American Psychological Association

Journal of Educational Psychology

© 2019 American Psychological Association
ISSN: 0022-0663

2020, Vol. 112, No. 7, 1303–1319
http://dx.doi.org/10.1037/edu0000436

# How Effective Is Peer Interaction in Facilitating Learning? A Meta-Analysis

Harriet R. Tenenbaum and Naomi E. Winstone
University of Surrey

Patrick J. Leman
King's College London

Rachel E. Avery
University of Surrey

Decades of research indicate that peer interaction, where individuals discuss or work on a task collaboratively, may be beneficial for children's and adolescents' learning. Yet, we do not know which features of interaction may be related to learning from peer interaction. This meta-analysis examined results from 62 articles with 71 studies into peer interaction, involving a total of 7,103 participants aged 4 to 18 years. Peer interaction was effective in promoting learning in comparison with other types of learning conditions, Hedges' $g = 0.40$, 95% confidence interval [CI: 0.27, 0.54], $p < .0001$, across different gender and age groups. In contrast, however, peer interaction was not more effective than child–adult dyadic interaction. Moderator analyses also indicated that peer interaction is more effective when children are specifically instructed to reach consensus than when they are not. Findings extend theoretical considerations by teasing apart the processes through which children learn from peer interactions and offer practical implications for the effective use of peer interaction techniques in the classroom.

***Educational Impact and Implications Statement***
We conducted a meta-analysis (statistical review of findings from previous research) of 71 studies to investigate whether children and adolescents learn from doing an educational task together in a group with other children and adolescents. We found that children and adolescents learned more when they completed a task with their peers compared with children and adolescents completing the same task alone or serving as a comparison group who were simply tested twice. Children and adolescents evidenced greater learning when they were asked to come to agreement than when they were not. The findings suggest that simple instructions for peer interaction in the classroom may increase learning.

*Keywords:* meta-analysis, peer interaction, conversation

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Individuals of different ages often learn through interaction with peers. In children, peer interaction can facilitate cognitive development (Topping, Buchs, Duran, & Van Keer, 2017). However, the dynamics of interaction between children who are engaged in learning activities are often not systematically managed or planned in the classroom (Baines, Blatchford, & Chowne, 2007). Moreover, peer interaction is not always effective (e.g., Chan, 2001; Stacey, 1992). In this article, we present a meta-analysis of the effectiveness of peer interaction across a variety of areas of learning to explore moderators of the effectiveness to inform theory and practice.

## Developmental Theories: Peer Interaction and Intersubjectivity

Peer interaction is typically conceptualized as stemming from constructivist theories of learning, such as the radical constructivism of Jean Piaget, and the social constructivism of Lev Vygotsky. Both theories and the subsequent research inspired by them place peer interaction at the heart of many developmental processes (Tudge & Rogoff, 1999). In both, the notion of intersubjectivity, that is, shared meaning between interactional partners, is key to explaining possible learning benefits from peer interaction. Piaget (e.g., Piaget, 1932, 1959) viewed peer interaction, distinct from adult–child interaction, as an important means of promoting intersubjectivity and subsequent social development. For Piaget, the child is an active participant in constructing knowledge. According to Piaget's approach, adult–child interaction is less effective in

This article was published Online First December 12, 2019.

Harriet R. Tenenbaum and Naomi E. Winstone, School of Psychology, University of Surrey; Patrick J. Leman, Institute of Psychiatry, Psychology and Neuroscience, King's College London; Rachel E. Avery, School of Psychology, University of Surrey.

We would like to acknowledge the help of Stefania Tzourmana and Orianna Bairaktari.

Correspondence concerning this article should be addressed to Harriet R. Tenenbaum, School of Psychology, University of Surrey, Surrey, GU2 7XH, United Kingdom. E-mail: h.tenenbaum@surrey.ac.uk

promoting learning because the adult's natural authority leads to asymmetry that renders the child a more passive recipient of knowledge and instruction (Piaget, 1932). With peers, in contrast, children are better able to question others as free and active participants in social discourse, argument, and learning (Castellaro & Roselli, 2015). It is this active engagement from all interlocutors that leads to gains in learning. Studies across a range of areas of learning, from both Piagetian (e.g., Dimant & Bearison, 1991; Druyan, 2001; Golbeck & Sinagra, 2000; Kruger, 1992; Light & Littleton, 1994; Slavin, 1992) and from Vygotskian (e.g., Garton & Pratt, 2001; Samaha & De Lisi, 2000; Tudge, Winterhoff, & Hogan, 1996; Underwood, Underwood, & Wood, 2000) frames have demonstrated benefits for peer interaction over and above individual or independent learning.

In line with Piaget's theory, research into peer interaction and social learning has sought to demonstrate the benefits of conversation between peers on developmental processes. For instance, Doise, Mugny, and Perret-Clermont (1975) demonstrated the benefits of peer interaction over independent learning for children's grasp of concepts of conservation. Doise and Mugny (1984) have argued that peer interaction is beneficial because it generates sociocognitive conflict (akin to cognitive conflict) that prompts children to appreciate and consider another child's different perspective and, in turn, develop by adjusting their understanding of a situation accordingly. Subsequent researchers have sought to explore how far social processes and aspects of the child's social identity can promote or inhibit development (e.g., Psaltis, Duveen, & Perret-Clermont, 2009) or the ways in which children at different ages respond to the social dynamics of peer interaction (Leman, 2015).

Similarly, Vygotsky (1978) emphasized the role of intersubjectivity in development. Subsequent approaches, developed out of Vygotsky's theories, have focused on development as a process imbued with social and cultural context. While the sociocultural context remains vital to conceptualize learning and development, Vygotsky proposed a zone of proximal development (e.g., Nyikos & Hashimoto, 1997) where interactions between children (interpsychological) are appropriated into cognitive and metacognitive processes (intrapsychological). Thus, children learn from peers with relative expertise through sharing of knowledge and a mutual understanding that is generated in social interaction (Topping, 2005).

A key difference between Vygotsky's (1978) and Piaget's (1932) approaches lies in the role of adults and peers in interaction. Piaget emphasized the ways in which adult–child interaction could present potential obstacles to engagement in learning. For instance, in adult–child interactions, adults may dominate and children may comply more readily with adults' than other children's instruction. Vygotsky was more positive in arguing that adults could stimulate individual children's learning. Indeed, some studies from a Vygotskian perspective suggest that individual children learn more from interaction with adults than peers (Radziszewska & Rogoff, 1988, 1991). This meta-analysis makes it possible to evaluate these contrasting positions through particular moderators, such as whether children need to build consensus in groups and to whom peer interaction is compared (i.e., individual vs. child–adult comparisons).

Peer interaction is central to programs such as peer tutoring and reciprocal whole-class peer tutoring, which involve structured forms of interaction. In peer tutoring programs, one child, who is typically older or more advanced cognitively, is trained to work with another child. Across 72 studies of peer tutoring, Leung (2015) found that peer tutoring was effective, with a medium effect size (Hedge's $g = .59$).

Classwide reciprocal peer tutoring has also been found to effective (e.g., Greenwood, Delquadri, & Carta, 1988) in increasing children's learning. One program that incorporates classwide tutoring is Peer Assisted Learning Strategies (PALS), which involves reciprocal whole-class peer tutoring. In PALS, two pupils with different levels of achievement work together while the roles of tutor and tutee are alternated (e.g., Fuchs, Fuchs, & Burish, 2000). Children are taught to work in pairs and additionally receive targeted subject lessons from teachers, such as summarizing paragraphs and making predictions (Fuchs et al., 2000; Fuchs, Fuchs, Mathes, & Martinez, 2002; Fuchs, Fuchs, Kazdan, & Allen, 1999). PALS programs have been found to produce positive learning outcomes, particularly with younger children, and low-income and ethnic minority urban children (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003).

One concern is that such programs require significant time and resource investment from teachers, and differ from the more common naturally occurring interaction that takes place in everyday learning contexts. In typical classrooms, children spontaneously engage in interaction while completing learning tasks, and teachers often directly encourage children to engage in conversation. It is the effectiveness of these conversations that take place between untrained peers that forms the focus of our analysis.

## Exploring the Efficacy of Peer Interaction

A key feature of our focus is that it is characterized by untutored, bidirectional interaction, involving "the use of small groups of students working together to achieve common goals of learning" (Topping et al., 2017, p. 5). This stands in contrast to situations where children are trained to learn through interaction (i.e., cooperative learning; see Roseth, Johnson, & Johnson, 2008), or where one peer tutors another (i.e., peer tutoring; see Leung, 2015).

We chose to focus specifically on these more untutored forms of interactive inquiry learning because this reflects the kind of learning that might be influenced by peer interaction in typical classrooms. From this perspective, a meta-analysis is important on both theoretical and practical grounds. On a theoretical level, such an approach will isolate the effects of peer interaction in typical learning situations from environments where children are trained before the peer conversation to engage in interaction. Indeed, understanding the effectiveness of untrained peer conversations compared with child–adult conversations informs us about the degree to which Piagetian versus Vygotskian theoretical perspectives capture children's learning. In particular, in this approach we are able to explore the cognitive mechanisms underpinning learning through peer interaction. From a Piagetian view, peer interaction facilitates learning because contrasting viewpoints create sociocognitive conflict. If it is the subsequent restoration of equilibrium that is fundamental to this process, then we would expect greater learning where the task specifically requires children to reach consensus through interaction. This approach also leads to very clear practical implications; as unstructured peer

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

interaction is more common in everyday pedagogic practice, understanding the moderators of its effectiveness can inform teachers how best to facilitate learning through peer interaction, without engaging in more resource-intensive schemes that involve training and further input.

Previous narrative reviews of collaborative learning identify factors that may act as moderators of the extent to which learning gains result from peer interaction (e.g., Davis & Winstone, 2017; Sills, Rowse, & Emerson, 2016). Such factors include the age of the children involved in interaction, the gender mix of the group, group size, and the area of learning. We also examined characteristics of the studies themselves. These factors fall into three broad categories of moderators: study design factors (assessment pattern, posttest delay, comparison group, number of sessions of peer interaction, and consensus required); interactional dynamics factors (age, gender, and group size); and area of learning.

## Study Design Factors

The potential moderating effects of these factors can inform both theoretical and practical understanding of the impact of peer interaction, and under which circumstances these effects are most reliably demonstrated.

**Assessment pattern.** Contrasting theoretical approaches, alongside practical considerations about how studies can be designed for developing classroom interventions, have also led to differing approaches to the design of studies into peer interaction. Classic Genevan studies from a Piagetian perspective typically used a classic pretest and posttest design (e.g., Doise & Mugny, 1984), often outside of a traditional classroom learning situation. In contrast, some studies, especially in more applied settings, may include only learning or developmental outcomes for children at posttest (i.e., without a measure of knowledge and ability before interaction). Given that pretest assignment controls for differences between groups, we expected greater change from peer interaction compared with the control when pretest measures were taken than when they were not.

**Posttest delay.** A further aspect of variation in studies of peer interaction concerns when learning gains are assessed. In many classic experimental designs, posttest measures are taken immediately after peer interactions (e.g., Messer, Joiner, Loveridge, Light, & Littleton, 1993). Differences between peers working together and individuals working alone may be small and statistically nonsignificant when tested immediately after an interaction (e.g., To, 2015) or even favor individuals working alone (e.g., Messer et al., 1993). Other times, assessments may be taken from when peers work together (Leman & Oldham, 2005). In such cases, better performance again may not be observed in peer interaction compared with individuals working alone. Howe, McWilliam, and Cross (2005) posited that learning gains are greater after a delay because children have time to incubate ideas discussed during peer interaction, which primes further learning. In support of this idea, Howe et al. (2005) assessed children's scientific reasoning after a delay arguing that gains are more apparent after a larger interval of time than immediately after an interaction. Based on Howe et al.'s incubation argument, we posit that learning gains should be greater after a delay than when assessed either during a group task or immediately afterward.

**Comparison group.** Another way in which studies differ is the comparison group. In the peer interaction learning literature, there are three separate groups to whom peers interacting together are compared. First, sometimes the comparison group completes an identical task individually. Second, the comparison group may be a true control group who do not complete a related task. For example, Leman, Skipper, Watling, and Rutland (2016) had children work on science tasks with others or not complete any related tasks. Not surprisingly, children who engaged in the related learning task outperformed children who did not engage in learning related science material. More typically, however, children learning in pairs or groups are compared with children who engaged in a similar activity independently (Almasi, 1995; Asterhan, Schwarz, & Cohen-Eliyahu, 2014; Bertucci, Conte, Johnson, & Johnson, 2010; Leman & Oldham, 2005). When To (2015) compared students working alone on the same task to those working in groups, there was no statistically significant advantage for those working in groups. Thus, we expect peer interaction to show greater effect sizes when students serve as a true control than study the material on their own.

The third comparison group used in the peer interaction literature is children working individually with an adult. For example, Radziszewska and Rogoff (1988) assigned 9- and 10-year-children to complete a planning task with their mother or a peer. Children who worked with their mothers were found to learn more efficient planning strategies than those who worked with peers. Vygotskian (Vygotsky, 1978) perspectives suggest that individual children learn more from adults than from peers who are of a more similar level. Through dyadic interaction, adults should be well-placed to scaffold children's learning (Wood & Middleton, 1975). More frequently, however, the literature on peer interaction compares peers to individual children working on their own (e.g., Blaye, Light, Joiner, & Sheldon, 1991; Russell, 1981; To, 2015). From a Vygotskian perspective, we expected that children would learn more from dyadic adult–child interaction than peer interaction.

**Frequency.** We would expect that greater opportunities for peer interaction might lead to greater learning gains. In a few studies, children worked collaboratively on more than one occasion (e.g., Howe et al., 2005). As a result, we might expect greater gains in studies with additional peer interaction sessions.

**Requirement for consensus.** Studies of peer interaction differ according to whether task instructions require children to reach consensus as part of their dialogue. For example, in the moral reasoning domain, children made greater gains on a moral judgment task where they were required to reach a consensual answer compared with engaging in open-ended discussion (Maitland & Goldman, 1974). Consensus building should lead to greater sociocognitive conflict and subsequent restoration of equilibrium (Piaget, 1932). We expect that the process of reaching consensus will be more beneficial for children's learning through interaction than not being asked to reach consensus.

## Interactional Dynamics

Interactional dynamics are likely to be important because peer interaction involves the coordination of shared activity toward a goal. From a theoretical perspective, the moderating effects of these factors can inform understanding of the impact of the degree of symmetry or asymmetry between interactional partners. From a

practical perspective, these factors can inform pedagogic practice involving peer interaction.

**Age.** The strategies used for collaboration may differ with age shifting (at 9–10 years) from a view of interaction as an opportunity to transmit knowledge, to a view of interaction as meaningful collaboration and coconstruction (Leman, 2015). As a result, we would expect age differences in children's ability to collaborate and their ability to learn from such collaborations. More specifically, we would expect older children to learn more from peer interactions than younger children.

**Gender.** Gender is another factor that might influence the effectiveness of peer collaboration. A significant body of research has explored the degree to which the gender of children engaged in interaction affects learning (e.g., Strough, Berg, & Meegan, 2001). Such research suggests that the conversation and interaction styles of girls and boys differ (Leman, Ahmed, & Ozarow, 2005; Leman & Tenenbaum, 2017) such that girls participate more when not outnumbered by or partnered with boys (Webb, 1989). However, certainly by later childhood, these conversation and interaction effects do not appear to feed through to influence learning outcomes (Leman & Björnberg, 2010). Understanding the role of gender in peer interaction adds an important dimension to developmental theory because it can demonstrate the significance of social identities and social dynamics in peer interaction. In addition, understanding the part played by gender is an important consideration for educators to optimize the learning gains of their pupils. Thus, we explored the gender composition of the dyad as a predictor of learning gains.

**Group size.** Another key aspect of social context is the size of a group engaged in interaction. Pairs of children (i.e., groups of two) will likely facilitate higher quality and more intensive interaction than larger groups because it is difficult for one child not to be engaged (Webb, 1989). However, more children involved in interaction also increases the possibility of different perspectives being introduced to a conversation. In this respect, more perspectives may bring greater opportunity for the transfer of knowledge or could generate greater sociocognitive conflict (Fuchs, Fuchs, Kazdan et al., 2000). It may be that there is an optimal group size; for instance, more than six children in a group may lead to *social loafing* (i.e., exerting less effort than optimal; see Topping et al., 2017). We expected that children interacting in larger groups would outperform those in smaller groups.

## Area of Learning

The area of learning also varies across studies of peer interaction. Research has examined peer interaction in areas including conservation, creativity, moral judgments, general reasoning, scientific reasoning, mathematical reasoning, spatial reasoning, spatial conservation, and memory tasks. Area effects may be attributable to the different learning objectives or complexity or specificity of the subject matter (e.g., Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) and, thus, be more or less amenable to learning in a social, interactional context.

To answer the question of the degree to which peer interactions is an effective pedagogical method, we conducted a meta-analysis seeking to address the following research aims:

1. Does research support the effectiveness of peer interaction in facilitating learning in comparison to other types of learning?
2. Is the effectiveness of peer interaction influenced by features of the study design (assessment pattern, posttest delay, comparison group, frequency of sessions, and instructions for consensus)?
3. Is the effectiveness of peer interaction influenced by interactional dynamics (age, gender, and group size)?
4. Does the effectiveness of peer interaction differ according to the area of learning?

# Method

## Study Sample

Our meta-analysis included 62 separate publications that included 71 separate samples. These studies included a total of 7,103 children between 4 and 18 years of age. Sample sizes ranged from 26 to 390 children in each study. The majority of studies were conducted in the United States ($n = 32$) and the United Kingdom ($n = 23$). Studies were also conducted in Australia, Canada, Chile, Cyprus, Germany, Israel, the Netherlands, Portugal, Singapore, Sweden, Switzerland, and Russia. Most studies ($n = 49$) provided no information about ethnicity, which may be a consequence of the date of publication. Although some included Asian and Asian American, African American, Latino/a, and Middle-Eastern young people, the majority of the samples that reported information about ethnicity were ethnically homogeneous samples of White European and European American children.

## Literature Search

We identified reports examining peer interaction through a variety of sources. We included reports from as early as we could find up to and including 2017. We did not have any geographical restrictions. All articles were published in English. The earliest article identified was Hudgins (1960). First, we looked for articles, book chapters, and dissertations using computerized searches of *PsycINFO*, *ERIC*, and *Dissertations Abstracts International*. Literature searches were conducted using different keywords (i.e., peer, peer learning, collaborative learning, cooperative learning, peer interaction, and peer conversation) alone and in combination with other words (i.e., children, adolescents, and young people), which yielded 11,570 results. Second, we compiled all the citations from Davis and Winstone (2017) and Sills et al. (2016), with these sources representing relevant narrative syntheses on this topic. Third, we then examined the profile pages of authors of each article included to see if they had conducted other studies that were not included in our database. Fourth, studies were identified from citations in articles and from forward citations in databases. Finally, we examined the conference schedules from the American Education Research Association, the European Association for Developmental Psychology, and the Society for Research in Child Development. The abstracts were scanned for relevance, which left 210 articles in the database. After using additional criteria (see

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

below), 62 articles with 71 separate studies were retained. At least two authors together checked each article before exclusion and a reason was noted on a spreadsheet shared between all authors, which was discussed between them. Figure 1 displays the PRISMA diagram of this procedure.

The selection criterion was that studies had to test directly for differences between peer interaction and a comparison group (i.e., control group, individuals working alone, individuals working with adults) in learning in children 18 years or younger.

Exclusion criteria precluded the use of several potentially relevant studies. First, studies that reported preinteraction, individual test scores and then scores after children worked in peers (posttest) were not included if they did not include a comparison group because children could improve on a task over time with no intervention. We excluded 68 studies for not having comparison conditions. Second, the comparison group needed to differ from the peer group only on the peer interaction element. If the peer group engaged in an additional task (e.g., presenting to the class in addition to peer interaction) and the control group did not (e.g., Bunrasi, 2012), we did not include the study because we were unable to isolate the effects of peer interaction from other factors. Because of the focus on peer interaction rather than programs that come from other theoretical traditions, we did not include studies using PALS (Fuchs et al., 2002), cooperative learning (e.g., Nichols & Miller, 1994), or peer tutoring (Leung, 2015). Third, studies that focused on participants whose mean ages were older than 18 years were excluded. Fourth, if the peer group received additional support different from the control group, we excluded the study. Fifth, studies that did not report quantitative data were excluded. Because we did not want to perform a sign test, which simply records whether the data supports or disconfirms a difference (see Dixon & Mood, 1946), we did not include articles that did not

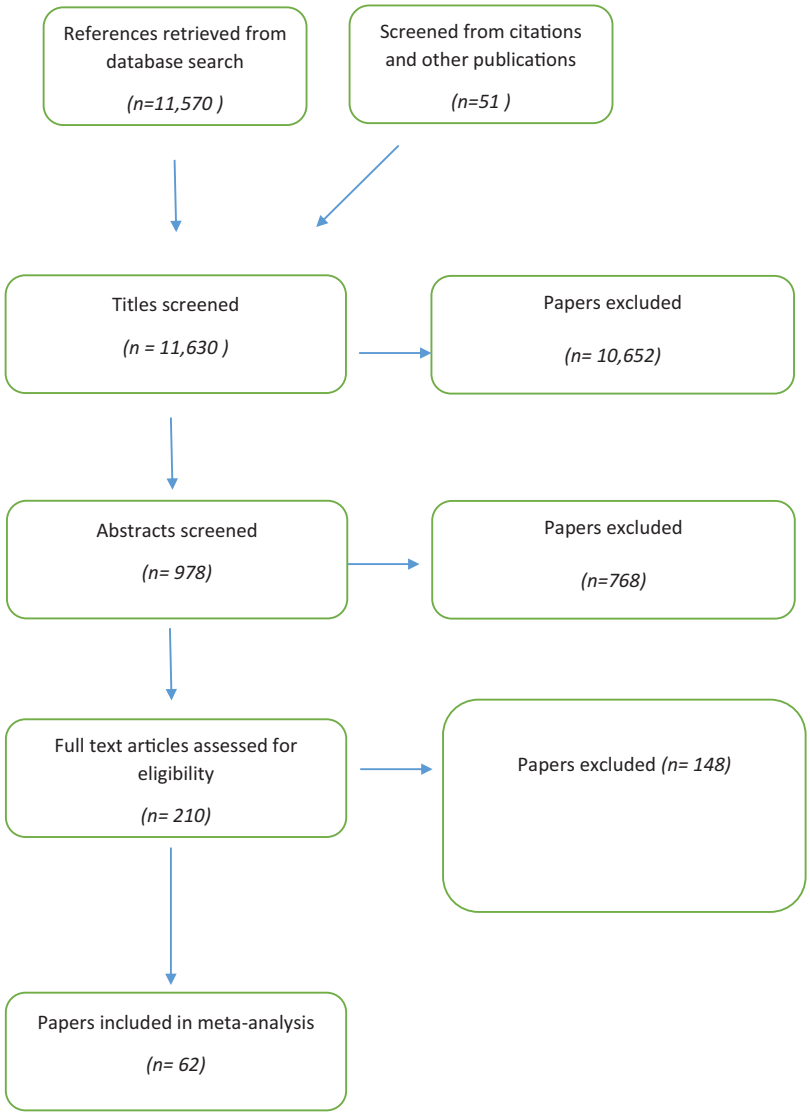



*Figure 1.* PRISMA checklist for systematic review of the literature. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

provide usable statistical information. However, before discarding any articles, authors were contacted for information that could be included in the meta-analysis. We wrote to 16 authors. Six replied with information about nine separate articles. Following recommendations by Borenstein, Hedges, Higgins, and Rothstein (2011), if a study reported no significant difference without further statistical information, data from that study was recorded as $d = 0.000001$ and added to the other comparison. Only two comparisons did not report exact $p$ values or statistics (e.g., $F$ value, means and standard deviations, etc.). Table 1 lists all studies.

## Information Extracted

A coding scheme was created based on variables that could serve as moderators of effects. Table 2 describes these moderators.

To start, we considered features of the design. First, we coded whether studies used a pretest–posttest design or a posttest only assessment pattern as part of the design. Second, we examined when the testing of interest occurred. We recorded whether it occurred while groups worked together, immediately after, after a delay of half an hour or more, or whether no information was reported separately for time. Third, we recorded whether the comparison group consisted of children working alone on the task, children who served as true controls who did not complete the task, or adults working with individual children. For example, Gauvain and Rogoff (1989) compared peer interaction to children working alone as well as children working in an adult–child dyad. We analyzed whether children engaging in peer interaction showed higher learning compared with children working individually with an adult, children working individually on their own on the same task, or children serving as a true control who did not complete the same task. We then examined the effect sizes from these three comparisons to see if they were significantly different from each other. Fourth, we examined over how many sessions peer interaction occurred. There were 62 studies that incorporated one session, one study with two sessions, one study with three sessions, two studies with four sessions, one study with five sessions, two studies with six sessions, one study with eight sessions, and one study with 12 sessions. As a result, we compared whether there was only one session or more than one session. Finally, we recorded whether instructions to students explicitly asked them to reach consensus or not. We included studies in the former category when they instructed students to try to arrive at consensus, but if this was not possible, they needed to understand the other person's argument (e.g., Kuhn, Shaw, & Felton, 1997; Murphy & Messer, 2000; Schwarz & Linchevski, 2007).

Age was the first moderator related to the interactional dynamics we considered. We split samples into whether they included primary-aged students (up to age 10 years) or older (11 to 18 years). Studies were coded as mixed if they spanned both age groups. If an article reported data separately for two different age groups independently (e.g., 6- and 12-year-olds), we counted these effects as separate samples. The youngest age group included in the studies was 4 years. Second, we recorded whether the groups were composed of single-gender, mixed-gender, both mixed- and same-gender groups, or not reported. If an article reported statistics separately for samples by gender of pairs (e.g., Samaha & De Lisi, 2000), we entered that information separately for the moderator analyses. We had originally aimed to look at gender of the participants, but only five studies reported statistics for single-gender samples so this was insufficiently powered. Third, we examined the size of the peer group by comparing studies that assigned children to dyads versus larger groups as the intervention group (e.g., Roazzi & Bryant, 1998).

Finally, we looked at the area of learning under investigation, which included general reasoning tasks that did not involve subject domains, content, or school topics (e.g., puzzle completion, deciding when to cross a road, sorting blocks by color and shape, etc.), conservation (e.g., liquid, mass), creativity (e.g., Torrance tests of creativity), memory (for words), moral reasoning, spatial planning and reasoning (e.g., route planning), and spatial conservation tasks (e.g., three mountains problem; Piaget & Inhelder, 1956). We also created two other categories based on domain knowledge (see Alexander, Jetton, & Kulikowich, 1995, for an explanation of domain vs. subject knowledge) for science and mathematics that included scientific reasoning (e.g., reasoning about topics within the scientific domain that involved scientific content, such as evolutionary theory, buoyancy, and photosynthesis), and mathematical procedures and reasoning (e.g., arithmetic, word problems). Although we had tried to split the mathematical domain into reasoning versus procedural knowledge, studies using word problems rely on arithmetical as well as reasoning knowledge (e.g., Hudgins, 1960) and thus, were coded as mathematical reasoning. Similarly, studies of science learning, such as Chan (2001) tested students' evolutionary knowledge, which relied on content knowledge as well as understanding of processes. Thus, such studies were coded as scientific reasoning. Table 2 lists all moderators.

## Reliability

Coding for moderators was accomplished with recommendations from the four authors who decided on moderator codes. The first two authors separately coded 16 articles for all the moderators except area of study. Area of study and requirement for consensus were coded by the first author with a trained research assistant. Reliability on each of the moderators and statistics across each article was found to be consistently high leading to $\kappa$s ranging from .73 (e.g., on whether the design was pretest posttest or pottest only) to 1.00 (e.g., size of the group, area of study, consensus), which indicates perfect agreement. We also discussed all articles with unclear statistical information. All disagreements were resolved through a discussion of how best to classify the variable in question both within the context of the study and the purposes of analysis. The first author then completed all coding.

## Effect Size Calculation

We used CMA Version 3 (Borenstein et al., 2011) to calculate effect sizes. We used a random effects model because there was variability in the design and location of studies conducted. Effect sizes were coded so that a negative effect size indicates that participants in the compared instructional conditions evidenced greater learning than participants in peer groups, whereas a positive effect size indicates that participants in the peer conditions evidenced greater learning than participants in the compared instructional conditions. The majority of studies reported means and standard deviations for each group.

As the unit of analysis, group samples of studies and comparisons were considered separately. *Studies* as a unit of analysis referred to

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



Table 1
*List of Studies Included in the Meta-Analysis*

| Author(s) | Total *n* | Hedge's *g* [LL, UL] | Assessment pattern | Delay | Same task | Consensus | Frequency | Adult peer | Age | Gender | Size | Area of learning |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ames and Murray (1982) | 48 | .71 [.12/1.31] | Pre | De | Yes | Yes | 1 | I | Y | M | 2 | CN |
| Arterberry, Cain, and Chopko (2007) | 192 | .26 [−.04/.56] | Post | Du | Yes | No | 1 | I | Y | S | 2 | GR |
| Asterhan, Schwarz, and Cohen-Eliyahu (2014) | 390 | .03 [−.23/.29] | Pre | De | Yes | No | 1 | I | O | NR | 2 | GR |
| Bearison, Magzamen, and Filardo (1986) | 99 | −.07 [−.48/.34] | Pre | De | Yes | Yes | 1 | I | Y | S | 2 | SC |
| Blaye, Light, Joiner, and Sheldon (1991) | 38 | .95 [.16/1.74] | Post | De | Yes | Yes | 2 | I | O | S | 2 | SLR |
| Botvin and Murray (1975) | 36 | 3.05 [2.1/4.00] | Post | I | No | Yes | 1 | I | Y | NR | >2 | CN |
| Chan (2001) | 108 | −.04 [−.41/.33] | Pre | I | Yes | No | 1 | I | O | MS | 2 | SR |
| Damon and Killen (1982) | 147 | .21 [−.35/.76] | Pre | De | Some | Yes | 1 | B | Y | M | >2 | ML |
| Diehl and Stroebe (1987) Study 1 | 42 | −1.86 [−3.21/−.51] | Post | I | Yes | No | 1 | I | O | S | >2 | CY |
| Doise and Mugny (1979) | 72 | .42 [−.06/.9] | Pre | De | Yes | Yes | 1 | I | Y | NR | 2 | SC |
| Doise, Mugny, and Perret-Clermont (1975) Study 1 | 60 | 1.13 [.33/1.92] | Post | Du | Yes | Yes | 1 | I | Y | S | 2 | SC |
| Doise, Mugny, and Perret-Clermont (1975) Study 2 | 49 | .71 [.11/1.32] | Pre | De | No | Yes | 1 | I | Y | NR | >2 | CN |
| Druyan (2001) | 338 | −.69 [−1.22/−.14] | Pre | De | Yes | Yes | 1 | B | Y | NR | 2 | SR |
| Duran and Gauvain (1993) | 48 | .54 [−.05/1.13] | Pre | I | Yes | No | 1 | I | Y | S | 2 | SLR |
| Emler and Valiant (1982) Study 1 | 80 | .017 [−.42/.45] | Pre | De | Yes | Yes | 1 | I | Y | NR | 2 | SLR |
| Fawcett and Garton (2005) | 100 | .95 [.50/1.46] | Pre | De | Yes | No | 1 | I | Y | S | 2 | GR |
| Garton and Pratt (2001) | 222 | .02 [−.26/.31] | Pre | De | Yes | No | 1 | I | Y | S | 2 | GR |
| Gauvain and Rogoff (1989) Study 1 | 48 | .07 [−.73/.58] | Post | Du | Yes | No | 1 | I | Y | S | 2 | SLR |
| Gauvain and Rogoff (1989) Study 2 | 18 | −.36 [−.91/.20] | Post | M | Yes | No | 1 | B | Y | S | 2 | SLR |
| Golbeck (1998) | 40 | .41 [−.28/1.10] | Pre | I | Yes | Yes | 1 | I | O | S | 2 | SC |
| Golbeck (1998) | 45 | −.43 [−1.07/.21] | Pre | I | Yes | Yes | 1 | I | Y | S | 2 | SC |
| Gummerum, Leman, and Hollins (2013) | 312 | −.34 [−.97/.29] | Post | Du | Yes | No | 1 | I | Y | S | >2 | MY |
| Howe, McWilliam, and Cross (2005) Study 2 | 138 | .36 [.023/.69] | Pre | De | some | Yes | 1 | I | M | NR | >2 | SR |
| Howe, Devine, & Taylor Tavares (2013) | 139 | .37 [−.27/.39] | Pre | De | some | No | 1 | I | M | NR | >2 | SR |
| Howe, McWilliam, and Cross (2005) Study 3 | 143 | .21 [−.12/.53] | Pre | De | Yes | Yes | 1 | B | Y | NR | 2 | SR |
| Howe, Taylor Tavares, and Devine (2016) | 168 | .06 [−.45/.48] | Pre | De | some | No | 1 | B | Y | NR | 2 | SR |
| Hudgins (1960) | 128 | .28 [−.07/.63] | Post | De | Yes | Yes | 1 | I | M | NR | >2 | MR |
| Kirschner, Paas, and Kirschner (2009) | 70 | .43 [−.38/1.25] | Post | NR | Yes | No | 1 | I | O | NR | >2 | SR |
| Kruger (1992) | 72 | .6 [.11/1.10] | Pre | I | Yes | Yes | 1 | A | Y | S | 2 | ML |
| Kuhn, Shaw, and Felton (1997) | 49 | .62 [−.13/1.37] | Pre | NR | No | Yes | 5 | I | O | S | 2 | GR |
| Leman and Oldham (2005) | 96 | −.35 [−.75/.047] | Post | Du | Yes | No | 1 | I | Y | S | 2 | MY |
| Leman, Skipper, Watling, and Rutland (2016) | 324 | .38 [.16/.60] | Pre | De | No | No | 1 | I | Y | S | 2 | SR |
| Light and Glachan (1985) Study 1 | 32 | .90 [.21/1.59] | Pre | De | Yes | Yes | 1 | I | O | NR | 2 | GR |
| Light and Glachan (1985) Study 1 | 32 | 1.27 [.50/2.04] | Pre | De | Yes | Yes | 1 | I | Y | NR | 2 | GR |
| Littleton, Light, Joiner, Messer, and Barnes (1992) | 120 | .46 [.00/.83] | Pre | M | Yes | No | 1 | I | O | MS | 2 | SLR |
| Lumpe and Staver (1995) | 52 | 1.61 [.88/2.34] | Pre | M | Yes | Yes | 4 | I | O | NR | >2 | SR |
| McCall (2017) | 66 | 2.46 [1.81/3.12] | Post | Du | Yes | Yes | 12 | I | O | NR | >2 | SR |
| Maitland and Goldman (1974) | 36 | −.01 [−.67/.65] | Pre | De | Yes | Both | 1 | I | O | NR | >2 | ML |
| Manion and Alexander (1997) | 90 | .34 [−.07/.75] | Pre | I | Yes | Yes | 1 | I | Y | NR | >2 | MY |
| Messer, Joiner, Loveridge, Light, and Littleton (1993) Study 1 | 40 | −.33 [−1.07/.42] | Pre | I | Yes | No | 1 | I | Y | S | 2 | SR |
| Meudell, Hitch, and Boyle (1995) | 96 | 1.34 [.81/1.87] | Pre | Du | Yes | Yes | 1 | I | O | NR | 2 | MY |
| Murphy and Messer (2000) | 122 | .39 [−.82/.03] | Pre | I | Yes | Yes | 1 | B | Y | NR | 2 | SR |
| Perlmutter, Behrend, Kuo, and Muller (1989) Study 3 | 42 | .16 [−.49/.81] | Post | De | Yes | No | 6 | I | Y | MS | 2 | SR |
| Phelps and Damon (1989) | 110 | .52 [.12/.93] | Pre | NR | Yes | No | 6 | I | Y | S | 2 | MR |
| Psaltis and Duveen (2007) | 213 | .50 [.22/.78] | Pre | De | No | Yes | 1 | I | Y | MS | 2 | CN |
| Psaltis (2011) | 266 | .39 [−.04/.81] | Pre | M | No | Yes | 1 | I | Y | MS | 2 | SC |
| Radziszewska and Rogoff (1988) | 32 | −.66 [−1.35/.039] | Post | I | Yes | No | 1 | A | Y | S | 2 | SLR |
| Radziszewska and Rogoff (1991) | 40 | −1.18 [−1.84/−.52] | Post | I | Yes | No | 1 | A | Y | S | 2 | SLR |
| Roazzi and Bryant (1998) | 78 | 1.55 [1.04/2.06] | Pre | De | Yes | No | 1 | I | Y | NR | >2 | SR |
| Russell (1981) | 161 | .30 [−.08/.67] | Pre | Du | Yes | Yes | 1 | I | Y | S | 2 | CN |
| Russell, Mills, and Reiff-Musgrove (1990) Study 1 | 99 | .93 [.29/1.58] | Pre | Du | Yes | No | 1 | I | Y | S | 2 | CN |

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



Table 1 (*continued*)

| Author(s) | Total *n* | Hedge's *g* [LL, UL] | Assessment pattern | Delay | Same task | Consensus | Frequency | Adult peer | Age | Gender | Size | Area of learning |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Russell et al. (1990) Study 2 | 64 | −.06 [−.79/.66] | Pre | Du | Yes | No | 1 | I | Y | S | 2 | CN |
| Russell et al. (1990) Study 3 | 64 | 2.20 [.64/3.77] | Pre | Du | Yes | No | 1 | I | Y | S | 2 | CN |
| Samaha and De Lisi (2000) | 88 | .67 [−.92/2.57] | Pre | M | Yes | Yes | 1 | I | O | both | >2 | GR |
| Schwartz (1995) Study 1 | 28 | 2.67 [.99/4.35] | Post | Du | Yes | Yes | 1 | I | O | S | 2 | SR |
| Schwartz (1995) Study 2 | 91 | .73 [.19/1.27] | Post | Du | Yes | Yes | 1 | I | O | NR | 2 | SR |
| Schwartz (1995) Study 3 | 40 | 1.43 [.50/2.35] | Post | Du | Yes | Yes | 1 | I | O | NR | 2 | SR |
| Schwarz and Linchevski (2007). | 60 | .60 [.084/1.11] | Pre | De | No | Yes | 1 | I | O | NR | 2 | MR |
| Stacey (1992) | 80 | −.001 [−.44/.44] | Post | Du | Yes | No | 1 | I | O | NR | >2 | MR |
| Teasley (1995) | 46 | .63 [.13/1.13] | Post | I | Yes | No | 1 | I | Y | S | 2 | SR |
| To (2015) | 112 | .15 [−.22/.52] | Pre | I | Yes | No | 1 | I | O | MS | 2 | SR |
| Tolmie, Thomson, Foot, Whelan, Morrison, and Mclaren (2005) Study 1 | 38 | −.14 [−.71/.43] | Pre | De | some | No | 4 | B | Y | M | >2 | GR |
| Torrance (1970) | 39 | .47 [−.11/1.04] | Post | Du | yes | No | 1 | I | Y | NR | 2 | CY |
| Tudge (1989) | 84 | .08 [−.61/.77] | Pre | De | yes | Yes | 1 | I | Y | NR | 2 | SR |
| Tudge (1992) | 148 | −.10 [−.41/.22] | Pre | De | yes | Yes | 1 | I | Y | S | 2 | SR |
| Tudge and Winterhoff (1993) | 79 | .01 [−.54/.56] | Pre | De | yes | Yes | 3 | I | Y | S | 2 | SR |
| Tudge, Winterhoff, and Hogan (1996) | 140 | .36 [−.094/.82] | Pre | De | yes | Yes | 1 | I | Y | S | 2 | SR |
| Valiant, Glachan, and Emler (1982) | 78 | .93 [.38/1.48] | Pre | De | yes | Yes | 1 | I | Y | NR | 2 | SC |
| Weinstein and Bearison (1985) | 80 | 1.66 [1.17/2.16] | Pre | De | yes | Yes | 1 | I | Y | S | 2 | CN |
| Williams and Tolmie (2000) | 96 | 1.29 [.75/1.83] | Pre | De | yes | Yes | 1 | I | Y | M | >2 | SR |
| Zawilinski (2012) | 75 | .28 [−.37/.91] | Pre | I | no | No | 8 | I | Y | MS | 2 | GR |

*Note*. For assessment pattern, Pre = pretest and posttest, Post = posttest only. For time; De = delay; Du = during; I = immediately; NR = not reported; M = more than one assessment time. In M, analyzed separately when we examined that moderator. For consensus, Both = we analyses separately for that moderator; for age, Y = 4 to 10 years; O = 11 to 18 years, M = spans both age groups. For gender, MS = mixed- and same-gender groups; S = same-gender groups; M = mixed-gender groups; NR = not reported. For adult peer, I = individual children; A = adult; B = both adults and individual children. In B, we computed separately when we examined that moderator. For area, CN = conservation; CY = creativity; G = general reasoning; MR = mathematical reasoning; MY = memory; ML = moral; SCN = spatial conservation; SLR = spatial reasoning; SR = scientific reasoning.

Table 2
*Moderators Included in the Meta-Analysis*

| Moderators | List of categories |
|---|---|
| Assessment pattern | Pretest posttest |
| | Posttest only |
| Time of assessment | During |
| | Immediately after |
| | Delay |
| | No information |
| | Combined |
| Comparison | Individual children working alone |
| | Individual children serving as a control |
| | Individual children working with an adult |
| Frequency | One session |
| | More than one session |
| Consensus | Asked to agree |
| | No instructions about consensus provided |
| Age | Young (up to and including age 10 years) |
| | Old (11 years to 18 years) |
| | Mixed (study includes both old and young children) |
| Gender of pairs | Mixed |
| | Same |
| | Both mixed and same |
| | Unknown |
| Group size | Two |
| | More than two |
| Area of learning | Conservation |
| | Creativity |
| | Mathematical reasoning |
| | Memory |
| | Moral |
| | General reasoning |
| | Scientific reasoning |
| | Spatial conservation |
| | Spatial reasoning |
| Publication characteristic | Published |
| | Unpublished |

individual experiments with different participants. Studies, thus, treat multiple experiments reported within a single article as separate studies if they involved different participants, which might be two age groups of participants reported separately within a single article. Three articles contributed two separate studies and two contributed three separate studies. Two articles reported age groups separately that fit into the ages examined separately in this meta-analysis. *Comparisons* were also used as a unit of analysis in two analyses in a separate dataset than the one used to examine studies. These three analyses (comparison group, consensus, and gender) were of interest to the present meta-analysis. Analysis at the level of comparisons refers to counting each individual statistical comparison as an independent contribution. Although multiple comparisons reported for a single sample violate assumptions of independence, analysis at this level was required to test for effects of these two moderating variables. Six studies had different comparison groups in which children in peer groups were either compared with individuals working alone or individuals working with adults (Druyan, 2001; Gauvain & Rogoff, 1989; Howe et al., 2013; Howe, Taylor Tavares, & Devine, 2016; Murphy & Messer, 2000; Tolmie et al., 2005). The latter group consisted of different children. In this case, we divided the number of children in the intervention condition by the number of comparison groups so that studies with more than one comparison would not weight the analysis disproportionate to its sample size following procedures set out by Borenstein et al. (2011). This data set was also used to examine whether children completed the same task or not because these studies used a control group in which children were tested twice without completing a task. However, two studies were not included in these analyses (Howe et al., 2005 Study 2 and Study 3) because comparisons were not broken down by whether children completed the same task on their own or not. To examine instructions for consensus, Maitland and Goldman (1974) contributed two studies because the instructions were varied with one condition asking participants to arrive at consensus, whereas one did not. In this case, we divided the control group in half for each comparison so that each participant only contributed one data point. For gender analyses, one study used two different gender groups (groups of boys, groups of girls, and mixed-gender groups) with different children in each peer interaction group (Samaha & De Lisi, 2000). In this case, we divided the number of children in the comparison group by three so that they would not have disproportionate weight.

Because some studies reported scores from each individual in the peer condition and others took a mean from the peer group or selected one child from the group to represent the group score, studies using the first strategy would have been weighted unequal to the other types of studies. After discussion with Michael Borenstein, we weighted studies by the number of participants in each study. Thus, if a study used one child's score for a group of children (e.g., Tudge, 1989), we used the mean and standard deviation for this group and counted the actual number of participants rather than the number reported by the authors.

Hedge's $g$ values are reported here as calculated by the CMA program. Hedge's $g$s between 0.20 and 0.50 indicate a small effect size, between 0.50 and 0.80 indicate a medium effect, and greater than 0.80 indicate a large effect (Hedges, 1982). Of course, the effect size alone does not determine statistical significance and we determined the statistical significance of effect sizes based on the $p$ values of the resultant $Z$-scores.

When multiple effect sizes were reported, they were averaged to form a single effect size using the formula recommended by Borenstein et al. (2011):

$$\overline{Y} = \frac{1}{m}\left(\sum_{j}^{m} Yj\right). \quad (1)$$

In this formula, $m$ signifies the number of tests in a study that we combined, $Y$ are the values of the different effect sizes in a study, and $j$ indicates the particular study (see Borenstein et al., 2011, Equation 24.4).

## Results

### Overall Findings

Does research support the effectiveness of peer interaction in facilitating learning in comparison with other types of learning? A total of 71 samples from 62 articles compared peer interaction to other types of learning. Using a random-effects model, studies computed separately for group samples had a mean weighted effect size of Hedges' $g = 0.40$, 95% confidence interval [CI: 0.27, 0.54], $p < .0001$. This effect constitutes a small but meaningful effect size. The positive sign indicates that children assigned to peer interaction conditions evidenced greater learning than those assigned to other conditions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The fail-safe *N*, which is the number of missing studies with a mean Hedges' *g* of 0 that would make the effect size no longer statistically significant, was 2,669 for the data set using studies as the unit of analysis. We also visually inspected the funnel plot (see Figure 2) to look for whether studies cluster around the mean effect size, which would suggest publication bias (Borenstein et al., 2011). We found convergence toward the mean in big studies, which would suggest minimal publication bias (Borenstein et al., 2011). However, there were quite a few outliers. We also conducted correlation coefficients using Spearman's rho in SPSS. Sample size was not normally distributed, which is why we used nonparametric statistics. There was not a statistically significant relation between sample size and Hedges' *g* effect sizes, $r(69) = -.21$, $p = .07$. There was also no effect of decade when used as a moderator, $Q_w(5) = 6.54$, $p = .26$. We also conducted a metaregression using year as covariate, $Q_w(1) = .84$, $p = .84$, $R^2$ analog = 0.001, which indicated that year of publication did not improve the fit of the model.

We also examined impact of publication bias on the size of the effect reported. We included three unpublished studies because research shows that studies with null effects are often not published, which is known as the file-drawer problem (Cook & Therrien, 2017; Polanin, Tanner-Smith, & Hennessy, 2016; Rosenthal, 1979). As a result, a common criticism of meta-analysis is that the approach miscalculates the overall effect size by not including unpublished studies (Esterhuizen & Thabane, 2016). To combat this issue, experts on meta-analysis recommend including unpublished studies (Cook & Therrien, 2017; Esterhuizen & Thabane, 2016; Polanin et al., 2016; Rosenthal, 1979). We compared published (Hedge's $g = 0.38$, 95% CI [.25, .50]) with unpublished contributions (Hedge's $g = 0.95$, 95% CI [−.42, 2.32]). There was no statistically significant difference in effect sizes, $Q(1) = .66$, $p = .42$.

To further assess publication bias, we used Duval and Tweedie's (2000) *Trim and Fill* technique. The *Trim and Fill* technique removes the most extreme small studies from either the negative or positive side of the population mean. Data are trimmed until a symmetrical funnel plot is produced, which computes a new effect size based on the mean of the recomputed symmetrical funnel. At this point, the trimmed studies are filled back into the plot and a mirror image created so that the plot remains symmetrical (Borenstein et al., 2011). If the effect size change is small, once the plot has been trimmed and imagined studies are accounted for, then the impact of publication bias in the sample is considered to be low. When we trimmed the studies from the left of the mean using a random effects model, there was no change in the effect size. The point estimate was 0.40, 95% CI [0.27, 0.54]. When we used a random effects model and trimmed from the right side of the distribution, the imputed point estimate was 0.42, 95% CI [0.29, 0.55]. Together, these indicators suggest minimal evidence of bias from missing studies.

We conducted sensitivity analyses by removing each study and recalculating the statistics. The Hedges' *g*s ranged from 0.37, 95% CI [0.25, 0.50] to 0.42, 95% CI [0.29, 0.55] when we did this. When the study with the largest number of participants was removed (Asterhan et al., 2014), the Hedge's g was 0.41, 95% CI [0.28, 0.55]. In summary, these indicators suggest minimal bias in the findings.

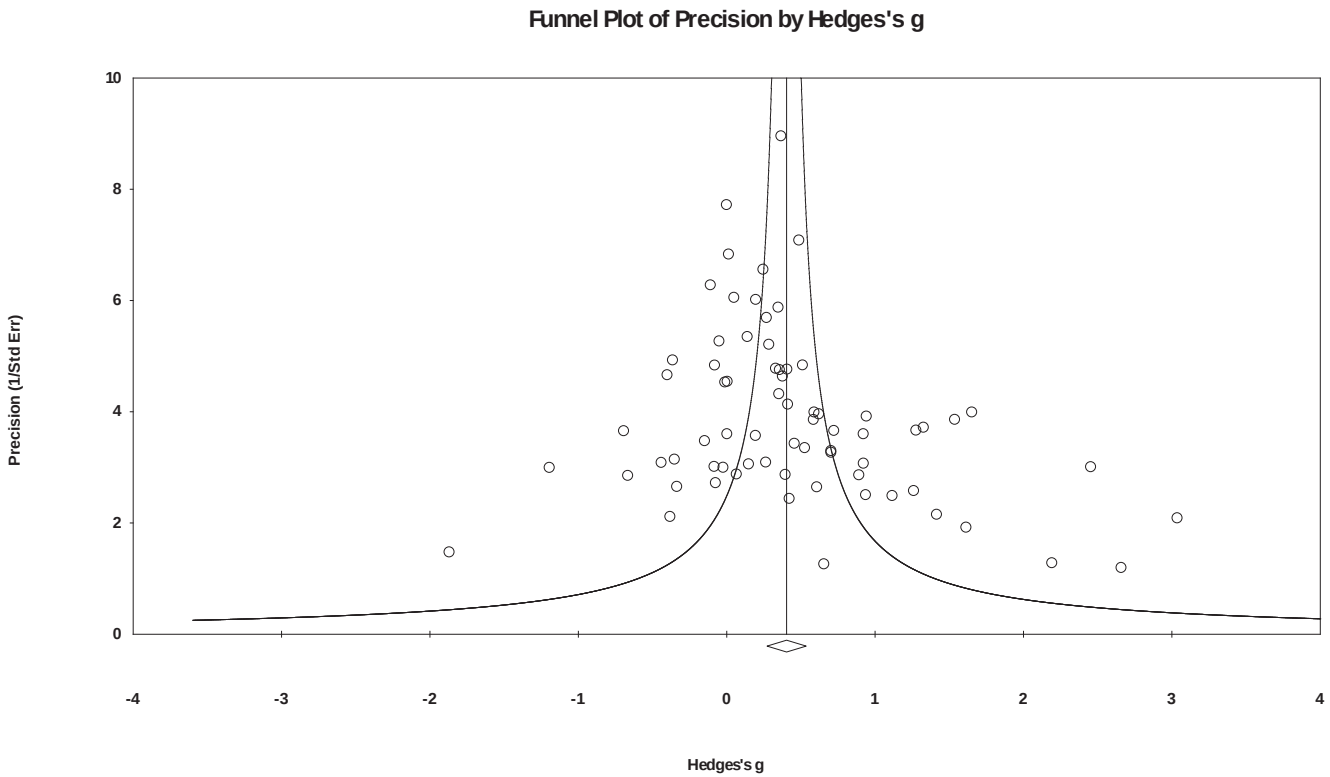



*Figure 2.* Funnel plot.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

## Moderators

Homogeneity analyses for the group sample indicated that the effects were highly heterogeneous, $Q_w$ (70) = 344.23, $p < .0001$. The large value suggests that variability in results may not be because of sampling error alone given the size of the sample (Rosenthal, 1991). For this reason, we were able to investigate different moderators.

An advantage of quantitative meta-analytic techniques is the ability to examine potential moderators of relations with ample statistical power. In the present meta-analyses, the following potential moderators were investigated: features of the design (i.e., assessment pattern, posttest delay, comparison group, frequency of peer sessions, and instructions for consensus), interactional dynamics (i.e., age of participants, gender of pairs, and group size), and the area of learning. Table 3 shows effect sizes, which comparisons were significantly different from each other, and whether the studies in each moderator category demonstrated a statistically significant effect.

**Is the effectiveness of peer interaction influenced by features of the study design?** The first moderators examined aspects of the design. First, we looked at the assessment pattern of the study and found that whether the study used a pretest–posttest or posttest only design did not influence the findings, $Q$ (1) = .02, $p$ = .89, which was in contrast to our hypothesis. Second, in contrast to our hypothesis that delays would increase learning gains in peer interactions, the time of the posttest (during, after, mixed, or with a delay) did not moderate the effect, $Q$ (2) = 4.35, $p$ = .23. Third, we looked at whether the comparison for peer interaction was children working individually on the same task, children serving as a true control who did not complete the same task, or children working individually with adults. Children working in peer groups outperformed children working individually and children serving as true controls. Children working in peer groups did not differ from children working individually with adults. Table 3 shows the effects sizes for each of these three comparisons. When these three effect sizes were contrasted with each other, there was an overall statistically significant effect of comparison type, $Q$ (1) = 12.95, $p < .0001$. We followed up this analysis to examine the effectiveness of peer interaction in comparison with (a) children working individually; (b) true controls; and (c) children working individually with an adult. Because there are three separate comparisons to peer interaction, we ran three follow-up tests to test if each

Table 3
*Moderator Effects*

| Moderator | Hedge's *g* | LL/UL | *Z* | *K* | Heterogeneity *Q* | Comparison by moderator *Q* |
|---|---|---|---|---|---|---|
| Assessment pattern | | | | 2 | | .01 |
| Posttest only | .42 | .06/.78 | 2.26* | 21 | 157.82*** | |
| Pretest posttest | .40 | .27/.53 | 6.17*** | 50 | 195.68*** | |
| Time of posttest | | | | | | 4.35 |
| During | .68 | .30/1.05 | 3.54*** | 16 | 100.78*** | |
| Delay | .40 | .24/.56 | 4.81*** | 31 | 133.32*** | |
| Immediately | .14 | −.21/.48 | .78 | 16 | 90.08*** | |
| Comparison | | | | | | 12.95** |
| Children working alone | .45 | .29/.61 | 5.56*** | 54 | 255.54*** | |
| Children serving as a control group | .51 | .26/.77 | 3.94*** | 12 | 36.76*** | |
| Individual child working with an adult | −.30 | −.70/.10 | −1.48 | 10 | 37.93*** | |
| Number of sessions | | | | | | 1.53 |
| One | .36 | .23/.49 | 5.34*** | 9 | 292.61*** | |
| Multiple | .65 | .18/1.23 | 2.61** | 62 | 54.74*** | |
| Consensus | | | | | | 11.74** |
| Yes | .61 | .42/.81 | 215.63*** | 39 | 6.12*** | |
| No | .17 | .02/.33 | 117.29*** | 33 | 2.14* | |
| Age | | | | | | 2.73 |
| Older age group (11–18 years) | .62 | .32/.91 | 4.08*** | 20 | 109.10*** | |
| Younger age group (4–10 years) | .33 | .18/.49 | 4.18*** | 48 | 240.03*** | |
| Gender of peer groups | | | | | | 1.82 |
| Mixed | .70 | .07/1.34 | 2.17* | 5 | 19.16*** | |
| Same | .24 | .05/.44 | 2.43*** | 33 | 140.82*** | |
| Group size | | | | | | 1.88 |
| More than 2 children | .61 | .25/.97 | 3.29*** | 18 | 128.89*** | |
| Two children | .34 | .20/.47 | 4.93*** | 53 | 207.08*** | |
| Area of learning | | | | | | 11.05 |
| Conservation | .98 | .51/1.47 | 4.04*** | 9 | 52.58*** | |
| Scientific reasoning | .45 | .22/.69 | 3.73*** | 24 | 141.62*** | |
| General reasoning | .39 | .12/.65 | 2.83** | 10 | 26.38** | |
| Spatial conservation | .37 | .01/.72 | 2.04* | 7 | 18.83** | |
| Mathematical reasoning | .34 | .09/.58 | 2.66** | 4 | 4.19 | |
| Moral | .31 | −.14/.67 | 1.73 | 3 | 2.38 | |
| Memory | .25 | −.49/.98 | .65 | 4 | 28.25*** | |
| Spatial reasoning | −.03 | −.47/.41 | −.14 | 8 | 28.88*** | |
| Creativity | −.61 | −2.89/1.66 | −.53 | 2 | 9.60* | |

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

comparison differed from the other two comparisons. As a result, we used a protected alpha of .017 (.05 divided by 3) to conduct the three contrasts. When we contrasted the effect sizes from the comparison groups of children working alone to true controls, there was no difference in effect sizes, $Q$ (1) = .17, $p$ = .68. When we contrasted the effects sizes of children working individually with adults to children working individually as a comparison, there was a statistically significant difference in effect sizes, $Q$ (1) = 11.72, $p$ = .001. When the effect sizes of children working individually with adults as a comparison was contrasted to true controls as a comparison, there was a statistically significant difference, $Q$ (1) = 11.34, $p$ = .001. In summary, there was support for one of our hypotheses, which was that the effect size for the comparison of children working individually with adults would differ from when the comparison was children working individually on a task or serving as a true control who did not complete a task. Fourth, again disconfirming our hypothesis, whether children engaged in peer interaction in one or more sessions did not moderate the findings, $Q$ (1) = 1.53, $p$ = .22. Fifth, confirming our hypothesis, when asked to achieve consensus, children in peer groups made greater gains than when no instructions were provided, $Q$ (1) = 11.74, $p$ = .0001.

**Is the effectiveness of peer interaction influenced by interactional dynamics?** In contrast to our hypothesis, age, $Q$ (1) = 2.73, $p$ = .10, did not moderate the effect. There was also no effect of gender composition when we compared children in mixed versus same-gender pairs, $Q$ (1) = 1.82, $p$ = .18. Finally, children learn the same amount in groups of two children and in larger groups, $Q$ (1) = 1.88, $p$ = .17. Thus, none of our hypotheses related to interactional dynamics were confirmed.

**Does the effectiveness of peer interaction differ according to the area of learning?** Finally, the area of learning did not significantly moderate the findings, $Q$ (8) = 11.05, $p$ = .20.

## Discussion

### The Effectiveness of Peer Interaction

The present study compared findings from 62 papers including 71 separate samples and 7,105 children. The outcomes of this meta-analysis confirm that peer interaction is associated with beneficial learning outcomes compared with other learning conditions, with a small to medium effect size. This finding is consistent with meta-analyses of other types of peer learning (e.g., Johnson, Maruyama, Johnson, Nelson, & Skon, 1981). There is no evidence of publication bias in these results (in terms of sample size or published vs. unpublished reports). However, the conclusion that peer interaction is an effective means of promoting academic advancement is qualified by a large amount of heterogeneity in terms of such effects. In other words, although peer interaction facilitates learning, the conditions and means by which that happens varies greatly and depends on a number of moderating factors. The analysis of the impact of various moderators offers evidence about the source of this variation that has both theoretical and practical implications.

### Interactional Dynamics

None of the findings on interactional dynamics supported our hypotheses. First, we found no evidence that the age of children engaged in peer interaction exerted an influence on its effectiveness. This finding indicates that children of any age may benefit from peer interaction; however, children of different ages may benefit for different reasons. Previous work has suggested that the strategies used for collaboration may differ with age, shifting from a representation of interaction as an opportunity to acquire knowledge that is transmitted from another, to interaction as a meaningful collaboration (Leman, 2015).

With greater age, however, children may engage in social loafing, through which less effort is expended in group than individual activities. Indeed, children younger than third grade (aged 8 to 9 years) are less likely than older students to engage in social loafing (Karau & Williams, 1993). Part of the reason that younger children may engage in less social loafing is because their understanding of others' mental states is less well developed than it is in older individuals (Thompson & Thornton, 2014). Thus, the underlying reasons for why peer interaction supports children at different ages may be complex and represents an important direction for future research.

Gender is another social factor that previous research has examined (e.g., Leman & Björnberg, 2010). Gender influences children's conversations from preschool to school age (Leaper & Smith, 2004; Tenenbaum, Ford, & Alkhedairy, 2011), but the present analysis indicates that the gender composition of groups does not feed through to affect learning outcomes. Although gendered conversation dynamics may be very much a feature of children's interactions, children can work around these in solving a problem and learning from the interaction (see Leman et al., 2005).

The third social factor we examined was group size. Previous meta-analyses identified that smaller group size was associated with higher effectiveness of cooperative peer learning (Johnson et al., 1981). The present analysis found no evidence for group size effects on the effectiveness of peer interaction. The lack of an effect has important implications for classroom contexts where children are often assigned to work in groups of varying sizes for pragmatic reasons. In summary, findings for the three interactional dynamics are promising for practice in that across each group, gender pairing, and group size, peer interaction is more effective than the comparison conditions.

### Area of Learning

In the present meta-analysis, we included studies using a wide variety of tasks from many areas of learning. There was no statistically significant effect of this moderator. Part of the reason may be the small number of studies in some areas, which may mask the effect. One limitation in meta-analysis is the reliance on previous literature. However, we note that there were large differences across these areas. For example, whereas peer interaction was not found to be effective for Creativity, there was a large effect for Conservation. This large effect for conservation tasks is consistent with a significant strand of work in European psychology (e.g., Doise et al., 1975) that sought to unpack the learning benefits of social interaction between peers. These tasks rely on a set of fundamental cognitive processes that can readily trigger sociocognitive conflict. Typically, conservation tasks set up disequilibrium by pairing a conserver with a nonconserver (e.g., Botvin & Murray, 1975; Weinstein & Bearison, 1985), and many

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Piagetian and neo-Piagetian tasks rely on qualitative shifts in understanding. These conditions provide scope for alternative representational systems (understanding) to come into conflict in discussion and debate, which can be a catalyst for learning. This sociocognitive conflict may not be a component of other areas of learning. We return to this explanation when we discuss instructions for consensus.

### Study Design

Aspects of study design including whether the study used a pretest–posttest or merely posttest design, the timing of the posttest (specifically, whether there was a significant delay after interaction), and the frequency of sessions did not moderate the effect of peer interaction.

In contrast, one element of study design which significantly moderated the effectiveness of peer interaction was the requirement for peers to reach consensus in putting forward a common answer to a task or problem. In tasks where peers were not explicitly instructed to reach consensus, the effect size was small. In contrast, when consensus was part of the task instructions, the effect size was of medium magnitude. This finding is in accord with arguments that cognitive conflict is necessary but not sufficient as a catalyst for learning. The key element seems to be the verbal interactions that facilitate negotiation of diverse viewpoints, testing ideas, and seeing things from a different point of view (Forman & Kraker, 1985). In this sense, the cognitive processes underpinning consensus generate social (interpersonal) disequilibrium (Damon & Killen, 1982). However, previous research suggests that consensus alone is not enough to generate learning gains but requires also some further appraisal of consensual positions (Howe & Tolmie, 2003). In this respect, consensus that is achieved as a social process (e.g., through processes of conformity) is less adequate from a perspective of learning and development than consensus achieved through sociocognitive processes. Involvement in a discussion requiring consensus also facilitates children's emotional and intellectual engagement and investment in the dialogue. Passivity is difficult in a task requiring consensus. Furthermore, social loafing is less likely when consensus is required; the higher cognitive demands have the potential to ensure that social dynamics in the interaction do not lead to domination of discussion and decision-making.

Supporting the role of study design, there was a moderating effect of whether peer interaction was compared against children working individually, true controls, or a child working individually with an adult. Peer interaction only had positive effects when peer interactions were compared with children working individually or a true control. In contrast, peer interaction was not more effective than a child working with an adult. This finding supports the Vygotskian perspective that adult–child dyads can support children's learning through careful scaffolding of children to help them reconstruct knowledge (Vygotsky, 1978; Wood & Middleton, 1975).

### Implications for Theory

The moderator analyses of study design support theoretical approaches that have proposed an important role in learning for social interaction (e.g., Piaget, 1932; Vygotsky, 1978), and provide further evidence that conversation and language play a primary role in scaffolding children's learning as others have argued (de Rosnay & Hughes, 2006; Tenenbaum, Leman, Aznar, & To, 2016). The processes through which children learn from peers versus adults might differ, but the findings suggest that children need to be actively engaged in conversation and indicate the central role of language in children's learning.

More specifically, our key finding that peer interaction is more effective when children are instructed to reach consensus points to the importance of intersubjectivity, sociocognitive conflict, and restoration of equilibrium as key cognitive processes underpinning the learning gains from peer interaction. Negotiating differences of opinion in the process of reaching consensus creates disequilibrium, requiring children to "try out new ideas" (Palincsar, 1998, p. 350). Similarly, Piaget (1985) argued that "disequilibrium forces the subject to go beyond his current state and strike out in new directions" (p. 10). Vygotsky (1978) also proposed that language can drive developmental processes, where language as a cultural tool facilitates the coconstruction of knowledge (Vygotsky, 1981). It appears that scaffolding children's interaction through the instruction to work toward consensus is a simple element of task design that can have powerful effects. While Piaget proposed that children working with adults is less effective because the degree of intellectual asymmetry is too great, our findings indicate that working one-on-one with adults can be effective in facilitating children's learning. In summary, this finding suggests that both Piaget and Vygotsky are correct in suggesting that children can learn from either adults or children as long as they are active members of the conversation.

### Implications for Practice

Of course there is a key pragmatic difference between child–peer and child–adult learning partnerships. Many classrooms have just a few adults to work with a large group of children; the process of children working one-on-one with adults is simply not practical. As a result, children are often asked to work simultaneously and individually on a particular task, which is less effective than peer interaction. That peers can also be effective teachers opens up the possibility for simultaneous learning conversations across an entire classroom of children, multiplying the number of potential instructors in a typical classroom.

Our analysis supports that peer interaction is important for pedagogic and not just practical reasons. Peer interaction appears to be effective across different areas of the curriculum, and in both elementary and secondary education. A key educational implication of our findings is that a simple amendment to task instructions, whereby children working in dyads or groups are instructed to reach consensus through interaction, can lead to greater gains in learning. These gains may also extend beyond the focal task. Howe (2015) proposed that the process of exploring and negotiating differences of opinion has the potential to develop metacognitive awareness, which can facilitate future learning in similar situations. Additionally, while peer interaction is associated with effective learning, peer interaction entails many positive social and personal outcomes as well for children including the development of relationships, positive intergroup attitudes, and the rehearsal of communication and collaboration skills (Leman & Lam, 2008; Leman et al., 2016). Working with others toward consensus is

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

likely to further facilitate perspective-taking and shared responsibility for task outcomes, which pose further social benefits to children.

## Limitations

One of the most common limitations in meta-analysis is the amount of information included in the primary literature. As a result of lack of information in many studies we could not code for the duration of activities so we could not investigate this factor. Nor could we examine whether groups composed of girls or boys made greater gains because there were few studies reporting these effects separately (e.g., Tudge et al., 1996). Few studies explicitly stated characteristics of individual children, such as whether they were at risk or had developmental disabilities. Given the lack of information, we would assume that most studies relied on populations of children who were not at risk. Another limitation is that studies differed from each other in many ways, which is known as the apple-and-oranges problem in meta-analysis (Borenstein et al., 2011). All of these limitations suggest arenas for future primary research.

## Future Research Directions

For researchers, the findings point to the significance of the social and developmental context for learning, and that children's capabilities for effective social interaction interrelate with their cognitive abilities in often subtle ways. Future research should seek to look more closely at the interplay between social and cognitive factors, as well as considering the influence of broader cultural norms and practices in peer interaction processes (see, again, Thanh, Gillies, & Renshaw, 2008). Finally, future work should seek to uncover our suggestion that reaching consensus may play a central role in effective peer interaction. Untangling the mechanisms underlying successful peer interaction can contribute to increased learning gains and more effective classroom pedagogy.

## Conclusion

While many pedagogic designs require children to work in dyads or groups for practical reasons, key developmental theories point to the potential benefits of this practice for learning on both cognitive and social levels. Palincsar (1998) argued that "verbal interaction is the key to co-construction and cognitive change" (p. 351). As a cultural tool, language can facilitate learning, enabling children to coconstruct knowledge with an interlocutor (Vygotsky, 1981). Furthermore, the act of working with a peer in the context of a learning task has the potential to create disequilibrium through sociocognitive conflict, with the process of reaching consensus creating the conditions for cognitive growth (Piaget, 1985). Our findings indicate that the benefits of peer interaction can be realized by educators if they create opportunities not just for discussion, but also for the negotiation of a shared understanding.

## References


Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology, 87,* 559–575. http://dx.doi.org/10.1037/0022-0663.87.4.559

Almasi, J. F. (1995). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussions of literature. *Reading Research Quarterly, 30,* 314–351. http://dx.doi.org/10.2307/747620

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology, 18,* 894–897. http://dx.doi.org/10.1037/0012-1649.18.6.894

Arterberry, M. E., Cain, K. M., & Chopko, S. A. (2007). Collaborative Problem Solving in Five-Year-Old Children: Evidence of social facilitation and social loafing. *Educational Psychology, 27,* 577–596, http://dx.doi.org/10.1080/01443410701308755

Asterhan, C. S., Schwarz, B. B., & Cohen-Eliyahu, N. (2014). Outcome feedback during collaborative learning: Contingencies between feedback and dyad composition. *Learning and Instruction, 34,* 1–10. http://dx.doi.org/10.1016/j.learninstruc.2014.07.003

Baines, E., Blatchford, P., & Chowne, A. (2007). Improving the effectiveness of collaborative group work in primary schools: Effects on science attainment. *British Educational Research Journal, 33,* 663–680. http://dx.doi.org/10.1080/01411920701582231

Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *The Journal of General Psychology, 137,* 256–272. http://dx.doi.org/10.1080/00221309.2010.484448

Bearison, D. J., Magzamen, S., & Filardo, E. K. (1986). Socio-cognitive conflict and cognitive growth in young children. *Merrill-Palmer Quarterly, 32,* 51–72. http://dx.doi.org/10.1353/mpq.2005.0016

Blaye, A., Light, P., Joiner, R., & Sheldon, S. (1991). Collaboration as a facilitator of planning and problem solving on a computer-based task. *British Journal of Developmental Psychology, 9,* 471–483. http://dx.doi.org/10.1111/j.2044-835X.1991.tb00890.x

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York, NY: David McKay Company.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* Chichester: John Wiley & Sons Ltd.

Botvin, G. J., & Murray, F. B. (1975). The efficacy of peer modeling and social conflict in the acquisition of conservation. *Child Development, 46,* 796–799. Retrieved from https://www.jstor.org/stable/1128583

Bunrasi, J. B. T. (2012). *Algebra I achievement of eighth grade Mexican American students using cooperative learning versus traditional instruction* (Doctoral dissertation). Walden University.

Castellaro, M. A., & Roselli, N. D. (2015). Peer collaboration in childhood according to age, socioeconomic context and task. *European Journal of Psychology of Education, 30,* 63–80. http://dx.doi.org/10.1007/s10212-014-0228-3

Chan, C. K. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science, 29,* 443–479.

Cook, B. G., & Therrien, W. J. (2017). Null effects and publication bias in special education research. *Behavioral Disorders, 42,* 149–158. http://dx.doi.org/10.1177/0198742917709473

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly, 28,* 347–367. Retrieved from http://www.jstor.org/stable/23086121

Davis, A., & Winstone, N. (2017). Educational implications. In A. Slater & G. Bremner (Eds.), *An introduction to developmental psychology* (3rd ed., pp. 613–644). Chichester: Wiley.

de Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology, 24,* 7–37. http://dx.doi.org/10.1348/026151005X82901

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and*


This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

*Social Psychology, 53,* 497–509. http://dx.doi.org/10.1037/0022-3514.53.3.497

Dimant, R. J., & Bearison, D. J. (1991). Development of formal reasoning during successive peer interactions. *Developmental Psychology, 27,* 277–284. http://dx.doi.org/10.1037/0012-1649.27.2.277

Dixon, W. J., & Mood, A. M. (1946). The statistical sign test. *Journal of the American Statistical Association, 41,* 557–566. Retrieved from https://www.jstor.org/stable/2280577

Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centrations in cognitive development. *European Journal of Social Psychology, 9,* 105–108, http://dx.doi.org/10.1002/ejsp.2420090110

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon Press.

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5,* 367–383. http://dx.doi.org/10.1002/ejsp.2420050309

Druyan, S. (2001). A comparison of four types of cognitive conflict and their effect on cognitive development. *International Journal of Behavioral Development, 25,* 226–236. http://dx.doi.org/10.1080/01650250042000276

Duran, R. T., & Gauvain, M. (1993). The role of age versus expertise in peer collaboration during joint planning. *Journal of Experimental Child Psychology, 55,* 227–242. http://dx.doi.org/10.1006/jecp.1993.1013

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Emler, N., & Valiant, G. L. (1982). Social interaction and cognitive conflict in the development of spatial coordination skills. *British Journal of Psychology, 73,* 295–303. http://dx.doi.org/10.1111/j.2044-8295.1982.tb01812.x

Esterhuizen, T. M., & Thabane, L. (2016). Con: Meta-analysis: Some key limitations and potential solutions. *Nephrology, Dialysis, Transplantation, 31,* 882–885. http://dx.doi.org/10.1093/ndt/gfw092

Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology, 75,* 157–169. http://dx.doi.org/10.1348/000709904X23411

Forman, E. A., & Kraker, M. J. (1985). The social origins of logic: The contributions of Piaget and Vygotsky. *New Directions for Child and Adolescent Development, 1985,* 23–39. http://dx.doi.org/10.1002/cd.23219852904

Fuchs, D., Fuchs, L. S., & Burish, P. (2000). Peer-assisted learning strategies: An evidence-based practice to promote reading achievement. *Learning Disabilities Research & Practice, 15,* 85–91. http://dx.doi.org/10.1207/SLDRP1502_4

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Martinez, E. A. (2002). Preliminary evidence on the social standing of students with learning disabilities in PALS and no–PALS classrooms. *Learning Disabilities Research & Practice, 17,* 205–215. http://dx.doi.org/10.1111/1540-5826.00046

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal, 99,* 201–219. Retrieved from https://www.jstor.org/stable/1002313

Fuchs, L. S., Fuchs, D., Kazdan, S., Karns, K., Calhoon, M. B., Hamlett, C. L., & Hewlett, S. (2000). Effects of workgroup structure and size on student productivity during collaborative work on complex tasks. *The Elementary School Journal, 100,* 183–212. Retrieved from http://www.jstor.org/stable/1002151

Garton, A. F., & Pratt, C. (2001). Peer assistance in children's problem solving. *British Journal of Developmental Psychology, 19,* 307–318. http://dx.doi.org/10.1348/026151001166092

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25,* 139–151. http://dx.doi.org/10.1037/0012-1649.25.1.139

Golbeck, S. L. (1998). Peer collaboration and children's representation of the horizontal surface of liquid. *Journal of Applied Developmental Psychology, 19,* 571–592. http://dx.doi.org/10.1016/S0193-3973(99)80056-2

Golbeck, S. L., & Sinagra, K. (2000). Effects of gender and collaboration on college students' performance on a Piagetian Spatial Task. *The Journal of Experimental Education, 69,* 22–35, http://dx.doi.org/10.1080/00220970009600647

Greenwood, C. R., Delquadri, J., & Carta, J. J. (1988). *Classwide peer tutoring*. Seattle, WA: Educational Achievement Systems.

Gummerum, M., Leman, P. J., & Hollins, T. S. (2013). Children's collaborative recall of shared and unshared information. *British Journal of Developmental Psychology, 31,* 302–317. http://dx.doi.org/10.1111/bjdp.12006

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92,* 490–499. http://dx.doi.org/10.1037/0033-2909.92.2.490

Howe, C. (2015). *Conceptual structure in childhood and adolescence: The case of everyday physics*. London: Routledge. http://dx.doi.org/10.4324/9781315626772

Howe, C., Devine, A., & Taylor Tavares, J. (2013). Supporting conceptual change in school science: A possible role for tacit understanding. *International Journal of Science Education, 35,* 864–883. http://dx.doi.org/10.1080/09500693.2011.585353

Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology, 96,* 67–93. http://dx.doi.org/10.1348/000712604X15527

Howe, C., Taylor Tavares, J., & Devine, A. (2016). Recognition as support for reasoning about horizontal motion: A further resource for school science? *Research in Science & Technological Education, 34,* 273–289. http://dx.doi.org/10.1080/02635143.2016.1168393

Howe, C., & Tolmie, A. (2003). Group work in primary school science: Discussions, consensus and guidance from experts. *International Journal of Educational Research, 39,* 51–72. http://dx.doi.org/10.1016/S0883-0355(03)00073-9

Hudgins, B. B. (1960). Effects of group experience on individual problem solving. *Journal of Educational Psychology, 51,* 37–42. http://dx.doi.org/10.1037/h0038574

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89,* 47–62. http://dx.doi.org/10.1037/0033-2909.89.1.47

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681–706. http://dx.doi.org/10.1037/0022-3514.65.4.681

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review, 21,* 31–42. http://dx.doi.org/10.1007/s10648-008-9095-2

Kruger, A. C. (1992). The effect of peer and adult-child transactive discussions on moral reasoning. *Merrill-Palmer Quarterly, 38,* 191–211.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction, 15,* 287–315. http://dx.doi.org/10.1207/s1532690xci1503_1

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology, 40,* 993–1027. http://dx.doi.org/10.1037/0012-1649.40.6.993

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Leman, P. J. (2015). How do groups work? Age differences in performance and the social outcomes of peer collaboration. *Cognitive Science, 39,* 804–820. http://dx.doi.org/10.1111/cogs.12172

Leman, P. J., Ahmed, S., & Ozarow, L. (2005). Gender, gender relations, and the social dynamics of children's conversations. *Developmental Psychology, 41,* 64–74. http://dx.doi.org/10.1037/0012-1649.41.1.64

Leman, P. J., & Björnberg, M. (2010). Conversation, development, and gender: A study of changes in children's concepts of punishment. *Child Development, 81,* 958–971. http://dx.doi.org/10.1111/j.1467-8624.2010.01445.x

Leman, P. J., & Lam, V. L. (2008). The influence of race and gender on children's conversations and playmate choices. *Child Development, 79,* 1329–1343. http://dx.doi.org/10.1111/j.1467-8624.2008.01191.x

Leman, P. J., & Oldham, Z. (2005). Do children need to learn to collaborate?: The effect of age and age differences on collaborative recall. *Cognitive Development, 20,* 33–48. http://dx.doi.org/10.1016/j.cogdev.2004.07.002

Leman, P. J., Skipper, Y., Watling, D., & Rutland, A. (2016). Conceptual change in science is facilitated through peer collaboration for boys but not for girls. *Child Development, 87,* 176–183. http://dx.doi.org/10.1111/cdev.12481

Leman, P. J., & Tenenbaum, H. (2017). Communication in children's and adolescents' social groups. In A. Rutland, D. Nesdale, & C. S. Brown (Eds.), *The Wiley-Blackwell handbook of group processes in children and adolescents* (pp. 144–164). Oxford: Wiley-Blackwell. http://dx.doi.org/10.1002/9781118773123.ch7

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology, 107,* 558–579. http://dx.doi.org/10.1037/a0037698

Light, P., & Glachan, M. (1985). Facilitation of individual problem solving through peer interaction. *Educational psychology, 5,* 217–225. http://dx.doi.org/10.1080/0144341850050305

Light, P., & Littleton, K. (1994). Cognitive approaches to group work. In P. Kutnik & C. Rogers (Eds.), *Groups in schools* (pp. 87–103). London: Cassell.

Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1992). Pairing and gender effects on children's computer-based learning. *European Journal of Psychology of Education, 7,* 311–324. http://dx.doi.org/10.1007/BF03172896

Lumpe, A. T., & Staver, J. R. (1995). Peer collaboration and concept development: Learning about photosynthesis. *Journal of Research in Science Teaching, 32,* 71–98. http://dx.doi.org/10.1002/tea.3660320108

McCall, M. O. N. (2017). *The effects of individual versus cooperative testing in a flipped classroom on the academic achievement, motivation toward science, and study time for 9th grade biology students* (Unpublished doctoral dissertation). University of South Alabama.

Maitland, K. A., & Goldman, J. R. (1974). Moral judgment as a function of peer group interaction. *Journal of Personality and Social Psychology, 30,* 699–704. http://dx.doi.org/10.1037/h0037454

Manion, V., & Alexander, J. M. (1997). The benefits of peer collaboration on strategy use, metacognitive causal attribution, and recall. *Journal of Experimental Child Psychology, 67,* 268–289. http://dx.doi.org/10.1006/jecp.1997.2409

Messer, D. J., Joiner, R., Loveridge, N., Light, P., & Littleton, K. (1993). Influences on the effectiveness of peer interaction: Children's level of cognitive development and the relative ability of partners. *Social Development, 2,* 279–294. http://dx.doi.org/10.1111/j.1467-9507.1993.tb00018.x

Meudell, P. R., Hitch, G. J., & Boyle, M. M. (1995). Collaboration in recall: Do pairs of people cross-cue each other to produce new memories? *The Quarterly Journal of Experimental Psychology, 48,* 141–152. http://dx.doi.org/10.1080/14640749508401381

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology, 20,* 17–31. http://dx.doi.org/10.1080/014434100110353

Nichols, J. D., & Miller, R. B. (1994). Cooperative learning and student motivation. *Contemporary Educational Psychology, 19,* 167–178. http://dx.doi.org/10.1006/ceps.1994.1015

Nyikos, M., & Hashimoto, R. (1997). Constructivist theory applied to collaborative learning in teacher education: In search of ZPD. *The Modern Language Journal, 81,* 506–517. http://dx.doi.org/10.1111/j.1540-4781.1997.tb05518.x

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49,* 345–375. http://dx.doi.org/10.1146/annurev.psych.49.1.345

Perlmutter, M., Behrend, S. D., Kuo, F., & Muller, A. (1989). Social influences on children's problem solving. *Developmental Psychology, 25,* 744–754. http://dx.doi.org/10.1037//0012-1649.25.5.744

Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology, 81,* 639–646. http://dx.doi.org/10.1037/0022-0663.81.4.639

Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.

Piaget, J. (1959). *The language and thought of the child*. London: Routledge & Kegan Paul.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago, IL: University of Chicago Press.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. New York, NY: Norton.

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86,* 207–236. http://dx.doi.org/10.3102/0034654315582067

Psaltis, C. (2011). The constructive role of gender asymmetry in social interaction: Further evidence. *British Journal of Developmental Psychology, 29,* 305–312. http://dx.doi.org/10.1111/j.2044-835x.2011.02029.x

Psaltis, C., & Duveen, G. (2007). Conservation and conversation types: Forms of recognition and cognitive development. *British Journal of Developmental Psychology, 25,* 79–102. http://dx.doi.org/10.1348/026151005x91415

Psaltis, C., Duveen, G., & Perret-Clermont, A. N. (2009). The social and the psychological: Structure and context in intellectual development. *Human Development, 52,* 291–312. http://dx.doi.org/10.1159/000233261

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaborators on children's planning skills. *Developmental Psychology, 24,* 840–848. http://dx.doi.org/10.1037/0012-1649.24.6.840

Radziszewska, B., & Rogoff, B. (1991). Children's guided participation in planning imaginary errands with skilled adult or peer partners. *Developmental Psychology, 27,* 381–389. http://dx.doi.org/10.1037/0012-1649.27.3.381

Roazzi, A., & Bryant, P. (1998). The effects of symmetrical and asymmetrical social interaction on children's logical inferences. *British Journal of Developmental Psychology, 16,* 175–181. http://dx.doi.org/10.1111/j.2044-835X.1998.tb00917.x

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95,* 240–257. http://dx.doi.org/10.1037/0022-0663.95.2.240

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine, 53,* 247–271. http://dx.doi.org/10.1097/00006842-199105000-00001

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134,* 223–246. http://dx.doi.org/10.1037/0033-2909.134.2.223

Russell, J. (1981). Dyadic interaction in a logical reasoning problem requiring inclusion ability. *Child Development, 52,* 1322–1325. http://dx.doi.org/10.2307/1129524

Russell, J., Mills, I., & Reiff-Musgrove, P. (1990). The role of symmetrical and asymmetrical social conflict in cognitive change. *Journal of Experimental Child Psychology, 49,* 58–78. http://dx.doi.org/10.1016/0022-0965(90)90049-E

Samaha, N. V., & De Lisi, R. (2000). Peer collaboration on a nonverbal reasoning task by urban, minority students. *The Journal of Experimental Education, 69,* 5–21. http://dx.doi.org/10.1080/00220970009600646

Schwarz, B. B., & Linchevski, L. (2007). The role of task design and argumentation in cognitive development during peer interaction: The case of proportional reasoning. *Learning and Instruction, 17,* 510–531. http://dx.doi.org/10.1016/j.learninstruc.2007.09.009

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The journal of the Learning Sciences, 4,* 321–354. http://dx.doi.org/10.1207/s15327809jls0403_3

Sills, J., Rowse, G., & Emerson, L. M. (2016). The role of collaboration in the cognitive development of young children: A systematic review. *Child: Care, Health and Development, 42,* 313–324. http://dx.doi.org/10.1111/cch.12330

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145–173). New York, NY: Cambridge University Press.

Stacey, K. (1992). Mathematical problem solving in groups: Are two heads better than one. *Journal of Mathematical Behavior, 11,* 261–275.

Strough, J., Berg, C. A., & Meegan, S. P. (2001). Friendship and gender differences in task and social interpretations of peer collaborative problem solving. *Social Development, 10,* 1–22. http://dx.doi.org/10.1111/1467-9507.00145

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31,* 207–220. http://dx.doi.org/10.1037/0012-1649.31.2.207

Tenenbaum, H. R., Ford, S., & Alkhedairy, B. (2011). Telling stories: Gender differences in peers' emotion talk and communication style. *British Journal of Developmental Psychology, 29,* 707–721. http://dx.doi.org/10.1348/2044-835X.002003

Tenenbaum, H. R., Leman, P. J., Aznar, A., & To, C. (2016). Studying children's conversation. In J. Van Herwegen & J. Prior (Eds.), *Developmental methods* (pp. 89–106). Oxford: Psychology Press.

Thanh, P. T. H., Gillies, R., & Renshaw, P. (2008). Cooperative learning (CL) and academic achievement of Asian students: A true story. *International Education Studies, 1,* 82–89. http://dx.doi.org/10.5539/ies.v1n3p82

Thompson, R. B., & Thornton, B. (2014). Gender and theory of mind in preschoolers' group effort: Evidence for timing differences behind children's earliest social loafing. *The Journal of Social Psychology, 154,* 475–479. http://dx.doi.org/10.1080/00224545.2014.933763

To, C. (2015). *Developmental changes in British children's reasoning about evolution* (Unpublished doctoral dissertation). Guildford, Surrey: University of Surrey.

Tolmie, A., Thomson, J. A., Foot, H. C., Whelan, K., Morrison, S., & McLaren, B. (2005). The effects of adult guidance and peer discussion on the development of children's representations: Evidence from the training of pedestrian skills. *British Journal of Psychology, 96,* 181–204. http://dx.doi.org/10.1348/000712604X15545

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. London: Routledge. http://dx.doi.org/10.4324/9781315695471

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25,* 631–645. http://dx.doi.org/10.1080/01443410500345172

Torrance, E. P. (1970). Influence of dyadic interaction on creative functioning. *Psychological Reports, 26,* 391–394. http://dx.doi.org/10.2466/pr0.1970.26.2.391

Tudge, J. (1989). When collaboration leads to regression: Some negative consequences of socio-cognitive conflict. *European Journal of Social Psychology, 19,* 123–138. http://dx.doi.org/10.1002/ejsp.2420190204

Tudge, J. R. (1992). Processes and consequences of peer collaboration: A Vygotskian analysis. *Child Development, 63,* 1364–1379. http://dx.doi.org/10.1111/j.1467-8624.1992.tb01701.x

Tudge, J., & Rogoff, B. (1999). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In P. Lloyd & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments* (pp. 32–56). London: Routledge.

Tudge, J., & Winterhoff, P. (1993). Can young children benefit from collaborative problem solving? Tracing the effects of partner competence and feedback. *Social Development, 2,* 242–259. http://dx.doi.org/10.1111/j.1467-9507.1993.tb00016.x

Tudge, J. R., Winterhoff, P. A., & Hogan, D. M. (1996). The cognitive consequences of collaborative problem solving with and without feedback. *Child Development, 67,* 2892–2909. Retrieved from http://www.jstor.org/stable/1131758

Underwood, J., Underwood, G., & Wood, D. (2000). When does gender matter?: Interactions during computer-based problem solving. *Learning and Instruction, 10,* 447–462. http://dx.doi.org/10.1016/S0959-4752(00)00008-6

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press.

Valiant, G., Glachan, M., & Emler, N. (1982). The stimulation of cognitive development through co-operative task performance. *British Journal of Educational Psychology, 52,* 281–288. http://dx.doi.org/10.1111/j.2044-8279.1982.tb02515.x

Vygotsky, L. S. (1981). The instrumental method in psychology. *The concept of activity in Soviet Psychology, 137,* 143.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21–39. http://dx.doi.org/10.1016/0883-0355(89)90014-1

Williams, J. M., & Tolmie, A. (2000). Conceptual change in biology: Group interaction and the understanding of inheritance. *British Journal of Developmental Psychology, 18,* 625–649. http://dx.doi.org/10.1348/026151000165896

Weinstein, B. D., & Bearison, D. J. (1985). Social interaction, social observation, and cognitive development in young children. *European Journal of Social Psychology, 15,* 333–343. http://dx.doi.org/10.1002/ejsp.2420150307

Wood, D., & Middleton, D. (1975). A study of assisted problem-solving. *British Journal of Psychology, 66,* 181–191. http://dx.doi.org/10.1111/j.2044-8295.1975.tb01454.x

Zawilinski, L. M. (2012). *An exploration of a collaborative blogging approach to literacy and learning: A mixed methods study* (Unpublished doctoral dissertation). University of Connecticut.

Received February 19, 2019
Revision received September 21, 2019
Accepted October 18, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.