback on two scales that are reminiscent of mainstream peer-mediation constructs already discussed—equality (similar to degree of authority or power) and mutuality (similar to interdependence). Storch reported that the more collaborative the dyads were on these two scales, the more opportunities for and success with language learning occurred. In the terms previously used in L1 research this would mean that conditions of equity and high interdependence produce the largest gains. Swain, et al. also note that for these approaches to be maximally beneficial, students must be explicitly taught how to interact effectively with one another, and for language learners this includes instruction in particular grammatical structures and vocabulary, as well as turn-taking norms, strategies for persuasion, and pragmatic norms for politeness.

Two recent meta-analyses of the effectiveness of interaction at promoting L2 learning outcomes offer additional warrant for using peer-mediated learning methods with ELLs; and in addition to providing overall estimates of the effectiveness of peer-mediated L2 learning, they provide considerable insight into important factors that mediate effectiveness. The first

of the two meta-analyses (Keck et al. 2006), included 14 experimental studies conducted between 1980 and 2003. The meta-analysis reported a large overall mean effect size for peer-mediated learning greater than a standard deviation (d=1.12), as compared to a more moderate overall effect size (d=.66) for the comparison/ control groups. Participant characteristics like first language and level of L2 proficiency were not important variables in the effectiveness of the interventions, and the type of measure used (i.e., institutional grade level, researcher-created measure, or standardized assessment) did not affect the magnitude of the reported effect size. Moreover, the authors found that task-type (i.e., jigsaw, information gap, and narrative) was not an important moderator, and lexical outcomes (d=.90) and grammatical outcomes (d=.94) were also of similar magnitudes. However, the extent to which the task required the use of target forms (i.e., past tense verb constructions) was an important predictor of both immediate and delayed post-test performance. Overall, the more that students had to use the target form to correctly accomplish the task, the larger and more durable were the effects. Moreover, the authors report that interventions that encouraged "forced output" of the participants proved more effective (d=1.05) than interventions that merely allowed the possibility of participant output (d=.61), a finding that offers tentative support for the claim that degree of participation among participants may be an important factor in language learning.

Mackey and Goo (2007) is intended to provide an update to the Keck et al. (2006) meta-analysis. Mackey and Goo included all 14 of Keck, et al.'s studies, and an additional 14 studies for a total of 28 included studies. Twelve of the additional studies were published after the 2002 cut-off date of the previous meta-analysis, indicating ongoing and increased interest in the field. Overall, the Mackey and Goo meta-analysis reports a large effect size for peer-mediated learning (d=.99) compared to a much smaller effect size for the comparison groups (d=.38). Additionally, the authors report that peer-mediated learning remains effective beyond post-test; like Keck, et al., these authors report that peer-mediated learning is even more effective at the first delayed post-test (d=1.02). Despite considerable variability in participant language background, language ability, and instructional setting (i.e., SL, immersion, and FL), no significant differences in overall effectiveness are reported*9. Similarly, no differences are reported for length of treatment or other study design characteristics (e.g., experimental versus quasi-experimental). However, studies conducted in the laboratory (d=.96) report larger effects on average than those conducted in classroom settings (d=.57). Also, the type of dependent measure proves to be an important moderator of the variability in effectiveness; prompted response (d=.24) is the least effective, while open-ended prompted production (d=.68) and closed-ended prompted production (d=1.08) tasks are much more effective overall, adding some support to the claim that cognitively complex tasks are the most effective.

These syntheses provide compelling evidence that peer-mediated methods are effective at promoting a wide variety of language outcomes for second language learners, though many issues raised in the L1 research remain largely unanswered in the L2 literature. For instance, ELLs are a highly heterogeneous population (i.e., language background, prior schooling, SES, race/ethnicity, age of arrival, and length of residence), but there is little research that discusses with which ELLs peer-mediated methods might be most effective,

though both Keck et al. (2006), Mackey and Goo (2007) suggest that a small subset of these are not significant moderators (i.e., language background and language ability). Nonetheless, the studies by Oxford (1997) and Than et al. (2008) raise the question of whether cultural norms might mediate the effectiveness of these interventions for linguistically and culturally diverse students. At best, individual studies have attempted to account for these variables by controlling for them during assignment and/or measuring and controlling for differences following assignment, though few studies did either.

Type of task matters in both the theoretical and empirical L1 and L2 literatures reviewed so far, but neither the qualitative nor the quantitative literatures offer much feedback about which kinds of tasks are best for which types of language or academic outcomes for ELLs. Importantly, Keck et al. (2006) indicate that the more the use of the target structures measured at post-test were required for participation in the activity, the greater the gains; nonetheless, this commonsense connection between the degree to which the assessment is related to the intervention is a well-recognized phenomenon, and performance on distant, broad-band measures remains notoriously difficult to improve (e.g., Bloom et al. 2008, Slavin and Madden 2011). Similarly, the moderating effect of contextual variables (e.g., foreign language vs. second language, segregated vs. integrated, program model) is rarely measured directly, though again, both of the language-oriented meta-analyses suggest that a small subset is unimportant (i.e., language setting and program model). Issues of equity and power relations between students appear important in the qualitative literature but are not discussed in the quantitative literature.

Summary and Unanswered questions

Peer mediated methods have consistently proven effective at promoting academic, social, and language outcomes with a wide variety of first- and second-language students in a wide variety of contexts, lending support to Slavin (1996) claim that cooperative learning is one of the greatest successes in the academic evaluation literature. When compared to individualistic or competitive models, cooperative and other peer-mediated methods typically produce much larger gains. Nonetheless, researchers disagree about the influence of a number of key variables, which are summarized in Table 1 below. Notably, there are a number of similarities between the L1 and L2 literatures, though the research is not completely congruent between these two fields. As discussed in more detail below, L2 researchers do not always measure variables important in the L1 literature, and L2 researchers are often focused on aspects of language acquisition generally not researched in the L1 literature.

Table 1.

Summary of key variables from literature review.

VARIABLE	L1 Research	L2 Research
Peer-mediated Method Matters	Cohen 1994, Johnson et al. 1981, Johnson et al. 2000, Slavin 1996, Slavin and Cooper 1999	Oxford 1997

VARIABLE	L1 Research	L2 Research
Peer-mediated Method does not Matter	Kluge 1999, Rohrbeck et al. 2003, Slavin and Cooper 1999	Genesee et al. 2005 Gersten and Baker 2000 Keck et al. 2006 Swain et al. 2002
High-structure is Best	Fantuzzo et al. 1989, Johnson et al. 2000, Rohrbeck et al. 2003	Calderón et al. 1998 Sáenz et al. 2005
Low-structure is Best	Cohen 1994, Johnson et al. 2000, Rohrbeck et al. 2003	
Interdependence is Needed	Cohen 1994, Slavin 1996, Johnson et al. 1981, Johnson et al. 2000, Rohrbeck et al. 2003, Than et al. 2008	Oxford 1997 Swain et al. 2002
Interdependence is not Needed		Swain et al. 2002
Content Area Matters		
Content Area does not Matter	Johnson et al. 1981, Rohrbeck et al. 2003	
Age of Students is Important	Rohrbeck et al. 2003, Roseth et al. 2008	
Age of Students is not Important		
Ethnicity of Students Matters	Cohen 1994, Rohrbeck et al. 2003, Slavin and Cooper 1999, Than et al. 2008	
Ethnicity of Students does not Matter		
Language Proficiency (i.e., L1 or L2) of Students Matters		Genesee et al. 2005 Gersten and Baker 2000 Swain et al. 2002
Language Proficiency (i.e., L1 or L2) of Students does not Matter		Mackey and Goo 2007
Culturally-relevant Instruction Matters	Than et al. 2008	Calderón et al. 1998, Oxford 1997
Culturally-relevant Instruction does not Matter		
SES of Students Matters	Rohrbeck et al. 2003	
SES of students does not Matter		
Size of Group Matters	Johnson et al. 1981	
Size of Group does not Matter		
Equality of Power among Students Matters	Rohrbeck et al. 2003	Oxford 1997 Swain et al. 2002
Equality of Power among Students does not Matter		
Duration of Intervention Matters	Johnson et al. 1981	

VARIABLE	L1 Research	L2 Research
Duration of Intervention does not Matter		Mackey and Goo 2007
Setting (i.e., segregated, cooperative, ESL or EFL, lab or classroom, urban or rural) Matters	Rohrbeck et al. 2003, Slavin and Cooper 1999	Mackey and Goo 2007
Setting does not Matter		Mackey and Goo 2007
Journal Quality Matters	Johnson et al. 1981 Roseth et al. 2008	
Journal Quality does not Matter		
Sample Size Matters	Johnson et al. 1981	
Sample Size does not Matter		

First, researchers disagree about the importance of the particular method, whether cooperative, collaborative, peer tutoring, or some set of specific approaches (e.g., Jigsaw, Learning Together, STAD, TGT). The clearest distinction appears to be between L1 researchers that generally agree the method matters (though which method is ultimately superior remains debatable) and L2 researchers that typically do not report differences between specific methods. To be fair, this largely reflects the nascent state of L2 research, and many of the studies listed in Table 1 did not make clear distinctions amongst methods and simply grouped them all together as peer-peer or cooperative approaches. On the other hand, Swain et al. (2002) explicitly grouped multiple methods together in their synthesis, providing a theoretical rationale that it is the presence of peer-peer dialog that matters most for L2 learners. Although L1 research would suggest that specific methods vary considerably in their effectiveness at promoting academic and social outcomes, the question of which peer-mediated method is most effective for ELLs remains largely unaddressed.

While considerable debate exists within and across L1 and L2 literatures about which peer-mediated method is most effective, there is strong consensus that more structured approaches produce bigger gains than less-structured approaches. Despite this strong consensus, theoretical (Cohen 1994) and empirical (Johnson et al. 2000, Rohrbeck et al. 2003) grounds exist to challenge this claim. Similarly, overwhelming support concludes that establishing interdependence promotes learning gains, though Swain et al. (2002) report ambivalent findings on this variable. Language proficiency, the cultural-relevance of the instruction, and the equality of power relations among students appear to be important variables for L2 learners, though the research base for these claims is not as substantive as for other variables. Similarly, age, ethnicity, and SES appear to be moderators for effectiveness, though L2 research can neither support nor challenge this claim for ELLs. Finally, study quality variables (i.e., duration of intervention, journal quality, sample size) also suffer from ambivalence or few studies in the literature base; so claims for these variables are also tentative, and additional research could potentially bolster the warrant for making claims about the importance of these variables as moderators for the effectiveness of peer-mediated approaches.

Notably, several variables of equity mentioned in the Statement of the Problem in Chapter 1 appear to be missing, or at least largely ignored, in the above list, including: adequate facilities, context of reception, preparation of teachers to work with ELLs, attitudes and beliefs of teachers towards ELLs, relations of power between teachers and ELLs, and length of residence of ELLs. To the extent possible, these variables will also be coded when reviewing studies for inclusion in this meta-analysis. However, the absence of these variables from the extant literature probably supports the assertion that the field of peer-mediated learning studies for ELLs remains largely driven by psychological theory and that sociological perspectives remain underrepresented (e.g., Cohen 1994, Firth and Wagner 1997), and this meta-analysis hopes to bridge the more traditional focus on intervention effectiveness with these variables of power and equity.

Methods

Research Questions

Chapter 1 presented the two fundamental research questions driving this meta-analysis; however, as indicated in the literature review in Chapter 2, there are a number of substantive theoretical, instructional, and methodological variables of potential interest. Consequently, formal hypotheses regarding the key variables of interest are presented below.

- Is peer-mediated instruction effective for promoting language, academic, or attitudinal learning for English language learners in K-12 settings?

a. Hypothesis 1a: Test of H_A: Interventions testing the effectiveness of peer-mediated forms of learning against teacher-centered or individualistic control groups report language outcome effect sizes that are significantly larger.

b. Hypothesis 1b: Test of HA: Interventions testing the effectiveness of peer-mediated forms of learning against teacher-centered or individualistic control groups report academic outcome effect sizes that are significantly larger.

c. Hypothesis1c: Test of H_O: Interventions testing the effectiveness of peer-mediated forms of learning against teacher-centered or individualistic control groups report attitudinal outcome effect sizes that are not significantly different.

- What variables in instructional design, content area, setting, learners, or research design moderate the effectiveness of peer-mediated learning for English language learners?

a. Hypothesis 2a: Test of HO: Interventions testing the effectiveness of cooperative, collaborative, and peer tutoring approaches report effect sizes that are not significantly different.

b. Hypothesis 2b: Test of HO: Interventions testing the effectiveness of peer-mediated approaches in English-as-Second Language (ESL) and English-as-a-Foreign Language (EFL) settings report effect sizes that are not significantly different.

c. Hypothesis 2c: Test of HO: Interventions testing the effectiveness of peer-mediated approaches in elementary, middle school, and high school settings report effect sizes that are not significantly different.

d. Hypothesis 2d: Test of HO: Interventions testing the effectiveness of peer-mediated approaches in laboratory and classroom settings report effect sizes that are not significantly different.

e. Hypothesis 2e: Test of HO: Interventions testing the effectiveness of peer-mediated approaches as part of complex interventions and those testing just peer-mediation report effect sizes that are not significantly different.

f. Hypothesis 2f: Test of HO: Interventions testing the effectiveness of peer-mediated approaches with students from different language backgrounds report effect sizes that are not significantly different.

g. Hypothesis 2g: Test of HO: Interventions testing the effectiveness of peer-mediated approaches with students from high- and low-SES backgrounds report effect sizes that are not significantly different.

h. Hypothesis 2h: Test of HA: High-quality studies report effect sizes that are significantly larger than low-quality studies.

i. Hypothesis 2i: Test of HA: Studies of longer duration report effect sizes that are significantly larger than short-duration studies.

- In what ways do select issues of power and equity impact the effectiveness of peer-mediated methods?

a. Hypothesis 3a: Test of HA: Studies conducted in settings where ELLs are segregated from their English-speaking peers will report significantly lower effect sizes than studies conducted in settings where ELLs are integrated with non-ELLs.

b. Hypothesis 3b: Test of HA: Studies conducted in settings that authors describe as having adequate facilities will report significantly higher effect sizes than studies conducted in settings that authors describe as inadequate.

c. Hypothesis 3c: Test of HA: Studies conducted with ELL-certified teachers will report significantly higher effect sizes than studies in which teachers do not possess specialized certifications to work with ELLs.

d. Hypothesis 3d: Test of HA: Studies testing interventions described by the authors as at least partially culturally-relevant will report larger effect sizes than studies that do not make culturally-relevant claims.

e. Hypothesis 3e: Test of HA: Years of teaching experience will be positively correlated with effect sizes.

f. Hypothesis 3f: Test of HA: Studies reporting interventions that utilize students' native language during instruction will report larger effect sizes than studies using only students' second language (i.e., English) for instruction.

Criteria for Inclusion and Exclusion of Studies

A number of researchers argue that not enough experimental evaluations of intervention effectiveness exist in the ELL literature (e.g. Slavin and Cheung 2005, August and Shanahan 2006). Therefore, this meta-analysis cast a relatively-wide net, and subsequent analyses attempted to identify biases and sources of variance.

Types of Studies

Both 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 which 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.

Although 20 years is a common standard for study inclusion, studies that are older than 20 years were included if they met the other criteria because scarcity of research suggests that older studies may be necessary to provide sufficient power for the detection of effects and moderator analyses.

Finally, 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 in order to facilitate direct comparisons to ELLs in US schools; however, participants may have represented any language background, and instruction could have occurred in any language, as well.

Types of Participants and Interventions

Studies must have tested the effects of peer-learning involving students between the ages of 3 and 18, again in order to facilitate comparisons to US students in K-12 educational settings. For example, in studies of peer tutoring, both students for whom outcomes are measured and students who act as tutors must have been within this age range to preserve the focus on "peer" interactions. Also, participants must have included students identified as English language learners (though methods of identification and definitions of ELL varied), and results must have been exclusively, or disaggregated, for ELLs. For example, the inclusion of studies conducted internationally necessitated the inclusion of students learning English as a Foreign Language (EFL) and students in the United States learning English as a Second Language (ESL). The difference in settings (e.g. immersed in an English-dominant environment for ESL students) makes the process of language

acquisition very different, but for purposes of this synthesis, both of these types of learners were subsumed under the ELL category.

Interventions may have utilized 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). 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 large amounts of peer assistance.

Types of Outcomes and Instruments

Cooperative learning has been used to improve almost every conceivable academic achievement outcome, but it has also been widely used to improve a number of behavioral and social outcomes. Therefore, nearly any outcome was coded, though some outcomes were not assessed frequently enough to allow inferential statistical analyses. To facilitate coding and analysis, outcomes will be divided into five conceptually-distinct categories; and while variety existed within categories (e.g. math and social studies within academic outcomes), it was presumed that enough similarity existed to facilitate comparative analyses. These categories are: oral language, written language, other academic, attitudinal and social. Oral language outcomes were those that focused on speaking and listening, while written language outcomes were those that included primarily reading and writing. Other academic outcomes included content-area outcomes from subjects like science, social studies, and mathematics. Attitudinal outcomes were psychological in nature and consisted almost entirely of measures of motivation, and social outcomes were behavioral measures of things like interactions with native speakers. In some cases, measures were broad-band, complex measures that included aspects of several of these categories. For instance, the Revised Woodcock-Johnson Test of Achievement is a widely-used instrument that explicitly measures oral language, reading fluency and comprehension, and academic achievement. In some cases, specific subtests were reported and when possible, these sub-test scores were coded separately into one of the above categories. However, in other cases, only composite scores were reported, and in some cases descriptions of the measure seemed to favor one category over another. In some cases, however, the measures were simply too inclusive to reliably choose one category over another. In these cases, in order to maintain inter-rater reliability and to provide a systematic coding approach that could be replicated later, written language was chosen as the default outcome category for complex outcomes that measured more than one category.

Similarly, a number of 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, and they followed the same construct-driven division of results just discussed.

Search Strategy for Identifying Relevant Studies

Multiple databases were searched using consistent combinations of keywords, though specific format 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: Dissertations at Vanderbilt University and 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.

“Near-miss” 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 above, 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. The “near miss” studies are included in the references section, 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 experimental and statistical training methods in the ExpERT program at Vanderbilt University. After some discussion of the inclusion and exclusion criteria and practice with an example, the other coders made inclusion/exclusion decisions for a sub-sample of 30 abstracts.

Description of Methods Used in Primary Studies

As already discussed, previous syntheses suggest that high-quality experimental studies are scarce in this field. Consequently, it seems appropriate to cast a wide net, a long-standing approach to social science syntheses (e.g. Smith et al. 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 tremendous impact on the final synthesis, and so, an attempt to assess the extent to which study quality is related to reported effects was made. Thus, studies were coded to reflect the extent to which they employed 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, and the approach used to adjust for pre-test differences was also coded. For the sake of moderator analysis, “study quality” was assessed on a three-level scale determined by this information, such that: a) high-quality studies assessed pre-test equivalence AND used a covariate to control of 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.

Criteria for Determination of Independent Findings

As is often the case in meta-analysis, some studies reported data on several outcomes, and occasionally multiple measures of the same construct were provided by individual studies. For instance, a study may have measured outcomes of reading comprehension, reading fluency, and attitudes toward reading. Furthermore, both researcher-specific and state-mandated measures of reading comprehension were sometimes reported. For all such cases of multiple measures, the following general approach was used. First, every measure was coded in order to provide simple descriptive summaries of the kinds and frequencies of outcomes reported in the literature. Then, as part of the coding, outcomes were categorized into one of the five primary constructs outlined above. Finally, for situations in which multiple outcomes and/or measures were provided for any given construct in a single study (e.g. two different academic outcomes), a focal measurement was identified. In general, the most reliable instrument was coded as the focal instrument, though in cases where reliability information was not provided, the most widely-used measure was chosen. If neither of these criteria could be employed, the first measure discussed was chosen as a default. Although many meta-analyses average effects across measures, individual measures were utilized in this review because the measures varied considerably within constructs (e.g. math, reading and science within academic) and because coding of individual measures preserves the possibility of additional analyses at a later time. In any case, only one measurement for each of the five main constructs was identified as a focal instrument, allowing analyses within constructs that did not violate assumptions of independence.

Details of Study Coding Categories

A number of study and outcome characteristics were coded in order to enable analyses of the primary research questions as well as a number of potentially-relevant moderator analyses. A brief summary of the variables coded is provided here. Essentially, the variables included: study descriptors like design and quality, participant descriptors like age and language background, treatment descriptors like duration and frequency, and a variety of outcome descriptors. Key outcome descriptors included primary data like means and standard deviations as well as secondary calculations like effect sizes. While effect size

statistics are discussed in more detail elsewhere, as much relevant information as necessary for effect size calculations was identified and coded, in keeping with guidelines provided by Lipsey and Wilson (2001).

Moderating variables are those that may affect overall effect size estimates leading to different effect sizes estimates for different values of the moderator. A number of study, treatment and participant variables were analyzed as moderators in 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. A potential limitation of multiple moderator analysis is that it does not account for covariation amongst moderators, and meta-regression is an alternative analysis that allows examination of the independent contributions of each variable to variance in the effect sizes. To the extent possible, meta-regression analyses of key moderators that affect outcomes was conducted to determine the unique contribution made to the variance of outcomes by methodological and substantive moderators. At minimum, single-variable regressions of potentially influential variables were run to test their viability as moderator variables, even if multivariate regression was untenable because of small sample size. Exploratory analyses of substantively important variables also included correlational analysis and descriptive statistics.

Finally, coding reliability was assessed through 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 again 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. For variables with reliability coefficients low enough to be close to chance agreement, variable constructs were reexamined and disagreements were examined case by case to reach consensus.

The effect size statistic (ES) calculated was the Standardized Mean Difference(ES_{SM}), which is appropriate for group contrasts made across a variety of dependent measures (Lipsey and Wilson 2001). The most frequently-coded variables were continuous variables (e.g. standardized test results) with results contrasting mean treatment and control group performance on focal outcomes. The following is the formula for calculating the ES_{SM}

$$\overline{ES} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}}_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Thus, the mean effect size is calculated by dividing the difference between the mean for the treatment (X_{G1}) and the mean for the control (X_{G2}) by the pooled standard deviation (s_{pooled}). We see in the second formula that the pooled standard deviation (s_{pooled}) is equal

to the square root of the sum of the weighted variance for the treatment group (s_1^2 * [n_1-1]) and the weighted variance for the control group (s_2^2 * [n_2-1]) divided by the pooled degrees of freedom (n_1 + n_2 - 2). In these formulas, s^2 is the observed variance and n is the sample size.

The ES_{SM} is known to be upwardly biased for small samples. Thus, the Hedges G transformation is traditionally used to correct for this bias

$$G = D\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

Where Cohen's D = ES_{SM}, the biased effect size estimate weighted by a correction for small sample bias. This adjusted effect size, ES'_{SM}, has its own SE and inverse variance weight formulas, as illustrated in Lipsey and Wilson (2001). The weight term is included to compensate for reliability differences resulting from different sample sizes. That is, small sample sizes generate less precise estimates, whereas larger sample sizes generate more reliable estimates, and this weight term adjusts the impact of the estimates based on their sample size-driven reliability. The following formulas display the calculations for computing the standard error and weights for use with the standardized mean difference effect size statistic:

$$se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{\overline{ES}_{SM}}{2(n_1 + n_2)}} = \frac{1}{se^2}$$

However, the illustrated weight formula is appropriate only for fixed effects models which assume invariate effect sizes across studies. These assumptions are untenable given the broad constructs included in the proposed meta-analysis; consequently, a random effects model will be utilized in this meta-analysis, and the formulas for this model include another variance component in the denominator of the weight formula:

$$w_i = \frac{1}{se_i^2 + \hat{\nu}_\theta}$$

In addition to the sampling error represented by the term se_i^2, the random effects weight includes a term for heterogeneous effect sizes, v_θ. This additional term is a constant weight applied to every study, and can be computed as a method of moments estimate using the Q statistic, which is a measure of the heterogeneity of effect sizes within the sample. The formula for v_θ is:

In this formula, Q is the heterogeneity statistic provided in standard CMA output, k is the number of effect sizes included in the analysis, and w is the fixed-effects weight calculated as before.

As indicated, heterogeneity was assessed using the Q statistic, which describes the degree to which effect sizes vary beyond the degree of expected sampling error. I^2 is another useful measure of heterogeneity, and it indicates the amount of heterogeneity that exists between studies (Higgins et al. 2003). Both statistics were used to determine the degree of

heterogeneity in the sample of included studies, which was expected to be considerable given the relative breadth of acceptable studies.

Additionally, outliers can be particularly problematic, with extreme observations affecting both effect size estimates by distorting the means of the distributions as well as calculations of variance. Furthermore, as meta-analysis is primarily a survey methodology interested in synthesizing studies and providing descriptions of typical effects, atypical results are not overly-informative. Consequently, Tukey's guidelines were employed to identify outliers (3*IQR+75th percentile and 25th percentile-3*IQR). Results above and below these values were Winsorized to these cut-off points.

Another source of potential error involves designs that utilize cluster randomization in which intact groups are assigned en masse to conditions, and unless corrected, the standard errors upon which the inverse variance weights are based would be incorrect (Hedges 2007). This is the result of cluster effects in which students nested within classrooms tend to be more similar than students in separate classrooms. This problem can occur if randomization occurs at any level other than the level of the student, and thus, McHugh adjustments were made for studies that employed cluster randomization (Lipsey et al. 2012). The effective n, which is usually much smaller than the observed n, was computed, and these adjusted sample sizes were then used to calculate more accurate standard error estimates. However, a number of assumptions were made that merit discussion. Primarily, the rho, or inter-class correlation, will be estimated at .2 for academic and language outcomes and .15 for all other outcomes. These values are loosely based on the range of intra-class correlations obtained in Hedges (2007), which reported results from a large sample of academic outcomes from cluster-randomized evaluations. Much is more is known about academic outcomes in educational evaluation studies than for others, so a slightly lower rho is used for other outcomes. While it seems likely that observed values of rho varied across studies, the data was often not reported. Similarly, the number of students per cluster was occasionally not reported; in these cases, the total sample was divided by the number of clusters to compute a mean cluster size. Due in part to limitations in the reporting of data as well as to the relative newness of cluster effect corrections in meta-analysis, the adjusted estimates are somewhat crude and imprecise; consequently, the results of these adjustments are likely overly-conservative and may be interpreted as a lower bound of sorts.

Similarly, in several studies, pre-test data was available, but the original researchers did not use pre-test data in their post data analyses. that is, pre-test differences were left unadjusted in final analyses. In these situations, post hoc adjustments were made by this researcher to control for pre-test differences. Simply, 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.

Finally, a number of alternate computations were occasionally necessary. For instance, some studies did not provide ES estimates, and a number of formulations exist for converting other commonly reported data into ES_{SM}. These other data include means and

standard deviations, t-tests and degrees of freedom, and p values and sample sizes, and effect sizes using these alternative data were calculated as necessary.

Statistical Procedures and Conventions

General statistical analyses were computed using CMA and SPSS software; in particular, overall effect size analyses, some publication bias, and moderator analyses were computed with CMA, and diagnostic and descriptive analyses were conducted with SPSS.

Results

Chapter Four presents the data obtained from descriptive, main effects, and moderator analyses, and Chapter Five will consider the extent to which the data answers the formal research questions detailed in Chapter Three. First, descriptive information is provided for the included sample of studies. Then, descriptive statistics, main effects analyses, and moderator analyses are provided for each of the outcome categories. Because each outcome category contains independent samples of effect sizes and because outcomes are assumed to be more conceptually similar within categories than between them, Chapter Four is organized primarily by outcome type to maintain statistical and conceptual clarity.

Included Sample

Initial keyword searches returned 17, 613 results, of which 148 were unique and potentially relevant. Additionally, extant meta-analyses and syntheses (e.g., Genesee et al. 2005, Keck et al. 2006, Mackey and Goo 2007) were mined for potentially relevant studies, as were studies included in the author's prior research. Similarly, key authors were contacted in a gray literature search to identify additional studies that might be potentially relevant. From these combined sources, ultimately 37 study reports were included. Initial agreement rates among coders for inclusion/exclusion decisions were 92.4%, and differences were resolved to achieve consensus in ultimate coding. Included studies and near-miss studies are listed in Suppl. material 2. Table 2 below provides a snapshot of the included sample and a few key variables.

Table 2.

Included sample of 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
August	1987	Journal	USA	Peer Tutoring	Quasi-Experiment	Elementary

Lead Author	Year	Publication Type	Country	Construct	Design	Grade Level
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
Bustos	2004	Dissertation	USA	Cooperative	Experiment	Elementary
Calderon	1997	Technical Report	USA	Cooperative	Quasi-Experiment	Elementary
Calhoun	2007	Journal	USA	Cooperative	Quasi-Experiment	Elementary
Chen	2011	Journal	USA	Cooperative	Quasi-Experiment	High School
Cross	1995	Technical Report	USA	Collaborative	Quasi-Experiment	High School
Dockrell	2010	Journal	England	Collaborative	Quasi-Experiment	Pre-K
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
Johnson	1983	Journal	USA	Peer Tutoring	Experiment	Elementary
Jung	1999	Dissertation	South Korea	Peer Tutoring	Quasi-Experiment	Elementary
Khan	2011	Journal	Pakistan	Cooperative	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
Lopez	2010	Journal	USA	Collaborative	Quasi-Experiment	Elementary
Mack	1981	Dissertation	USA	Collaborative	Quasi-Experiment	Elementary
Martinez	1990	Dissertation	USA	Cooperative	Quasi-Experiment	Elementary
Prater	1993	Journal	USA	Cooperative	Experiment	Elementary
Sachs	2003	Journal	Hong Kong	Cooperative	Experiment	High School
Saenz	2002	Dissertation	USA	Peer Tutoring	Quasi-Experiment	Elementary
Satar	2008	Journal	Turkey	Collaborative	Experiment	High School
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
Tong	2008	Journal	USA	Collaborative	Quasi-Experiment	Elementary
Uludag	2010	Dissertation	Jordan	Collaborative	Quasi-Experiment	Middle/ High School
Vaughn	2009	Journal	USA	Peer Tutoring	Quasi-Experiment	Middle School

The 37 included studies reported relevant data on 44 independent samples (i.e., several reports described multiple experiments or included independent samples) and contained a total of 132 outcomes. As indicated in the full coding manual (in the Excel spreadsheet that accompanies this dissertation Suppl. material 1), numerous methodological, study-level,

sample-level, and outcome variables were coded for the included sample. Inter-rater reliability varied considerably across variables; mean Cohen's Kappa for categorical variables was (K=.787) with a range of (K=.318 to K=.1.0). Pearson's r was calculated for continuous variables, and mean agreement amongst raters was (r=.927) for continuous variables, though inter-rater reliability for continuous variables ranged between (r=.85 and r=1.0). Problematic variables were discussed and revised, and ultimately, all differences were resolved to consensus. Key variables are summarized in the tables below; Table 3 details several methodologically and theoretically important variables, and Table 4 summarizes key outcome data for the included sample.

As indicated in Table 3, peer-mediated learning for ELLs is currently an active field of research; in fact, more studies were conducted in the most recent decade than either of the previous decades. Moreover, the included sample is evenly composed of published (n=22) and unpublished (n=21) studies, and the sample contains nearly the same number of international studies (n=21) as studies conducted in the United States (n=22). Similarly, all three peer-mediated constructs are well-represented in the included sample, though there are fewer peer tutoring studies than cooperative or collaborative. However, some variables are less balanced; for instance, there are far more high-quality studies (as operationally defined) than medium or low-quality studies, and every study was conducted in a school setting, meaning that no lab studies are included in the sample. In many ways, it is what is missing in the included sample that is most striking. Very little information about the teachers was reported, and very few studies reported information about students' SES or length of residence. Similarly, contextual variables like the adequacy of facilities or the context of reception were typically not reported. Not only does the absence of this information limit the potential to conduct moderator analyses for these variables, it potentially limits the external validity of this meta-analysis. That is, findings are relevant only for a constrained set of variables, and the general effectiveness of peer-mediation may vary across a number of unmeasured, or unreported, variables.

Table 3.

Summary of Key Variables in Included Sample

Year (n=43)	Pre1980-1989 = 4	1990-1999 = 10	2000-2012 = 29	
Publication Type (n=43)	Dissertation = 15	Journal = 22	Technical Report = 6	
Country (n=43)	USA = 22	Other = 21		
Setting (n=43)	ESL= 23	EFL= 20		
Design (n=43)	Experimental = 8	Quasi-experimental= 35		
Quality (n=43)	High = 26	Medium = 13	Low = 4	
Dosage (Total Contacts) (n=43)	0-30 = 17	31-90 = 13	91+ = 13	
Construct (n=43)	Cooperative = 17	Collaborative = 16	Peer Tutoring = 10	

Component (n=43)	Yes =19	No =24		
Adequate Facilities (n=23)	Yes = 2	No = 3	Unknown = 18	
Segregated (n=23)	Yes = 9	No = 14		
Culturally Relevant (n=23)	Yes = 5	No =18		
Language of Instruction (n=43)	L1 only = 2	Bilingual = 14	L2 only = 14	Unknown = 13
In School (n=43)	Yes = 43	No = 0		
Teacher Certification (n=43)	ELL Certified = 12	Not ELL Certified = 2	Unknown =29	
Teacher Experience (n=43)	0-5 years= 3	6-10 years= 4	11+ years= 4	Unknown= 32
Teacher Ethnicity (n=43)	Same as Students'= 7	Different than Students' = 1	Unknown = 35	
Grade Level (n=43)	Elementary = 22	Middle = 8	High = 13	
Student Ethnicity (n=43)	Spanish = 20	Asian = 8	Other = 15	
Student SES (n=43)	Low = 21	High = 3	Mixed = 1	Unknown = 18
Student Length of Residence (n=23)	0-2 years = 1	2+ = 0	Unknown = 22	

Table 4 indicates that language outcomes were far more prevalent than academic or attitudinal outcomes, and social outcomes are completely absent from the included sample. In fact, too few studies are reported for academic outcomes to reliably conduct moderator analyses, and the samples for attitudinal and oral language are only marginally large enough. Thus, the presented moderator analyses for all three of these outcome types should be considered exploratory; however, the sample of written language outcomes is large enough to conduct moderator analyses with some degree of confidence, and tentative meta-regression results should be sufficiently powered to enable insight into which moderators are most influential.

Table 4.

Key outcome variables.

Total Outcomes= 62	**Number of Independent Outcomes by Construct**	**Number of Participants in Treatment Groups**	**Number of Participants in Control Groups**
Oral Language	14	843	787
Written Language	30	919	863
Other academic	6	220	451
Attitudinal	10	397	394
Social	0	0	0

Oral Language Outcomes

Summary of Included Studies and Main Effects

A random effects model of the un-corrected and un-Winsorized data provided a mean effect size estimate for the thirteen oral language outcomes of (.587, SE=.141, p<.001); however, after adjustments for outliers, pre-test differences, and cluster randomization, the mean effect size estimate decreased slightly and the variance decreased slightly (.578, SE=.136, p<.001), suggesting that the larger-than-average outliers and the effects of cluster randomization had very little impact on the original estimates. The adjusted distribution is illustrated by the forest plot in Fig. 2. It is notable that only one study (i.e., August 1987) has a mean below zero. Also, this distribution highlights one of the real strengths of meta-analysis; more than half of the studies have confidence intervals that cross the zero threshold, meaning that individually they are statistically indistinguishable from an effect size of zero. However, taken together, they provide enough statistical power to identify a strong, positive effect with a great deal of confidence.

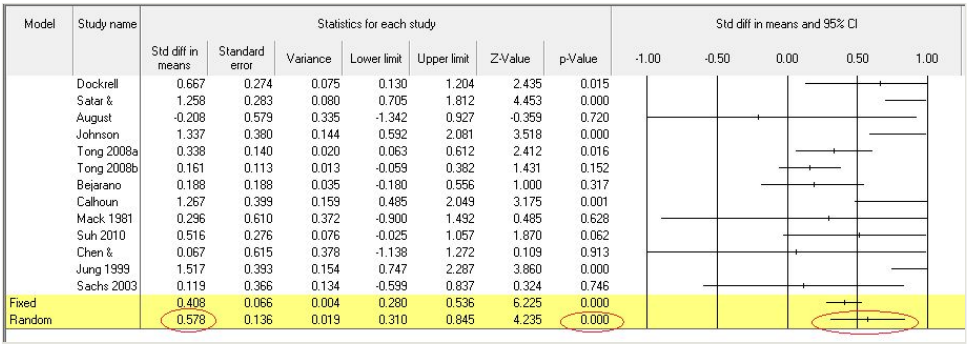


Figure 2. doi

Forest plot of oral language outcomes.

Throughout the paper, random effects models are the default, primarily because the assumptions of the fixed model are generally untenable. Empirically, homogeneity analysis of the fixed model illustrates the considerable heterogeneity that exists within the observed sample, offering some empirical justification for the use of a random effects model. The Q statistic (37.213, df=12, p<.001) indicates that the observed effect sizes vary more than would be expected by sampling error alone, and the I2 statistic (67.753) indicates that approximately 68% of the observed variance in effect sizes exists between studies. Together, this suggests that moderator analyses might provide insight into what factors influence the effectiveness of peer-mediated learning for ELLs.

Publication Bias for Oral Language Outcomes

The possibility of publication bias remains a persistent concern in meta-analysis, and the following analysis examines empirical evidence for the presence of publication bias in this sample and the extent to which it might distort the estimates. Lipsey and Wilson (1993, as

cited in Lipsey and Wilson 2001) demonstrated that published studies tend to report larger mean effect sizes than unpublished studies. While it is impossible to determine if this is the result of bias on behalf of journal editors or researchers, it is potentially problematic if the under-representation of unpublished studies induces significant bias. And while it is likely impossible that any literature review could be thorough enough to locate every study ever written on a given topic, the conceptual possibility that other studies *could have been written* is sufficient to suggest that the true population parameter could differ systematically from the retrieved sample. Similarly, given these vagaries, practically and conceptually, it is not possible to empirically demonstrate publication bias with complete certainty; rather, one can demonstrate the possibility of publication bias and estimate the potential effects of such bias on main effects analysis.One way to check for possible publication bias is to compare the means of published and unpublished studies in the sample; because unpublished studies represent only a fraction of the total empirical literature on a topic, the simple difference between the mean effect size estimates of the published and unpublished samples provides a sort of upper bound for publication bias.

A recoding of the type of publication variable into a dummy-coded variable with 1=published and 0=unpublished, indicated that 84.6% of the included sample had been published, while the other 15.4% were dissertations. The mean effect size for published studies (.377, SE=.067) is surprisingly much smaller than the mean effect size for unpublished studies (1.159, SE=.330). The difference between the mean effect sizes of -.782 provides a crude estimate of the upper bounds of potential publication bias. Of course, this simple difference does not adequately account for small sample bias nor does it employ inverse variance weights; consequently, appropriately meta-analytic tests of publication bias must also be utilized.

A look at 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. If there is no such bias, we expect small studies with negative and null results to be as frequent as small studies with positive results. The following funnel plot in Fig. 3 includes black circles for studies that have been imputed to achieve a symmetric distribution, the “trim and fill” technique, and we notice that both imputed studies fall in quadrant one, which is inconsistent with the possibility of publication bias. We also notice that when these studies are imputed, the mean effect size estimate remains relatively unchanged.

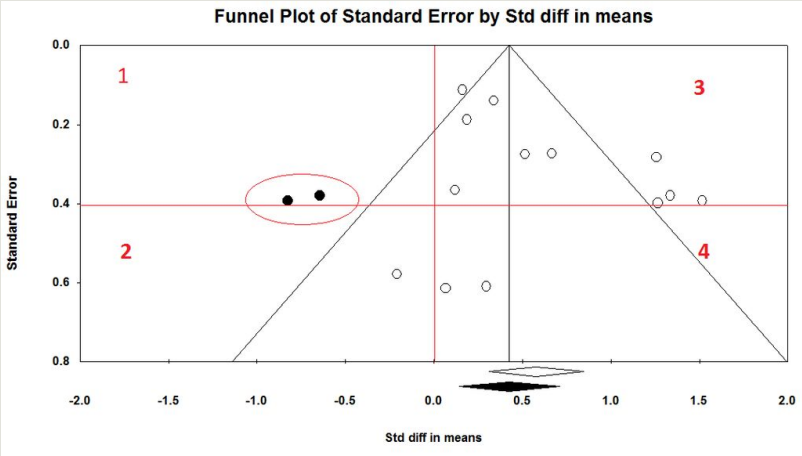


Figure 3. doi

Funnel plot of oral language outcomes with missing studies imputed

A computational alternative to visual inspection of the distribution is Egger's regression intercept, as discussed in Sterne and Egger (2006); (Fig. 4):

Egger *et al.* (1997) introduced a linear regression approach in which the standard normal deviate z_i (defined as $z_i = 0_i/s_i$) is regressed against its precision $prec_i$ (defined as $prec_i = 1/s_i$):

$$E[z_i] = \beta_0 + \beta_1 prec_i.$$

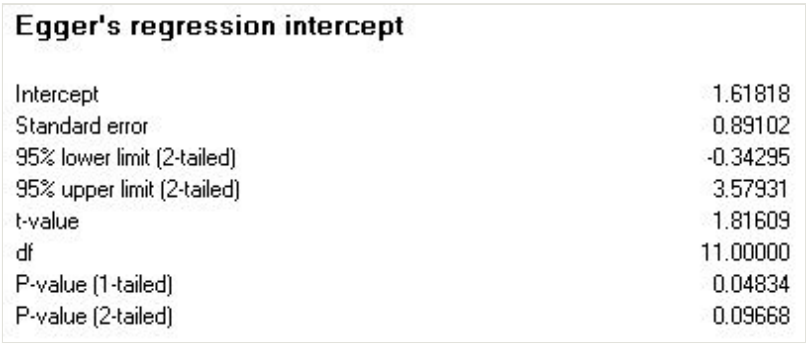
Figure 4. doi

Egger's regression intercept.

Because we assume that publication bias will be positive, that is, in the direction of significantly positive effects and because it provides a more conservative estimate of significance, the p value of the single-tailed test at α=.05 is typically reported. The null hypothesis tests whether the ratio of the ES/se is > 0. While some debate exists about whether the single-tailed or two-tailed test is more appropriate, we see in Fig. 5, that in this case the two estimates provide conflicting evidence of publication bias in the oral language outcome distribution. The intercept is significantly greater than zero for only the one-tailed test (1.618, t-value=1.816, p=.048) but not the two-tailed test (p=.097), thus providing limited evidence that smaller sample sizes are associated with larger effect size estimates.

In conclusion, these varied analyses provide very little evidence for the possibility that publication bias is likely for the distribution of studies reporting oral language outcomes. Furthermore, the potential bias induced is small enough that if a sufficient number of small sample studies with null or negative results were included to make the distribution more symmetrical, the mean effect size estimate would hardly change. As indicated, very few studies in the sample have null or negative effect size estimates; as such, it remains

distinctly possible that the literature search failed to uncover those studies that for one reason or another simply were not published because they failed to yield significantly positive results.



Egger's regression intercept

Intercept	1.61818
Standard error	0.89102
95% lower limit (2-tailed)	-0.34295
95% upper limit (2-tailed)	3.57931
t-value	1.81609
df	11.00000
P-value (1-tailed)	0.04834
P-value (2-tailed)	0.09668

Figure 5. doi

Egger's regression for oral language outcomes.

Moderator Analyses for Oral Language Outcomes

The distribution of oral language effect sizes was heterogeneous, as indicated by the Q and I^2 statistics; consequently, we might expect post hoc examination of moderator variables to uncover some statistically-significant moderator variables. However, the sample is modest (n=13) and underpowered for meta-regression analysis of the partial contributions for multiple independent variables. Given these limitations, analysis of moderators is primarily motivated by a priori questions of interest, and findings are qualified by the recognition that small differences may be difficult to detect with the small sample employed and confounding and lurking variables may temper any observed differences between sub-groups. Occasionally, when a categorical variable had too few studies on one or more categories, the category was recoded, often into a binary variable, to enable a more reliable comparison. Table 5 summarizes the results for measured variables reported in all thirteen studies, and the presence of significant bivariate correlations (i.e., chi square test) with other measured variables is indicated in the last column.

As indicated in the Q-between column, only two moderators were statistically significant at the p=.05 level: post hoc researcher adjusted and segregated. In cases where post-test effects sizes were unadjusted for pre-test differences by authors in the original study reports, the researcher of this meta-analysis adjusted post-test effect sizes post hoc. In these cases, post hoc adjustments resulted in much smaller effect sizes on average (G=.174) than unadjusted (G=.675). This finding indicates that methodological rigor and care in synthesizing previous research can exert a large influence on reported results. The other significant moderator of the effectiveness of peer-mediated learning for improving oral language outcomes was whether or not the intervention occurred in settings where ELLs were segregated from their non-ELL peers. ELLs in segregated settings performed much lower (G=.230) than they did in settings that were not segregated or in settings for which

segregation was unreported (G=.636). Some care should be taken when interpreting this result, in particular. First, the confluence of segregated settings with ambiguous settings (i.e., researchers did not report if segregated) presents some conceptual challenges in interpreting the results because some of the ambiguous settings may very well have been segregated in practice. Secondly, the number of studies that reported that they were segregated was relatively small (n=2), and so the estimate is not as precise as it could have been.

Table 5.

Summary of moderator analyses for oral language outcomes.

Moderator (Sub-group)	**Number in sub-group**	**Effect Size Point-estimate**	**Standard Error of estimate**	**p-value of estimate**	**Q-within of Sub-group**	**I^2 of Sub-group**	**Q-between in Random Effects Model**	**Observed Inter-correlation**
Published							**.601 (p=.438)**	**Yes**
Yes	11	.377	.067	.000	29.005 (p=.001)	65.523		
No	2	1.159	.330	.099	3.683 (p=.09)	64.681		
Study Quality							**4.089 (p=.129)**	**Yes**
High	7	.587	.164	.000	18.544 (p=.005)	67.644		
Medium	4	.761	.364	.036	8.266 (p=.041)	63.077		
Low	2	.174	.167	..299	.028 (p=.866)	.000		
Instrument Type							**2.513 (p=.285)**	**Yes**
Researcher-created	5	.478	.238	.045	10.408 (p=.034)	61.570		
Standard-Narrow	6	.743	.204	.000	25.583 (p=.000)	80.456		
Standard-Broad	2	.031	.420	.941	.0359 (p=.549)	.000		
Post Hoc Researcher Adjusted							**4.634 (p=.031)**	**Yes**
Yes	2	.174	.167	.299	.028 (p=.866)	.000		
No	11	.675	.162	.000	34.863 (p=.000)	71.136		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Construct							**2.503 (p=.286)**	**Yes**
Cooperative	2	.105	.315	.738	10.283 (p=.068)	51.378		
Collaborative	6	.506	.157	.001	.005 (p=.942)	.000		
Peer Tutoring	5	.837	.348	.016	18.721 (p=.001)	78.634		
Component							**1.035 (p=.309)**	**Yes**
Yes	4	.388	.172	.024	7.406 (p=.06)	59.494		
No	9	.651	.193	.001	24.013 (p=.002)	66.684		
Setting							**.380 (p=.538)**	**Yes**
EFL	5	.691	.269	.010	17.426 (p=.002)	77.045		
ESL	8	.498	.161	.002	17.332 (p=.015)	59.612		
Segregated							**5.412 (p=.020)**	**Yes**
Yes	2	.230	.088	.009	.966 (p=.326)	.000		
Other (Not and Unknown)	11	.686	.175	.000	26.944 (p=.003)	62.866		
Language of Instruction							**.681 (p=.711)**	**Yes**
L1 (L1-only and bilingual)	7	.649	.186	.000	24.282 (p=.000)	75.291		
L2 Only	4	.427	.215	.047	2.36 (p=.501)	.000		
Unknown	2	.702	.535	.189	9.946 (p=.002)	89.946		
Culturally Relevant							**.739 (p=.691)**	**Yes**
Yes	3	.413	.196	.035	7.405 (p=.025)	72.933		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
No	5	.572	.264	.03	6.701 (p=.153)	40.309		
Not U.S.A.	5	.691	.269	.01	17.426 (p=002)	77.045		
Grade Level							**.240 (p=.624)**	**Yes**
Elementary	9	.628	.164	.000	25.846 (p=.001)	69.047		
Other	4	.454	.314	.148	11.320 (p=.010)	73.499		
SES							**.194 (p=.908)**	Yes
Low	5	.518	.193	.007	6.821 (p=.146)	41.36		
High	2	.788	.582	.176	3.099 (p=.078)	67.731		
Unknown	6	.550	.202	.007	19.731 (p=.001)	74.659		
Student Hispanic							**.541 (p=.462)**	
Hispanic	7	.472	.181	.009	15.801 (p=.015)	62.027		
Other (Asian, Arabic, Bangladeshi, Israeli)	6	.68	.217	.002	17.535 (p=.004)	71.486		
Student Asian							**.139 (p=.71)**	
Asian	3	.696	.376	.064	7.206 (p=.027)	72.244		
Other	10	.545	.15	.000	28.272 (p=.001)	68.166		

For all other variables, differences in mean effect sizes were evident across variables, but none proved to be significant moderators. Because the sample size for oral language outcomes is relatively small, this general lack of statistically significant moderators likely represents a lack of statistical power to detect meaningful differences. Thus, some of these moderators might prove significant if additional studies were included, and future meta-analyses may benefit from larger sample sizes as the field continues to produce experimental and quasi-experimental evaluations of peer-mediated learning.

Written Language Outcomes

Summary of Included Studies and Main Effects

A random effects model of the un-corrected and un-Winsorized data provided a mean effect size estimate for the twenty eight written language 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 had some noticeable impact on the original estimates. The adjusted distribution of written language outcomes is illustrated by the forest plot in Fig. 6. Unlike oral language outcomes already discussed, the distribution of written language outcomes includes eight studies with means equal to or less than zero. This really highlights the importance of publishing studies with null or negative findings, as they contribute to more accurate and meaningful syntheses.

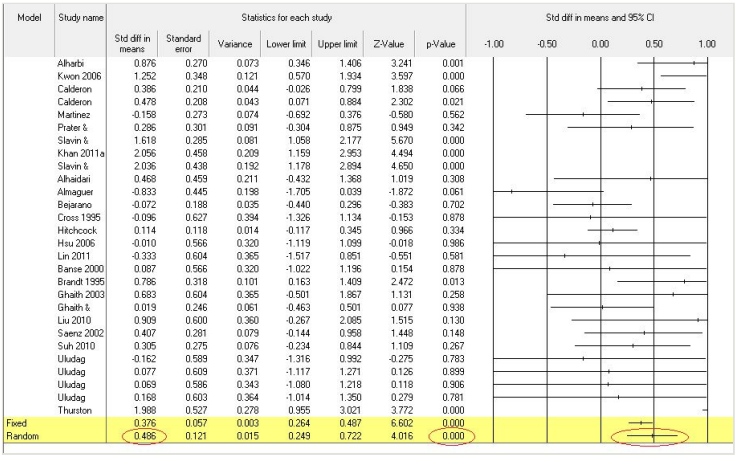


Figure 6. doi

Forest plot of written language outcomes.

The distribution of effect sizes for written language outcomes was even more heterogeneous than the distribution of oral language outcomes. The Q statistic (97.135, df=27, p<.001) indicates that the observed effect sizes vary more than would be expected by sampling error alone, and the I2 statistic (72.204) indicates that approximately 72% of the observed variance in effect sizes exists between studies. Together, this suggests that moderator analyses might provide insight into what factors influence the effectiveness of peer-mediated learning for ELLs for written language outcomes.

Publication Bias for Written Language Outcomes

A recoding of the type of publication variable into a dummy-coded variable with 1=published and 0=unpublished, indicated that 64.3% of the included sample were unpublished (i.e., technical reports and dissertations), while the other 36.7% were

dissertations. 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). The difference between the mean effect sizes of -.082 provides a crude estimate of the upper bounds of potential publication bias.

The funnel plot in Fig. 7 includes black circles for studies that have been imputed to achieve a symmetric distribution, the “trim and fill” technique, and we notice that there are no studies imputed to achieve a symmetric distribution, which is inconsistent with the possibility of publication bias. Similar, the black diamond indicates that the anticipated mean did not change at all under publication bias conditions.

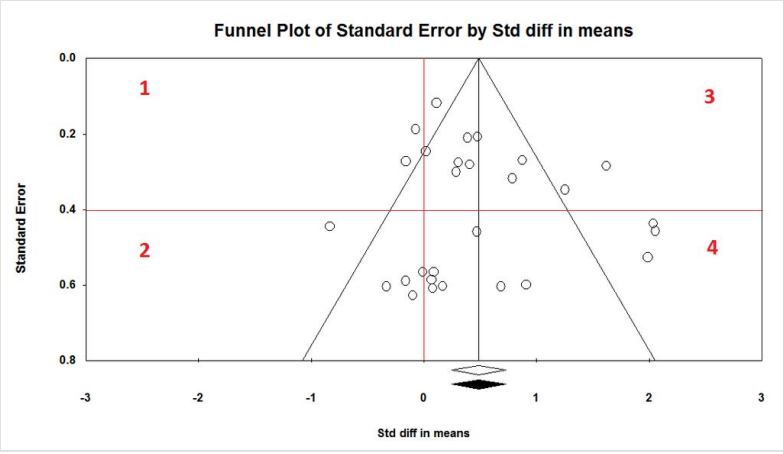


Figure 7. doi

Funnel plot of written language outcomes with missing studies imputed.

We see in Fig. 8, that the Egger’s regression test provides confirmatory evidence of the improbability of publication bias in the written language outcome distribution. The intercept is not significantly greater than zero for the one-tailed test (1.02, t-value=1.338, p=.096) or the two-tailed test (p=.193).

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 8. doi

Egger’s Regression for Written Language Outcomes.

In conclusion, these analyses provide no evidence for the possibility that publication is likely for the distribution of studies reporting written language outcomes. Additionally, several studies in the sample have 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 because they failed to yield significantly positive results, and as indicated by the funnel plot and the difference in means between published and unpublished studies, the possible impact of studies lurking in the "the file drawer" on the mean effect size estimates appears relatively minor in this case.

Moderator Analyses for Written Language Outcomes

The distribution of oral language effect sizes was heterogeneous, as indicated by the Q and I2 statistics; consequently, we might expect post hoc examination of moderator variables to uncover some statistically-significant moderator variables. The sample is large enough (n=28) and sufficiently powered for meta-regression analysis of the partial contributions for at least a few, (e.g., 2-3) independent variables. As before, analysis of moderators is primarily motivated by a priori questions of interest, and findings remain qualified by the recognition that small differences may be difficult to detect with the size of the sample employed and confounding and lurking variables may temper any observed differences between sub-groups. Table 6 summarizes the results for measured variables reported in the 28 studies included for this outcome type, and the presence of significant bivariate correlations, analyzed as chi square statistics, with other measured variables is indicated in the last column.

Like the distribution of oral language outcomes, the distribution of written language outcomes demonstrated few significant moderators, indicating that peer-mediated learning is effective across a number of methodological, setting, and participant variables. However, three moderators were statistically significant at the p=.05 level: study quality, post hoc researcher adjusted, and grade level. As with oral language outcomes, post hoc adjustments of written language 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. For this distribution, study quality was also a significant moderator; as study quality increased, so did the magnitude of the mean effect size, a finding that is somewhat counterintuitive. One might actually 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 low quality studies tended to report lower effect sizes (e.g., Keck et al. 2006). Finally, the other significant moderator of the effectiveness of peer-mediated learning for improving written language outcomes was grade level. Notably, middle school students showed much smaller gains (G=-.007) than high school (G=.7) or elementary (.539). It is worth noting that there were far more middle and high school studies in the written language distribution, so the categories were not collapsed as with oral language outcomes. Consequently, comparisons between the two are somewhat complicated by the differences in coding.

Table 6.

Summary of Moderator Analyses for Written Language Outcomes

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Published							**.086 (p=.770)**	**Yes**
Yes	10	.442	.240	.065	38.89 (p=.000)	76.858		
No	18	.524	.142	.000	55.851 (p=.000)	60.562		
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)	.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)	.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		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Component							**1.07 (p=.301)**	**Yes**
Yes	12	.633	.184	.001	30.714 (p=.001)	64.186		
No	16	.385	.154	.012	55.422 (p=.000)	72.935		
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.A.	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		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
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		

Other Academic Outcomes

Summary of Included Studies and Main Effects

A random effects model of the un-corrected and un-Winsorized data provided a mean effect size estimate for the twenty eight written language outcomes of (.234, SE=.079, p=.003); however, after adjustments for outliers, pre-test differences, and cluster randomization, the mean effect size estimate and the variance increased slightly (.250, SE=.13, p=.054), suggesting that outliers and cluster randomization had more impact on the standard error estimate than the mean effect size estimate. Heterogeneity for the observed sample of other academic outcomes was statistically indistinguishable from zero (Q=1.882, p=.757, I2=0.00). thus, not only were there too few studies to reliably conduct moderator analyses for this distribution, empirical evidence indicates that there is

insufficient heterogeneity for moderators to explain the variance in effect sizes. Fig. 9 illustrates the distribution of effect sizes for Other Academic Outcomes.

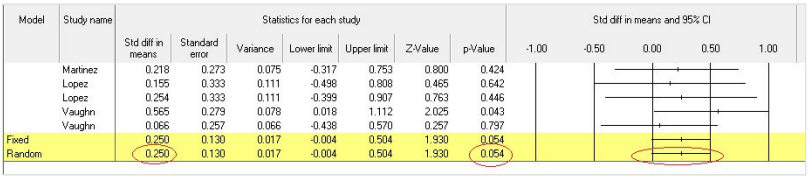


Figure 9. doi

Forest Plot of Other Academic Outcomes

Publication Bias for Other Academic Outcomes

A recoding of the type of publication variable into a dummy-coded variable with 1=published and 0=unpublished, indicated that 80% of the included sample were published in journals; the other study was a dissertation. The difference in the mean effect size for published studies (G=.260, p=.078) and the mean of unpublished studies (G=.218, p=.424) is .042 and provides a conceptual limit of the effect of publication bias on the mean effect size estimate. A funnel plot of effect sizes plotted against the standard errors in Fig. 10 shows no studies imputed. While this would suggest that publication bias is unlikely, it should be interpreted with caution given the small number of studies used for the analysis. Similarly, it should be noted that there are no studies in either quadrant one or two, suggesting that the absence of null or negative outcomes indicates that there might very well be such studies lurking in the unrecovered gray literature.

Egger's regression test provides confirmatory evidence that publication bias is not a significant threat to the validity of the mean effect size estimate. As demonstrated in Fig. 11, the intercept is not significant for either the one tailed (.352, SE=3.367, p=.462) or the two tailed test (p=.923). Again, the small sample size suggests that caution should be used when interpreting these results; nonetheless, consistently across the difference in means, funnel plot, and the Egger's regression test, empirical evidence suggests that publication bias is unlikely for the distribution of other academic outcomes.

In conclusion, the small sample of other academic outcomes shows a modest effect size of one quarter of a standard deviation that appears uninfluenced by publication bias. The small sample limits the viability of moderator analyses, and the lack of heterogeneity further discourages even exploratory analysis of the influence of moderators. The lack of included studies reporting outcomes for content areas like math, science or social studies is similar to the What Works Clearinghouse, which reports far more language outcomes than math outcomes. Similarly, a number of near-miss studies reported other academic outcomes but were excluded because they failed to meet methodological or other inclusion criteria. In general, it appears that this an emergent field of study, and future meta-analyses may prove useful as the field develops.

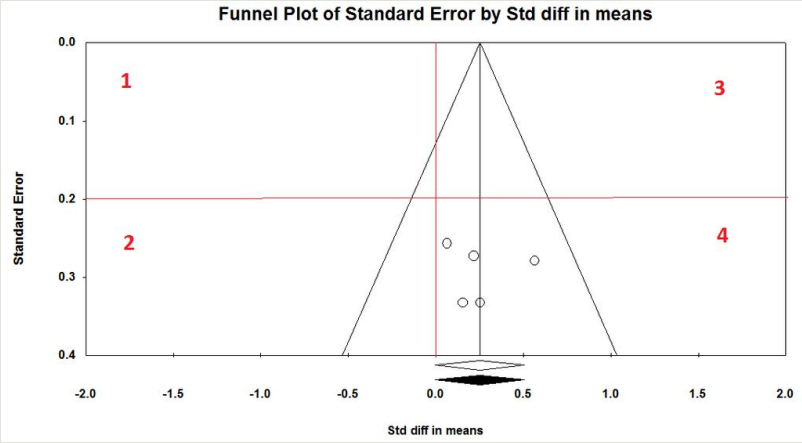


Figure 10. doi

Funnel Plot of Other Academic Outcomes

Egger's regression intercept

Intercept	0.35154
Standard error	3.36735
95% lower limit (2-tailed)	-10.36488
95% upper limit (2-tailed)	11.06796
t-value	0.10440
df	3.00000
P-value (1-tailed)	0.46172
P-value (2-tailed)	0.92344

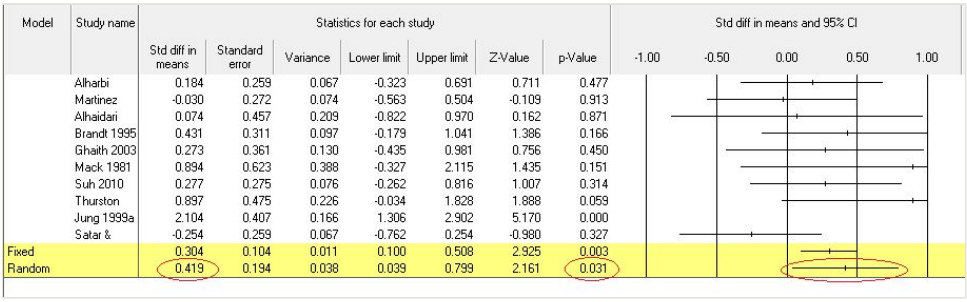
Figure 11. doi

Egger's Regression for Other Academic Outcomes

Attitudinal Outcomes

Summary of Included Studies and Main Effects

A random effects model of the un-corrected and un-Winsorized data generated a mean effect size estimate for the ten attitudinal outcomes of (.309, SE=.123, p=.012); however, after adjustments for outliers, pre-test differences, and cluster randomization, the mean effect size estimate and the variance increased noticeably (.419, SE=.194, p=.031), suggesting that outliers and cluster randomization had a moderate impact on the original estimates. Heterogeneity analysis indicate that the sample of effect sizes varies more than would be expected from sampling error alone, with about 60% of the variance occurring between studies (Q=28.806, p=.001, I^2=68.756); thus, moderator analyses might be able to explain some of this variance. The forest plot of Attitudinal outcomes is depicted in Fig. 12 below.



Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Alharbi	0.184	0.259	0.067	-0.323	0.691	0.711	0.477
	Martinez	-0.030	0.272	0.074	-0.563	0.504	-0.109	0.913
	Alhaidari	0.074	0.457	0.209	-0.822	0.970	0.162	0.871
	Brandt 1995	0.431	0.311	0.097	-0.179	1.041	1.386	0.166
	Ghaith 2003	0.273	0.361	0.130	-0.435	0.981	0.756	0.450
	Mack 1981	0.894	0.623	0.388	-0.327	2.115	1.435	0.151
	Suh 2010	0.277	0.275	0.076	-0.262	0.816	1.007	0.314
	Thurston	0.897	0.475	0.226	-0.034	1.828	1.888	0.059
	Jung 1999a	2.104	0.407	0.166	1.306	2.902	5.170	0.000
	Satar &	-0.254	0.259	0.067	-0.762	0.254	-0.980	0.327
Fixed		0.304	0.104	0.011	0.100	0.508	2.925	0.003
Random		0.419	0.194	0.038	0.039	0.799	2.161	0.031

Figure 12. doi

Forest Plot of Attitudinal Outcomes

Publication Bias for Attitudinal Outcomes

A recoding of the type of publication variable into a dummy-coded variable with 1=published and 0=unpublished, indicated that 40%of the included sample were published, and the other 60% were dissertations. The mean effect size for published studies (.201, se=.216) is considerably smaller than the mean effect size for unpublished studies (.565, se=.305). The difference between the mean effect sizes of -.364 provides a crude estimate of the upper bounds of potential publication bias.

Visual inspection of the funnel plot in Fig. 13 includes black circles for studies that have been imputed to achieve a symmetric distribution, and we notice that again there are no studies imputed to achieve a symmetric distribution, which is inconsistent with the possibility of publication bias. Thus, the black diamond indicates that the anticipated mean does not change at all. Moreover, we see that there are some, mostly larger, studies reporting null and negative effect sizes; this mitigates the possibility that such studies are languishing in file drawers somewhere. However, the included sample is small, and the results should therefore be treated with some caution.

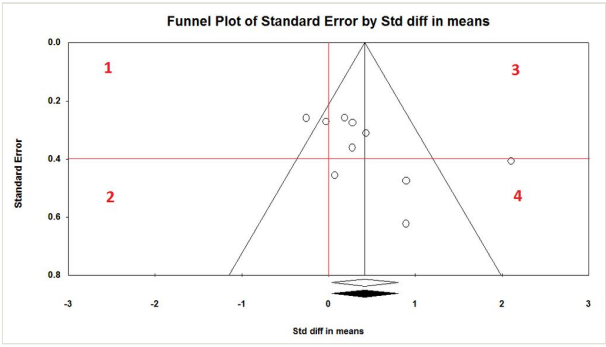
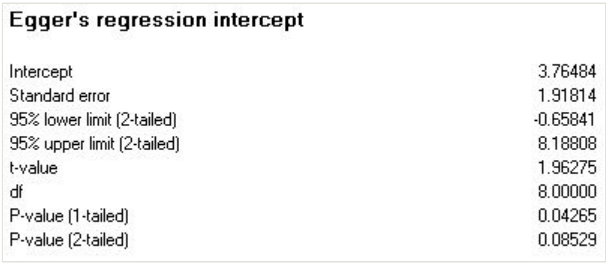


Figure 13. doi

Funnel Plot of Attitudinal Outcomes

Egger's regression test offers some evidence of the probability of publication bias for the included sample of attitudinal outcomes and provides confirmatory analysis to support the fairly large difference in means between published and unpublished studies already presented. As illustrated below in Fig. 14, the intercept is significant at α=.05 for the one tailed test (3.765, SE=1.918, p=043) and at α=.1 for the two wailed test (p=.085). Thus, there is conflicting but overall support for the probability that the main effect sizes estimates for attitudinal outcomes are influenced by publication bias.



Egger's regression intercept

Intercept	3.76484
Standard error	1.91814
95% lower limit (2-tailed)	-0.65841
95% upper limit (2-tailed)	8.18808
t-value	1.96275
df	8.00000
P-value (1-tailed)	0.04265
P-value (2-tailed)	0.08529

Figure 14. doi

Egger's regression for attitudinal outcomes.

Moderator Analyses for Attitudinal Outcomes

The distribution of attitudinal effect sizes was heterogeneous, as indicated by the Q and I2 statistics; consequently, we might expect post hoc examination of moderator variables to uncover some statistically-significant moderator variables. However, the sample is fairly small (n=10) and underpowered for meta-regression analysis of the partial contributions for multiple independent variables. Given these limitations, analysis of moderators is primarily motivated by a priori questions of interest, and findings are qualified by the recognition that small differences may be difficult to detect with the small sample employed and confounding and lurking variables may temper any observed differences between sub-groups. Table 7 summarizes the results for measured variables reported in the ten studies, and the presence of significant bivariate correlations with other measured variables is indicated in the last column.

Table 7.

Summary of moderator analyses for attitudinal outcomes.

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Published							**.947 (p=.330)**	**Yes**

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Yes	4	.201	.216	.064	5.232 (p=.156)	42.666		
No	6	.565	.305	. 352	21.834 (p=.001)	77.1		
Study Quality							**5.422 (p=.020)**	**Yes**
High	7	.650	.254	.011	19.624 (p=.003)	69.426		
Medium	3	-.058	.167	.728	1.424 (p=.491)	.000		
Low	0							
Instrument Type							**2.382 (p=.123)**	**Yes**
Researcher-created	5	.711	.36	.048	17.538 (p=.002)	77.192		
Standardized (Broad and Narrow)	5	.108	.151	.475	4.954 (p=.292)	19.257		
Post Hoc Researcher Adjusted							**5.383 (p=.020)**	**Yes**
Yes	1	-.254	.259	.327	.000 (p=.1.0)	.000		
No	9	.509	.202	.012	23.275 (p=.003)	65.628		
Construct							**4.845 (p=.089)**	**Yes**
Cooperative	5	.181	.14	.196	1.366 (p=.85)	.000		
Collaborative	3	.141	.275	.608	3.879 (p=.144)	48.442		
Peer Tutoring	2	1.525	.603	.011	3.723 (p=.054)	73.142		
Component							**.134 (p=.715)**	**Yes**
Yes	2	.523	.278	.06	.442 (p=.506)	.000		
No	8	.391	.23	.089	27.643 (p=.000)	74.677		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Setting							**.336 (p=.562)**	**Yes**
EFL	7	.466	.267	.08	26.195 (p=.000)	77.095		
ESL	3	.264	.225	.239	2.461 (p=.292)	18.745		
Segregated							**.918 (p=.338)**	**Yes**
Yes	2	.176	.229	.442	1.243 (p=.265)	19.543		
Other (Not and Unknown)	8	.5	.249	.045	26.984 (p=.000)	74.059		
Language of Instruction							**.973 (p=.615)**	**Yes**
L1 (L1-only and bilingual)	4	.651	.4	.104	19.997 (p=.000)	84.998		
L2 Only	3	.316	.258	.22	1.155 (p=.561)	.000		
Unknown	3	.169	.281	.547	4.78 (p=.092)	58.157		
Culturally Relevant							**.336 (p=.562)**	**Yes**
Yes	0							
No	3	.264	.225	.239	2.461 (p=.292)	18.745		
Not U.S.A.	7	.466	.267	.08	26.195 (p=.000)	77.095		
Grade Level							**2.237 (p=.135)**	**Yes**
Elementary	6	.667	.333	.045	21.943 (p=.001)	77.213		
Middle	0							
High	4	.119	.153	.434	3.322 (p=.345)	9.073		
SES							**.919 (p=.338)**	Yes
Low	3	.168	.205	.412	1.97 (p=.373)	.000		

Moderator (Sub-group)	Number in sub-group	Effect Size Point-estimate	Standard Error of estimate	p-value of estimate	Q-within of Sub-group	I^2 of Sub-group	Q-between in Random Effects Model	Observed Inter-correlation
Other (Includes High and Unknown)	7	.487	.261	.062	45.141 (p=.000)	77.138		
Student Hispanic							**.004 (p=.95)**	
Hispanic	4	.387	.221	.081	4.096 (p=.251)	26.76		
Other (Arabic, Asian, and Turkish)	6	.41	.292	.16	24.666 (p=.000)	79.729		
Student Asian							**1.166 (p=.280)**	
Asian	2	1.166	.913	.202	13.835 (p=.000)	92.772		
Other	8	.171	.125	.170	7.735 (p=.357)	9.497		

As with the other outcomes already discussed, most of the moderators proved insignificant predictors of variability in the effectiveness of peer-mediated learning at promoting attitudinal outcomes for ELLs; most likely, the low power prevented the detection of other meaningful effects. Nonetheless, a few variables proved to be significant (or nearly significant) moderators of attitudinal outcomes: post hoc researcher-adjusted, study quality, and the type of peer-mediated learning. The only variable to consistently prove significant as a moderator across outcome types was post hoc researcher adjustment for effect sizes that were unadjusted by the original researchers, and as before, post hoc adjustment resulted in much smaller average effect sizes (G=-.254) than unadjusted effect sizes (G=.509). Another methodological variable proved a significant moderator of attitudinal outcomes; in this case, study quality proved significant, and as with written outcomes, higher quality studies were associated with higher effect sizes. Finally, the type of peer-mediated learning (i.e., Construct) approached statistical significance, with peer tutoring studies (G=1.525) reporting much larger effect sizes than either cooperative (G=.181) or collaborative (G=.141). However, only two studies in this distribution of outcomes reported using peer-mediated learning, and consequently, caution should be used when interpreting this result. Nonetheless, given the reliability of the estimate (p=.011), it seems likely that an effect size of this magnitude is fairly meaningful despite the small sample size upon which the estimate is based.

Discussion

While Chapter 4 was organized by outcome type, the remainder of the paper is organized by the research questions presented in Chapter 3. As such, Chapter 5 is intended to synthesize findings across outcome types, and this requires a fairly organic combination of quantitative, formal hypothesis testing analysis and qualitative, pattern-seeking analysis. After addressing each of the research questions, a final section presents important limitations of this study and provides some recommendations for future research.

Research Question 1: Is peer-mediated instruction effective at promoting language, academic, or attitudinal learning for English language learners in K-12 settings?

Research Question 1 is the core question of the meta-analysis, and everything else is secondary or exploratory in comparison. Essentially, this question asks if peer-mediated learning works for ELLs, which is the most basic of effectiveness questions. Taken together, the results of the main effects analyses for all four of the available outcome types support the assertion that peer-mediated learning is very effective at promoting a number of learning outcomes for ELLs.

Specifically, the results for oral language outcomes (.578, SE=.136, $p<.001$) and written language outcomes (.486, SE=.121, $p<.001$) confirm Hypothesis 1a, which asserted that language outcomes would be significantly larger for interventions utilizing peer-mediated learning than control conditions. Both estimates are highly reliable at $\alpha=.001$, and both estimates appear unaffected by publication bias. Thus, data indicate that the alternative hypothesis of a significant difference favoring peer-mediated learning over teacher-centered or individualistic learning for ELLs cannot be rejected. Moreover, these effect sizes are of large enough magnitude to be practically significant. Compared to previous meta-analyses of cooperative learning which found effect sizes in the range of .13-1.04 (Johnson et al. 2000), the effect sizes for oral language (.578) and written language (.486) appear to be in the upper half of the distribution of effect sizes reported in Johns, et al. When compared to the effect size reported in meta-analyses of interaction for second language learners (Keck et al. 2006, Mackey and Goo 2007), the effect sizes for oral and written language found in this meta-analysis are of essentially the same magnitude as the difference between cooperative and individualistic effect sizes reported in the earlier meta-analyses. Thus, these results are largely confirmatory of the previous research on effectiveness of cooperative learning.

Similarly, the main effects analyses for other academic outcomes supports the assertion in Hypothesis 1b that peer-mediated learning would produce larger academic gains than control conditions. The mean effect size for other academic outcomes (.250, SE=.13, $p=.054$) is just significant at $\alpha=.05$, though the estimate is based on a modest sample that appeared somewhat influenced by outliers and methodological concerns. After post hoc adjustments were made, the reliability of the estimate dropped from $p=.003$ to $p=.054$, suggesting that some caution should be given to strong claims about the reliability of the estimate. Moreover, the correction of bias induced by cluster randomization reduced

heterogeneity in the sample to zero, indicating that moderator analyses were unsuitable for this distribution. Nonetheless, publication bias seems unlikely for this distribution of outcomes. The magnitude of the mean effect size of .250 appears a little smaller than the effect sizes of cooperative learning on academic outcomes reported by Slavin (1996).

Finally, the main effects analysis of attitudinal outcomes indicates that peer-mediated learning is effective at promoting motivation and similar psychologically-oriented outcomes for ELLs. The mean effect size estimate (.419, SE=.194, p=.031) is large and statistically significant at α=.05. However, it appears likely that the estimate is affected by publication bias, thus the magnitude of the estimate may be larger than it would be if all studies conducted had been published. As it stands, the current mean effect size estimate is comparable to the magnitude of previous syntheses of cooperative learning, in general (Johnson et al. 2000), as well as syntheses of interaction for second language learners (Keck et al. 2006, Mackey and Goo 2007).

In conclusion, analysis of all four outcome types indicates that the answer to research question 1 is yes, peer-mediated learning is effective at promoting a number of learning outcomes for ELLs. In fact, the estimates tended to be quite large in comparison to other instructional approaches, suggesting that peer-mediated learning is especially effective for ELLs. That effects for language outcomes are larger than effects for academic outcomes is consistent with previous syntheses supporting the linguistic rationale for peer-mediated learning. On the other hand, a sociocultural theory of learning would explain the difference by arguing that academic learning is largely mediated by language, and thus, ELLs must learn the language of the content areas before they can master the academic content. However, it could simply be that the small sample of academic outcomes simply needs to include more studies to accurately capture the effectiveness of peer-mediated learning at promoting academic learning. Unfortunately, the design of this study is insufficient to definitively discern the correct answer, and these explanations remain largely speculative.

Nonetheless, the results of the first research question answer the call of the National Reading Panel on Minority-language Youth and Children to determine if the various aspects of effective instruction highlighted by qualitative research are individually effective "…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" (August and Shanahan 2006, p.520).

Research Question 2: What variables in instructional design, content area, setting, learners, or research design moderate the effectiveness of peer-mediated learning for English language learners?

The second research question is intended to provide a more nuanced understanding for the answer to research question 1; essentially, the first question answers "What works?", and the second question attempts to answer "For whom, and under what conditions?." The following section details the answers to a large number of specific hypotheses of the influence of particular moderators and concludes with a summarizing synthesis of the effects of moderators across outcome types.

Given ambivalence in the previous literature regarding the effectiveness of specific cooperative, collaborative, and peer-mediated approaches, Hypothesis 2a suggested that there would be no significant difference among the three peer-mediated constructs, and the results of moderator analyses across the three outcome types generally support this hypothesis. For oral and written language outcomes, Construct was insignificant as a predictor, and Construct only approached significance as a predictor for attitudinal outcomes. Notably, the ES estimate for peer-mediated learning was very large (ES=1.525) for the attitudinal distribution, and it was based on only two studies. Thus, the fact that the moderator appeared nearly significant for this outcome distribution may very well reflect a larger-than-average estimate resulting from a very small sample of studies. Moreover, while peer-mediated learning provided the largest effect sizes in two of the three distributions (attitudinal and oral language), cooperative was the largest in written language outcomes, which was the distribution with the largest sample of included studies. Thus, even a qualitative analysis of the rank order of the three constructs suggests that no single version of peer-mediated learning was consistently more effective than the others. This actually affirms a theoretical orientation of this meta-analysis, which posits that a sociocultural explanation of the effectiveness of peer-mediated learning, in general, is that it is through mediated interaction that ELLs learn best. However, the fact that peer tutoring and cooperative learning are the two most structured forms of peer-mediated learning also lends tentative support to claims in the literature that high structure promotes the most learning (eg., Oxford 1997, Slavin 1996).

Hypothesis 2b claimed the language setting EFL or ESL, would not significantly moderate the effectiveness of peer-mediated learning for ELLs. Despite significant differences in the two types of settings (e.g., availability of native speakers and amount of exposure to the target language), both fields advocate the use of interactive methods, and consequently, a null hypothesis was forwarded. Empirical evidence across all three available outcome types suggests that the null hypothesis of no difference between EFL and ESL settings cannot be rejected. Setting was not a significant moderator for any of the outcome types; in fact, the significance of the moderator did not even approach significance for any of the distributions. Interestingly, mean effect sizes were actually larger in EFL settings across all three outcome types (i.e., oral language, written language, and attitudinal). This is surprising given that EFL settings provide less exposure to English input and fewer native language models; however, it supports output models of second language acquisition (e.g., Keck et al. 2006, Long 1981, Long 1996, Gass and Mackey 2006, Pica 1994) that suggest that opportunities to formulate meaningful output are as important as opportunities for comprehensible input.

Hypothesis 2c posited no significant difference in the effectiveness of peer-mediated learning at different grade levels. To some extent, this is a participant-level question about the effectiveness of peer-mediated learning with students of different ages, but it is analyzed here as a setting-level moderator to reflect differences in pedagogy and instructional delivery associated with these various grade levels. In practice, this moderator addresses aspects of both setting and participant.

Results of moderator analyses across outcome types provide ambivalent support for this hypothesis. For oral language and attitudinal outcomes, Grade was not a significant moderator, though it was analyzed as different bivariate variables for oral outcomes (i.e., elementary vs. other) and attitudinal outcomes (elementary vs. high school) because of availability of data in each distribution. However, for written language outcomes, which contained sufficient studies to analyze all three grade levels, Grade proved to be a significant moderator of effectiveness ($Q=10.863$, $p=.004$), mostly because the mean effect size was very low for middle school. In fact, middle school was consistently lower than elementary or high school estimates, suggesting that peer-mediated learning might not be as effective for middle school ELLs. This is markedly different than the general pattern for educational intervention studies which tend to report larger effect sizes for middle school than either elementary or high school (Lipsey et al. 2012). This is a particularly troublesome finding because of evidence that suggests middle school ELLs are a vulnerable population at tremendous risk of dropping out as they are confronted with increasingly difficult texts and as the focus of education shifts from learning to read to reading to learn (August et al. 2009, Capps et al. 2005, Cummins 2007, Rubinstein-Avila 2003, Short and Fitzsimmons 2007).

Hypothesis 2d could not be directly tested as a moderator in this meta-analysis because the sample of studies included only studies conducted in classrooms.

Hypothesis 2e posited no significant difference between interventions that were entirely peer-mediated (e.g., Jigsaw) and those for which peer-mediated learning was one component of a complex intervention (e.g., Bilingual Cooperative Integrated Reading Comprehension), and this moderator was intended to test a claim by Slavin that complex interventions like Success for All provide the greatest benefits (e.g., Cheung and Slavin 2005). Moderator analyses across all three outcome types suggest that the null hypothesis of no significant difference cannot be rejected. Similarly, no consistent pattern can be found in a qualitative analyses of the results, as interventions for which peer-mediated learning was just one component were larger on average in two of the distributions (attitudinal and written language) but those for which the entire intervention was peer mediated were larger on average in the distribution or oral language outcomes. This finding does not entirely dismiss claims that there are advantages associated with these large, complex interventions. Rather, as the primary focus of this meta-analysis is determining the effectiveness of peer-mediated learning for ELLs, it appears that peer-mediated learning is effective for ELLs across a number of intervention types, including those that use peer-mediated learning exclusively.

Hypothesis 2f posited no significant difference of the effectiveness of peer-mediated learning for students from differing language backgrounds. Due to limitations in the included sample and the reported data and because culture and language interact in complex ways, student ethnicity was used as a proxy measure of language background. Moderator analyses for all three outcomes suggest that the null hypothesis of no significant difference cannot be rejected. In fact, this variable was tested in two different ways: Hispanic vs. Other and Asian vs. Other. A number of important limitations of these coding categories should be mentioned. First, neither Hispanic nor Asian are monolithic

categories; each contains a wide diversity of language, cultural, and geographic variability. Secondly, comparing these two categories to all others faces the same limitation of masking important variability in language and cultural difference. However, these two were chosen because the included sample contained a particularly large number of Hispanic, or Spanish-speaking, participants, Latinos are the largest group of ELLs in the United States, Asians are the fastest growing group of ELLs in the United States, and because at least some research suggested peer-mediated learning may be ineffective for Asians (e.g.,Than et al. 2008). Regarding the last point, that Asians may be culturally averse to cooperative, Western-based approaches and may actually prefer teacher-centered approaches, qualitative analyses of the Student Asian variable indicate that across all three outcome types, Asian students actually performed better on average than their non-Asian peers. In fact, a majority of these studies were conducted in Asian EFL settings, where cultural norms should be strongest. Thus, the findings of this meta-analysis offer tentative evidence to contradict the claim by Than et al. (2008) that cooperative methods may be culturally inappropriate and ineffective for Asian ELLs.

Hypothesis 2g predicted no significant difference in the effectiveness of peer-mediated learning for students from high- or low-SES backgrounds, and moderator analyses across all three outcome types support this null hypothesis. Notably, SES was analyzed somewhat differently for written language outcomes (i.e., low vs other) than for oral language or attitudinal outcomes because of a lack of sufficient studies in the other two categories. Also, it is noteworthy that for all three outcome types, Unknown was the most frequently coded category, suggesting that findings are somewhat tentative and reflect a lack of careful reporting in the literature base.

Finally, Hypotheses 2h and 2i predicted a significant difference favoring high quality studies. Specifically, 2h posited that high-quality studies (i.e., tested for pre-test differences AND adjusted for pre-test differences) would outperform medium or low-quality studies, and moderator analyses for written language and attitudinal outcomes support this alternative hypothesis. However, study quality was not a significant predictor for oral language outcomes, and medium quality studies actually reported the highest average effect sizes. Thus, moderator analyses provide somewhat ambivalent support for Hypothesis 2h. Hypothesis 2i predicted a significant difference favoring higher dosage studies (i.e., total number of contacts) than for lower dosage studies, and moderator analyses across all three outcome types failed to support this hypothesis. Thus, the null hypothesis of no significant difference could not be rejected for the moderating influence of dosage.

Finally, another study quality moderator, for which there was no a priori hypothesis, proved important: post hoc researcher adjustment, which indicated that this researcher subtracted the post-test mean from the pre-test mean in order to control for unadjusted pre-test differences. Actually, this is the only moderator variable that proved a significant moderator for all three outcome types, and this finding indicates that not controlling for pre-test differences can have a very large impact on effect size estimates.

Research Question 3: In what ways do select issues of power and equity impact the effectiveness of peer-mediated methods?

This third research question is intended to situate the more typical effectiveness findings just discussed within the equity-oriented statement of the problem presented in Chapter 1; that is, the intention of this research question is to expand the typical effectiveness questions of what works, for whom, and under what conditions to include equity-driven variables that the literature indicates are crucial for the academic success of ELLs. To that end, the following hypotheses examine the influence of a number of equity moderators; however, to be clear, the included variables are not exhaustive not does the operationalization of equity implicit in the selection of moderating variables represent the most complex conception of equity available. Rather, these are explorations of equity and how equity-oriented variables may influence the effectiveness of a particular kind of instruction for ELLs.

Hypothesis 3a was an alternative hypothesis that predicted lower effect sizes for ELLs in settings where they are segregated from their peers. This hypothesis is complicated by the fact that many bilingual models intentionally segregate ELLs in order to provide extended, targeted language instruction. Nonetheless, exposure to native language peers offers linguistic, social, and academic advantages that motivate the prediction that ELLs will perform worse in segregated settings. Moderator analyses across the three outcome types offer ambivalent evidence that generally failed to support this hypothesis. However, for oral language outcomes, segregation was a significant moderator, and ELLs demonstrated larger oral language gains in non-segregated settings, as predicted. In fact, qualitative analyses of the written language and attitudinal distributions indicate that non-segregated settings reported higher average effect sizes, which taken with the significant effect for oral language outcomes offers some tentative support to the hypothesis.

As indicated in Table 2, only 5 studies in the included sample indicated whether or not facilities were adequate. Consequently, formal moderator analyses were not possible to test Hypothesis 3b that predicted lower effect sizes for inadequate facilities. Qualitative analysis of the reported effect sizes compared to the means for each of the outcome types also fails to support the hypothesis. Two studies reporting written language outcomes (ES=.386 and ES=.478) were quite close to the mean of .486. Similarly, two studies reporting academic outcomes (ES=.254 and ES=.155) were similar in magnitude to the mean of .25. Finally, one study reporting an oral language outcome (ES=.667) was actually larger than the mean of .578. Given the small number of studies actually reporting the adequacy of facilities, the strongest finding for this hypothesis was the lack of information in the extant literature base.

Similarly, Hypotheses 3c and 3e posited that higher quality teachers would result in more learning gains for ELLs, but very few studies actually reported this information and formal moderator analyses were not possible to test these two hypotheses.

Hypothesis 3d, on the other hand, predicted that culturally-relevant instruction would lead to high learning gains for ELLs. Again, very few studies coded this information, but because the coding was dichotomous and identified whether or not authors made even a cursory claim of cultural relevance, it was possible to code no even when authors did not report the information. Moderator analyses failed to support the hypothesis, however. For attitudinal outcomes, not one study claimed to be even slightly culturally-relevant. For oral language and written language outcomes, qualitative analysis indicates that those studies claiming any cultural relevance actually reported lower effect sizes on average. Overall, the very low bar for coding studies as culturally-relevant resulted in surprisingly few studies coded as culturally relevant, indicating that very little can be said about the moderating effect of strong forms of culturally-relevant instruction on the effectiveness of peer mediation for ELLs.

Finally, Hypothesis 3f predicted that interventions using students' native language would be more effective than those using only English. This represents an empirical test of the application of the largest literature base on equity-oriented effectiveness research for ELLs. That is, five meta-analyses of the effectiveness of using students' native language have consistently found that bilingual models outperform English-only models, and this hypothesis is intended to extend that to a particular instructional approach. As coded for these analyses, moderator analysis across all three outcomes consistently failed to support the assertion that using students' native language produced larger effects than interventions that used only English. Notably, for all three outcome types, one study reported using students' L1 exclusively (see Suppl. material 2). In each case, the effect size for the single study using L1 exclusively was much larger than for bilingual or English-only approaches; however, to provide sufficiently large samples in each moderator category, L1-only and bilingual approaches were combined for moderator analyses. Similarly, qualitative analyses of all three outcome types indicate that interventions using students' native language reported higher mean effect sizes than those using only English. Thus, qualitative analysis across all three outcome types offers some tentative support for the claim that the use of students' native language during instruction promotes the effectiveness of peer mediation for ELLs. Importantly, this variable only measures whether instruction utilized students' native language, but it does not measure whether or not students actively used their L1 during activities or if learning outcomes were greater for students' use of L1.

Overall, the hypotheses about the importance of equity demonstrate that effectiveness research continues to focus on academic and psychological factors to the exclusion of issues of power and equity. Very few studies reported sufficient information to code these variables, and consequently, the claims that could be tested or supported are relatively few and tentative. Despite these shortcomings, analyses offer some support to claims that that equity variables moderate the effectiveness of peer mediation for ELLs. For instance, segregation proved to be a significant moderator for oral language outcomes, and in all three outcome types, segregated settings produced smaller effect sizes that non-segregated settings. Similarly, effect sizes in all three outcome types were larger for interventions that used students' native language for instruction.

Limitations and Future Directions

These findings consistently indicate that peer-mediated learning is effective for ELLs nonetheless, there a number of important limitations to consider. For instance, this meta-analysis is limited by reporting in the original studies, and as discussed many important variables were either excluded from formal analyses or modified in some way because of limitations in the extant literature base. Similarly, these findings are based on a modest sample of studies; and analyses of some outcome types were severely limited by sample size. Future research may benefit from a growing literature base. The lack of statistically significant moderators, for instance, likely represents a lack of statistical power to detect practically meaningful differences rather than strong evidence that no difference actually exists. Future meta-analyses may benefit from the inclusion of additional studies that seem likely to be conducted given the ongoing interest in cooperative learning research for ELLs indicated by the large proportion of recent studies included in this sample.

Furthermore, the inclusion of low- and medium-quality studies may influence the findings, and there are certainly those that argue only the highest-quality studies should be included in research syntheses. As argued, ELLs represent an emergent field of research, and much effort was made to analyze the influence of study quality on the effects reported in this meta-analysis. Of course, all secondary data analyses are limited by the quality of the data they analyze, and this limitation is hardly unique to this particular meta-analysis.

Another limitation common to meta-analyses was availability of studies and data. Considerable effort was made to identify and retrieve the entire population of studies conducted on the effectiveness of peer-mediation, but certainly, some studies were missed. Moreover, some studies deemed relevant and qualified were missing data. Even after attempts to contact the authors, occasionally the studies were too old and even the original authors no longer had access to the data. Similarly, this meta-analysis is a product of its particular time, and search tools (e.g., electronic databases and e-mail) are likely biased towards more recent research. Thus, the findings reported in this meta-analysis are limited by the availability of data, and missing data may affect the internal validity of the result, as well as the ability of the sample to accurately estimate general population parameters.

Finally, a number of variables of interest were operationalized in ways that reflected availability of data or that allowed for reliable coding. However, the operationalizations of these variables likely simplified constructs of interest (e.g., equity); consequently, the findings presented in this study may only be of limited use for those doing research within any one of these fields. Similarly, the expansion of certain constructs (e.g., ELL) to include multiple variables (e.g., ESL and EFL) may affect the generalizability of these findings.

Future research should examine other potential moderators, including setting (e.g., laboratory settings), instructional variables (e.g., task type), teacher (e.g., beliefs and attitudes), and student (e.g., social capital and student use of L1) that are known to influence the effectiveness of peer-mediated methods and the learning of ELLs. Similarly, study quality variables (e.g., fidelity of implementation) were generally under-reported in this sample, and future research should examine the moderating influence these may exert

on the mean effect size. Additionally, future research should explore in more detail the mechanisms that make peer-mediated learning effective for ELLs; for example, why does peer-mediated learning appear more effective at promoting language outcomes than academic outcomes? Clearly, more attention should be paid to important factors like the certification and experience of teachers, the adequacy of the facilities, and the length of residence or previous schooling of ELLs. The nearly complete absence of this data in the literature base for this study marks a knowledge gap that is unacceptable, especially given a clear literature base demonstrating the importance of these variables for ELLs.

Acknowledgements

This work was supported, in part, by Vanderbilt's Experimental Education Research Training (ExpERT) grant (David S. Cordray, Director; grant number R305B040110). I am grateful for the faith the ExpERT folks demonstrated in selecting me, and the financial and technical support and training they provided were instrumental in the completion of this study. In particular Mark W. Lipsey, Director of the Peabody Research Institute, and David S. Cordray, director of the EspERT Program at Vanderbilt University, provided countless hours of guidance and support throughout my time at Vanderbilt.

I am also grateful to the support and guidance of my Dissertation Committee. To David Dickinsono, thank you for the tough, insightful readings of my my MAP and dissertation. You did exactly what I asked you to do. To Bridget Dalton, the kindness with which you tempered your feedback was always welcome, and your questions and comments always drove me to think deeply about some aspect of my work. To Mark Lipsey, your technical expertise was invaluable, and even when suggesting significant changes or advancing challenging critiques, you always made me feel that this work was important and valued. Mostly, to my advisor, mentor, and friend, Robert Jiménez, you have taught me more than I expected to be able to learn these last five years, and in the process you've become more than just another colleague to me. I am inspired by your example and more than grateful for all you've done to prepare me. I look forward to a long career as your colleague.

Finally, I am grateful beyond words for the love and support of my family. To my wife and daughter, you've sacrificed more hours than I care to remember in support of this accomplishment. Your love has been my refuge and my strength through these last five years, and I hope to share with you the fruits that your seeds of love have nurtured. To my parents, both by blood and by bond, your financial and emotional support made so much of this journey possible, and your own lives' works are the models for the work I still hope to accomplish.

References

- Allison B, Rehm M (2007) Effective Teaching Strategies for Middle School Learners in Multicultural, Multilingual Classrooms. Middle School Journal 39 (2): 12-18. https://doi.org/10.1080/00940771.2007.11461619
- August D (1987) Effects of Peer Tutoring on the Second Language Acquisition of Mexican American Children in Elementary School. TESOL Quarterly 21 (4): 717-736. https://doi.org/10.2307/3586991
- August D, Shanahan T (2006) Developing literacy in second-language learners: Report of the National Literacy Panel on language minority children and youth. Lawrence Erlbaum Associates, Inc, Washington, DC: Mahwah, NJ.
- August D, Barnett S, Christian D, Fix M, Frede E, Francis D (2009) The American Recovery and Reinvestment Act: Recommendations for addressing the needs of English language learners. Working Group on ELL Policy URL: http://www.migrationpolicy.org/research/american-recovery-and-reinvestment-act-recommendations-addressing-needs-english-language
- Baca L, Bransford J, Nelson C, Ortiz L (1994) Training, development, and improvement (TDI): A new approach for reforming bilingual teacher preparation. The Journal of Educational Issues of Language Minority Students 14: 1-22.
- Baker KA, Kanter AA (1981) Effectiveness of Bilingual Education: A Review of Literature. Office of Planning, Budget, and Evaluation, U.S. Department of Education., Washington, D.C.
- Ballantyne KG, Sanderman AR, Levy J (2008) Educating English language learners: Building teacher capacity. Washington, DC.
- Bloom HS, Hill CJ, Black AR, 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. https://doi.org/10.1080/19345740802400072
- Calderón M, Hertz-Lazarowitz R, Slavin R (1998) Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. The Elementary School Journal 99 (2): 153-165. https://doi.org/10.1086/461920
- Capps R, Fix M, Murray J, Ost J, Passel J, Herwantoro S (2005) The new demography of America's schools: Immigration and the No Child Left Behind Act. Urban Institute (NJ1) https://doi.org/10.1037/e723122011-001
- Cheung A, Slavin RE (2005) Effective reading programs for English language learners and other language-minority students. Bilingual Research Journal 29 (2): 241-270. https://doi.org/10.1080/15235882.2005.10162835
- Cohen EG (1994) Restructuring the classroom: Conditions for productive small groups. Review of Educational Research 64 (1): 1-35. https://doi.org/10.3102/00346543064001001
- Cummins J (2001) Empowering minority students: A framework for intervention. Harvard Educational Review 71 (4): 649-655. https://doi.org/10.17763/haer.71.4.j261357m62846812

- Cummins J, Bismilla V, Chow P, Cohen S, Giampapa F, Leoni L, Sandhu P, Sastri P (2005) Affirming identity in multilingual classrooms. Educational Leadership 63 (1): 38-43.
- Cummins J (2007) Rethinking monolingual instructional strategies in multilingual classrooms. Canadian Journal of Applied Linguistics 10: 221-240.
- Davies CE (2003) How English learners joke with native speakers: An interactional sociolinguistic perspective on humor as collaborative discourse across cultures. Journal of Pragmatics 35: 1361-1385. https://doi.org/10.1016/S0378-2166(02)00181-9
- Deyhle D (1995) Navajo youth and Anglo racism: Cultural integrity and resistance. Harvard Educational Review 65 (3): 403-444. https://doi.org/10.17763/haer.65.3.156624q12053470n
- Digest of Education Statistics (2009) Average reading scale scores of 4th- and 8th-graders in public schools and percentage scoring at or above selected reading achievement levels, by English language learner (ELL) status and state or jurisdiction: 2007. http://nces.ed.gov/programs/digest/d09/tables/dt09_124.asp. Accessed on: 2010-12-13.
- Dion E, Fuchs D, Fuchs LS (2007) Differential effects of peer-assisted learning strategies on students' social preference and friendship making. Behavioral Disorders 4.
- Duff P (2001) Language, literacy, content, and (pop) culture: Challenges for ESL students in mainstream courses. Canadian Modern Language Review/Revue canadienne des langues vivantes 58 (1): 103-132. https://doi.org/10.3138/cmlr.58.1.103
- Echevarria J, Short D, Powers K (2006) School reform and standards-based education: A model for English-language learners. The Journal of Educational Research 99: 195-210. https://doi.org/10.3200/JOER.99.4.195-211
- Fantuzzo JW, Riggio RE, Connely S, Dimeff LA (1989) Effects of reciprocal peer teaching on academic achievement and psychological adjustment: A component analysis. Journal of Educational Psychology 81 (2): 173-177. https://doi.org/10.1037/0022-0663.81.2.173
- Firth A, Wagner J (1997) On discourse, communication, and (some) fundamental concepts in SLA research. The Modern Language Journal 81 (3): 285-300. https://doi.org/10.1111/j.1540-4781.1997.tb05480.x
- Fradd SH, Lee O (2003) Teachers' roles in promoting science inquiry with students from diverse language backgrounds. Educational Researcher 28 (6): 14-20. https://doi.org/10.2307/1177292
- Francis DJ, 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. Portsmouth, New Hampshire.
- Gándara P (2000) In the aftermath of the storm: English learners in the post-227 era. Bilingual Research Journal 24: 1-14. https://doi.org/10.1080/15235882.2000.10162748
- Gándara P, Rumberger R, Maxwell-Jolly J, Callahan R (2003) English learners in California schools: Unequal resources, unequal outcomes. Educational Policy Analysis Archives 11 (36): 1-54.
- Gass SM, Mackey A (2006) Input, Interaction and Output: An Overview. AILA Review 19: 3-17. https://doi.org/10.1075/aila.19.03gas
- 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 (4): 363-386. https://doi.org/10.1207/s15327671espr1004_2

- Gersten R, Baker S (2000) What we know about effective instructional practices for English-language learners. Exceptional Children 66 (4): 454-470. https://doi.org/10.1177/001440290006600402
- Gifford F, Valdés G (2006) The linguistic isolation of Hispanic students in California's public schools: The challenge of reintegration. The Annual Yearbook of the National Society for the Study of Reintegration 125-154.
- Gitlin A, Buendía E, Crosland K, Doumbia F (2003) The production of margin and center: Welcoming-unwelcoming of immigrant students. American Educational Research Journal 40 (1): 91-122. https://doi.org/10.3102/00028312040001091
- Goodlad S (1998) Mentoring and Tutoring by Students. Routledge, 330 pp. https://doi.org/10.4324/9780203761212
- Greene J (1998) A Meta-Analysis of the Effectiveness of Bilingual education. Tomas Rivera Policy Institute, Claremont, CA.
- Gutiérrez KD, Larson J, Kreuter B (1995) Cultural tensions in the scripted classroom: The value of the subjugated perspective. Urban Education 29 (4): 410-442. https://doi.org/10.1177/0042085995029004004
- Gutiérrez KD, Baquedeño-Lopez P, Asato J (2000) "English for the Children": The new literacy of the old world order, language policy and educational reform. Bilingual Research Journal 24: 87-116. https://doi.org/10.1080/15235882.2000.10162753
- Harklau L (2000) From the 'good kids' to the 'worst': Representations of English language learners across educational settings. TESOL Quarterly 34 (1): 35-67. https://doi.org/10.2307/3588096
- Harper CA, Jong EJ (2009) English language teacher expertise: The elephant in the room. Language and Education 23 (2): 137-151. https://doi.org/10.1080/09500780802152788
- Hedges LV (2007) Effect sizes in cluster-randomized designs. Journal of Educational and Behavioral Statistics 34 (4): 341-370. https://doi.org/10.3102/1076998606298043
- Hertz-Lazarowitz R, Kirkus VB, Miller N (1992) An overview of the theoretical anatomy of cooperation in the classroom. In: N. Miller RH (Ed.) Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning. Cambridge University Press, New York.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327: 557-560. https://doi.org/10.1136/bmj.327.7414.557
- Iddings AC, McCafferty SG (2007) Carnival in a mainstream kindergarten classroom: A Bahktinian analysis of second language learner's off-task behaviors. The Modern Language Journal 91: 31-44. https://doi.org/10.1111/j.1540-4781.2007.00508.x
- Johnson DW, Maruyoma G, Johnson R, Nelson D, Skon L (1981) The effects of cooperative, competitive, and individualistic goal structures on achievement: a meta-analysis. Psychological Bulletin 89 (1): 47-62. https://doi.org/10.1037/0033-2909.89.1.47
- Johnson DW, Johnson RT, Stanne MB (2000) Cooperative learning methods: A meta-analysis. http://www.co-operation.org/pages/cl-methods.html. Accessed on: 2009-1-21.
- Kamberelis G (1986) Emergent and Polyphonic Character of Voice in Adolescent Writing. Emergent and Polyphonic Character of Voice in Adolescent Writing. National Reading Conference, Austin, Texas.

- Kamberelis G (2001) Producing heteroglossic classroom (micro)cultures through hybrid discourse practice. Linguistics and Education 12: 85-125. https://doi.org/10.1016/S0898-5898(00)00044-9
- Keck C, Iberri-Shea G, Tracy-Ventura N, Wa-Mbaleka S (2006) Investigating the empirical link between task-based interaction and acquisition. Synthesizing research on language learning and teaching 13: 91. https://doi.org/10.1075/lllt.13.08kec
- Kluge D (1999) A brief introduction to cooperative learning. In: Kluge D, McGuire S, Johnson D, Johnson R (Eds) Cooperative Learning. Japan Association for Language Teaching, Tokyo, 16-22 pp.
- Lantolf JP (2000) Second language learning as a mediated process. Language Teaching 33: 79-96. https://doi.org/10.1017/S0261444800015329
- Leki I (2001) “A narrow thinking system”: Nonnative-English-speaking students in group projects across the curriculum. TESOL Quarterly 35 (1): 39-67. https://doi.org/10.2307/3587859
- Lensmire TJ (1998) Rewriting student voice. Journal of curriculum Studies 30 (3): 261-291. https://doi.org/10.1080/002202798183611
- Lipsey MW, Wilson DB (2001) Practical Meta-Analysis. Sage Publications, Inc, CA.
- Lipsey MW, Puzio K, Yun C, Hebert MA, Steinka-Fry K, Cole MW, Roberts M, Anthony KS, Busick MD (2012) Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf
- Lipsky M (2010) Street-level Bureaucracy: Dilemmas of the Individual in Public Services. Russell Sage Foundation, New York.
- Long MH (1981) Input, interaction, and second-language acquisition. Annals of the New York Academy of Sciences 379: 259-278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x
- Long MH (1996) The Role of the Linguistic Environment in Second Language Acquisition. Handbook of Second Language Acquisition. https://doi.org/10.1016/b978-012589042-7/50015-3
- Macedo D (1994) Literacies of power: What Americans are not allowed to know. Westview Press
- Mackey A, Goo J (2007) Interaction research in SLA: A meta-analysis and research synthesis. Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies. Oxford University Press, New York.
- Mathews RS, Cooper JL, Davidson N, Hawkes P (1995) Building bridges between cooperative and collaborative learning. Change 27 (4): 34-40.
- Maxwell-Jolly J (2000) Factors influencing implementation of mandated policy change: Proposition 227 in seven northern California school districts. Bilingual Research Journal 24 (1): 37-56. https://doi.org/10.1080/15235882.2000.10162750
- McKeon D (2005) Research Talking Points on English Language Learners. 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
- Moll LC, Amanti C, Neff D, Gonzales N (1992) Funds of Knowledge for teaching: Using a qualitative approach to connect homes and classrooms. Theory into Practice 31 (2): 132-141. https://doi.org/10.1080/00405849209543534

- Morita N (2004) Negotiating participation and identity in second language academic communities. TESOL Quarterly 38 (4): 573-603. https://doi.org/10.2307/3588281
- National Center for Education Statistics (2011) The Condition of Education 2011. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011033
- Norton B (1997) Language, identity, and the ownership of English. TESOL Quarterly 31 (3): 409-429. https://doi.org/10.2307/3587831
- Norton B, Toohey K (2001) Changing perspectives on good learners. TESOL Quarterly 35 (2): 307-322. https://doi.org/10.2307/3587650
- Ogbu JU, Simons HD (1998) Voluntary and involuntary minorities: A cultural-ecological theory of school performance with some implications for education. Anthropology & Education Quarterly 29 (2): 155-188. https://doi.org/10.1525/aeq.1998.29.2.155
- Oortwijn MB, Boekaerts M, Vedder P, Strijbos J (2008) Helping behavior during cooperative learning and learning gains: The role of the teacher and of pupils' prior knowledge and ethnic background. Learning and Instruction 18: 146-159. https://doi.org/10.1016/j.learninstruc.2007.01.014
- Ovando CJ (2003) Bilingual education in the United States: Historical development and current issues. Bilingual Research Journal 27 (1): 1-24. https://doi.org/10.1080/15235882.2003.10162589
- Oxford R (1997) Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. The Modern Language Journal 81 (40): 456.
- Pavlenko A, Norton B (2007) Imagined communities, identity, and English language learning. In: Cummins J, Davies C (Eds) International Handbook of English Teaching. Springer, Dordrecht, Netherlands.
- Pica T (1994) Research on Negotiation: What Does It Reveal About Second-Language Learning Conditions, Processes, and Outcomes? Language Learning 44 (3): 493-527. https://doi.org/10.1111/j.1467-1770.1994.tb01115.x
- Platt E, Harper C, Mendoza MB (2003) Dueling philosophies: Inclusion or Separation for Florida's English language learners. TESOL Quarterly 37 (1): 105-133. https://doi.org/10.2307/3588467
- Portes A, Rumbaut RG (2001) Legacies: The Story of the Immigrant Second Generation. The Russell Sage Foundation, New York.
- Prior P (2001) Voices in text, mind, and society. Journal of Second Language Writing 10: 55-81. https://doi.org/10.1016/S1060-3743(00)00037-0
- Ramírez JD, Yuen SD, Ramey DR (1991) Final report: Longitudinal study ofStructured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. A technical report prepared for the United States Department of Education, Washington, DC.
- Rohrbeck CA, Fantuzzo JW, Ginsberg-Block MD, Miller TR (2003) Peer-assisted learning interventions with elementary school students: A meta-analytic review. Journal of Educational Psychology 95 (2): 240-257. https://doi.org/10.1037/0022-0663.95.2.240
- Rollinson P (2003) Using peer feedback in the ESL writing class. ELT Journal 59 (1): 23-30. https://doi.org/10.1093/elt/cci003
- Rolstad K, Mahoney K, Glass GV (2005) The big picture: A meta-analysis of program effectiveness research on English language learners. Educational Policy 19: 572-594. https://doi.org/10.1177/0895904805278067

- Roseth CJ, Johnson DW, Johnson RT (2008) Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. Psychological Bulletin 134 (2): 223-246. https://doi.org/10.1037/0033-2909.134.2.223
- Rossell CH, Baker K (1996) The educational effectiveness of bilingual education. Research in the Teaching of English 30 (1): 7-74.
- Rubinstein-Avila E (2003) Conversing with Miguel: An adolescent English language learner struggling with later literacy development. Journal of Adolescent & Adult Literacy 47 (4): 290-301.
- Ruiz VL (2001) South by southwest: Mexican Americans and segregated schooling, 1900-1950. Magazine of History 15: 23-27. https://doi.org/10.1093/maghis/15.2.23
- Rumbaut RG, Portes A (2001) Ethnicities: Children of Immigrants in America. California: University of California Press
- Sáenz LM, Fuchs LS, Fuchs D (2005) Peer-assisted learning strategies for English language learners with learning disabilities. Exceptional Children 71 (3): 231-247. https://doi.org/10.1177/001440290507100302
- Schmid CL (2001) The Politics of Language: Conflict, Identity, and Cultural Pluralism in Comparative Perspective. Oxford University Press, New York.
- 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. https://www.carnegie.org/media/filer_public/bd/d8/bdd80ac7-fb48-4b97-b082-df8c49320acb/ccny_report_2007_double.pdf
- Slavin RE (1986) Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher 15 (9): 5-11. https://doi.org/10.3102/0013189X015009005
- Slavin RE (1990) Achievement effects of ability grouping in secondary schools: A best evidence synthesis. Review of Educational Research 60 (3): 471-499. https://doi.org/10.3102/00346543060003471
- Slavin RE (1996) Research on cooperative learning and achievement: What we know, what we need to know. Contemporary Educational Psychology 21: 43-69. https://doi.org/10.1006/ceps.1996.0004
- Slavin RE, Cooper R (1999) Improving intergroup relations: Lessons learned from cooperative learning programs. Journal of Social Issues 55 (4): 647-663. https://doi.org/10.1111/0022-4537.00140
- Slavin RE, Cheung A (2005) A synthesis of research on language of reading instruction for English language learners. Review of Educational Research 75 (2): 247-284. https://doi.org/10.3102/00346543075002247
- Slavin RE, Madden NA (2011) Measures inherent to treatments in program effectiveness reviews. Journal of Research on Educational Effectiveness 4 (4): 370-380. https://doi.org/10.1080/19345747.2011.558986
- Sleeter C (2001) Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness. Journal of Teacher Education 52: 94-106. https://doi.org/10.1177/0022487101052002002
- Smith ML, Glass GV, Miller TI (1980) Benefits of psychotherapy. Johns Hopkins University, Baltimore, MD.
- Solano-Flores G (2008) Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language

learners. Educational Researcher 37: 189-199. https://doi.org/10.3102/0013189X08319569

- Sterne JA, Egger M (2006) Regression Methods to Detect Publication and Other Bias in Meta-Analysis. Publication Bias in Meta-Analysis. https://doi.org/10.1002/0470870168.ch6
- Stritikus T, Garcia E (2000) Education of limited English proficient students in California schools: An assessment of the influence of Proposition 227 on selected teachers and classrooms. Bilingual Research Journal 24: 75-85. https://doi.org/10.1080/15235882.2000.10162752
- Swain M, Brook L, Tocalli-Beller A (2002) Peer-peer dialogue as a means of second language learning. Annual Review of Applied Linguistics 22: 171-185.
- Talmy S (2004) Forever FOB: The cultural production of ESL in a high school. Pragmatics 14: 149-172. https://doi.org/10.1075/prag.14.2-3.03tal
- Talmy S (2008) The cultural productions of the ESL student at Tradewinds High: Contingency, multidirectionality, and Identity in L2 socialization. Applied Linguistics 29 (4): 619-644. https://doi.org/10.1093/applin/amn011
- Than PT, Gillies R, Renshaw P (2008) Cooperative learning (CL) and Academic achievement of Asian students: A true story. International education Studies 1 (3): 82-88.
- Thomas WP, Collier VP (2004) The astounding effectiveness of dual language education for all NABE. Journal of Research and Practice 2 (1): 1-20.
- Tijerino A, Asato J (2002) The implementation of Proposition 227 in California schools: A critical analysis of the effect on teacher beliefs and classroom practices. Equity & Excellence in Education 35 (2): 108-118. https://doi.org/10.1080/713845279
- Valdés G (2001) Learning and not learning English: Latino students in American schools. Teachers College Press, New York.
- Valenzuela A (1999) Subtractive schooling: U.S. Mexican youth and the politics of caring. State University of New York Press, New York.
- Voloshinov VN (1973) Marxism and the philosophy of language. Marxism and the Philosophy of Language. Harvard University Press, Cambridge, MA.
- Wiese A, García E (1998) The Bilingual Education Act: Language minority students and equal educational opportunity. Bilingual Research Journal 22 (1): 1-18. https://doi.org/10.1080/15235882.1998.10668670
- Willig A (1985) A Meta-Analysis of Selected Studies on the Effectiveness of Bilingual Education. Review of Educational Research 55 (3): 269. https://doi.org/10.2307/1170389
- Yoon B (2008) Uninvited Guests: The Influence of Teachers' Roles and Pedagogies on the Positioning of English Language Learners in the Regular Classroom. American Educational Research Journal 45 (2): 495-522. https://doi.org/10.3102/0002831208316200

Supplementary materials

Suppl. material 1: Coding forms doi

Authors: Mikel Cole
Data type: coding forms
Filename: Coding Forms_Master Document.xlsx - Download file (215.53 kb)

Suppl. material 2: Included and Near-Miss Studies doi

Authors: Mikel Cole
Data type: List of references
Brief description: These are studies that were potentially-relevant to the meta-anlayses, but that were ultimately excluded during inclusion coding. Future researchers might find this list especially valuable.
Filename: Included and Near Miss Studies_Dissertation-3.pdf - Download file (569.45 kb)

Endnotes

*1 English language learner is only one of many terms that refer to linguistically diverse students. Other terms like Limited-English proficient and language minority convey deficiency-oriented or disempowering.

*2 It is worth noting that "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 et al. 2007).

*3 The important theoretical issues raised in this meta-analysis are largely distinct from the questions analyzed and synthesized in the Major Area Paper to which this comment refers. However, the idea that sociocultural theory might prove heuristically useful is explored in this paper. Thus, little explanation for this bias is given here, and readers are encouraged to examine the evidence that warrants this presumption.

*4 Nonetheless, this meta-analysis primarily employs the term effectiveness to emphasize the ability of peer-mediated approaches to improve outcomes for ELLs on discrete measures or instruments, even when those measures assess constructs like out-group relations.

*5 Notably, this is the same research that informed the historic *Brown v Board* decision that created the legal foundation for the desegregation, if not integration, of public schools in the United States.

*6 The authors actually report the inverse-variance adjustment for small samples as d+, but it is based on Hedge's original work and is more commonly referred to as Hedges' *g*; as such, figures are reported here as *g*.

*7 It is important to distinguish this assertion from a deficit view of ELLs. Asserting that English proficiency is a barrier to mainstream instruction is not intended to be equivalent to an assertion that ELLs are deficient learners. All ELLs come to school proficient in at least one language, and many are proficient in several. Rather, like the landmark ruling in Lau v Nichols, the assertion is intended to indicate that most

instruction in the US is provided in English by monolingual, White teachers; and without affirmative efforts to make the curriculum accessible to ELLs, these students do not generally have a chance to succeed in most US classrooms.

*8 While theoretically distinct, the more individualistic and cognitive orientations (e.g., traditional second language acquisition interaction and cooperative learning) and the more socially-oriented (e.g., sociocultural theory) perspectives share conceptual common ground. Thus, although the theoretical differences are acknowledged, the assertion of a conceptual common ground enables the inclusion of studies from all three theoretical orientations.

*9 While there was some discussion of second language and foreign language differences in the results, the authors report too few FL settings to make substantial claims.