Educational Research Review 28 (2019) 100284

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014

Heisawn Jeong[a,*], Cindy E. Hmelo-Silver[b], Kihyun Jo[c]

[a] Hallym University, Chuncheon, Republic of Korea
[b] Indiana University, Bloomington, USA
[c] Korea University, Seoul, Republic of Korea

ARTICLE INFO

The current paper is an extension of an earlier work presented at HICSS (Hawaii International Conference on System Science) 2016 in which a preliminary analysis based on a subset of the data was reported.

Keywords:
Computer-supported collaborative learning (CSCL)
meta-Analysis
Collaborative technology
Pedagogy
Domains
Educational levels

ABSTRACT

The goal of this paper is to report on a meta-analysis about the effects of Computer-Supported Collaborative Learning (CSCL) in STEM education. The analysis is based on 316 outcomes from 143 studies that examined the effects of CSCL published between 2005 and 2014. Our analysis showed that the overall effect size of STEM CSCL was 0.51, a *moderate* but notable effect size in educational research. The effect was greatest on process outcomes, followed by knowledge outcomes, and affective outcomes. The sizes of the effects were moderated by types of technology and pedagogy, educational levels of learners, and learning domains. Moderators further interacted so that effects of technology and pedagogy varied depending on the modes of collaboration, learners' educational levels, and domains of learning. The current study demonstrates the overall advantage of CSCL in STEM education and highlights a need to understand how these variables may interact to contribute to overall CSCL effectiveness.

1. Introduction

Computer-Supported Collaborative Learning (CSCL) emerged as a means of learning and instruction that can foster the social nature of learning using a variety of technological and pedagogical strategies (Dillenbourg, Järvelä, & Fischer, 2009; Stahl, Koschmann, & Suthers, 2014). As a research field, CSCL is now more than 20 years old and is coming of age as a mature field. Although there is extensive research regarding the use of CSCL in STEM education (Jeong & Hmelo-Silver, 2012), there have been few systematic assessments on the effects of CSCL. The goal of this research is to examine the effects of CSCL in Science, Technology, Engineering, and Mathematics (STEM) education using a statistical meta-analysis (Cooper, Hedges, & Valentine, 2009; Glass, 1976, 1977). We focus on STEM in particular, as that has been the focus of much CSCL research (Jeong & Hmelo-Silver, 2012).

1.1. Theoretical backgrounds of CSCL

CSCL is built on the premise that collaborative knowledge construction and problem solving can effectively be assisted by technology. This premise was not always obvious to instructors and learning researchers. Early theories emphasized conditioning behaviors and/or strengthening memory traces through repeated associations and practices by individual learners. Researchers soon

* Corresponding author. Department of Psychology, Hallym University, 1 Hallymdaehak-gil, Chuncheon, Gangwon-do, 24252, South Korea.
E-mail address: heis@hallym.ac.kr (H. Jeong).

https://doi.org/10.1016/j.edurev.2019.100284
Received 21 August 2017; Received in revised form 10 February 2019; Accepted 25 April 2019
Available online 14 June 2019

noted that learners' active engagement with learning materials and strategy use is critical to learning success (Chi, 2009; O'Donnell & King, 1999). Researchers also realized the fundamental role of social interaction in learning and development. It was noted that social interactions, even interactions that involve social conflicts, can lead to cognitive reorganization and development (Diose & Mugny, & Perret-Clement, 1975). Vygotsky (1978) argued higher psychological functions emerge first in social interaction, and skillful scaffolding within the child's zone of proximal development is critical for cognitive development. These theoretical advancements eventually led to the proposition that co-construction of shared understanding and intersubjective meaning-making is the key objectives of collaborative learning (Stahl, 2006; Suthers, 2006).

Collaborative learning, an essential element of CSCL, can take a number of forms depending on the pedagogical emphasis and contexts of learning. In many constructivist pedagogical approaches such as problem-based learning and inquiry learning, students solve problems, design artifacts, or engage in inquiry in small groups. Group work and collaboration are encouraged as they help students to engage with the learning materials and develop deep disciplinary understanding (Engle & Conant, 2002; Hmelo-Silver, 2004). Pedagogical approaches and techniques by Johnson and Johnson (2009) and Slavin (1983) focused on how students should be grouped and rewarded, and emphasized conditions to ensure productive modes of collaborative engagements. In responses to the distraction and unruliness that can arise in peer collaboration, O'Donnell and King (1999) proposed structuring group processes to ensure productive interaction. Online collaborative learning presents special challenges in terms of regulating students' behaviors because students are often outside the supervision of teachers and instructors. CSCL scripts focus on the regulation issues that arise in CSCL settings (Kollar, Fischer, & Hesse, 2006).

A diverse variety of tools ranging from discussion boards to simulations, and wikis to robots have been used to support collaborative learning in CSCL research. Learning has always been assisted by tools and technologies, but the recent proliferation of digital technologies makes the role of those tools even more critical. Vygotsky (1978) was one of the first scholars who noted the importance of tools for their capacity to mediate thinking. Modern researchers noted the fundamental roles of external representations and tools in carrying out cognitive functions (Hutchins, 1995; Zhang & Norman, 1994). These realizations helped CSCL researchers make the distinction between learning “with” and “from” computers, and focused the development of tools that can amplify, extend, and enhance what learner groups and communities understand and do in the process of learning and solving problems (Kirschner & Erkens, 2006). It has also been noted that humans and tools form a system of distributed cognition, which constitutes part of the socio-cultural and historical activity systems along with other components such as actors, artifacts, community, and rules that regulate the activities in the system (Arnseth & Ludvigsen, 2006; Greeno, 2006). Researchers now liken learning to the process of participating in the practice of cultural significance in these systems instead of having knowledge or skills. This means that the very nature and notion of learning is closely tied to the contexts within which it takes place and may vary across cultural and historical contexts (Sfard, 1998).

1.2. Assessing the effectiveness of CSCL

The diverse traditions in CSCL research present overlapping and yet different perspectives on what it means to learn collaboratively with the help of computers. Some approaches focus on developing and assessing the role of specific technology or design features of CSCL applications/environments, whereas others are more intent on understanding and evaluating pedagogical models to foster CSCL. Still others are interested in examining the influences of cultural and historical contexts on CSCL (Arnseth & Ludvigsen, 2006). In searching for the effectiveness of CSCL, some researchers focused on learners' cognitive outcomes and processes, while others were more interested in demonstrating how CSCL affects student motivations and attitudes toward subject domains or learning partners (Serrano-Cámara, Paredes-Velasco, Alcover, & Velazquez-Iturbide, 2014). This means that there is a diverse array of empirical CSCL research, each examining the workings of different variables and assessing different outcomes. Integrating evidence from these diverse research practices demands that we use common metrics that can transcend differences in individual studies, while paying attention to the complexity of CSCL practices. Statistical meta-analysis is one of the most widely used synthesis methods that make it possible to combine and compare results from different studies on a single uniform scale of effectiveness (Cooper et al., 2009). In addition to assessing the overall effect of CSCL, it is also possible to evaluate the role of potential moderator variables (Glass, 1976, 1977).

The positive outcomes of collaborative learning are well established by a number of meta-analyses (Kyndt et al., 2013; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003; Springer, Stanne, & Donovan, 1999), but this does not necessarily mean that the outcomes of CSCL would also be positive. Computers and digital technology can provide a number of supports for collaborative learning, but taking advantage of these affordances requires overcoming a number of challenges. In addition to having to deal with technical issues that might arise when using digital tools, technologies may distract learners from learning (Kershner, Mercer, Warwick, & Staarman, 2010; Mateu & Alaman, 2013). Technologies can also demand that learners and teachers change their behaviors and practices. Not everyone is comfortable with posting messages and their work online. Helping them to overcome these difficulties may require a significant amount of demonstration and guidance (Kafai, Fields, & Burke, 2010).

A few existing meta-analyses provide important background to our current work and demonstrate that collaborative learning is effective even when computers are used. Lou, Abrami, and D'Apollonia (2001) carried out a meta-analysis comparing the outcomes of small group learning with technology to outcomes of individual learning with technology. Based on 486 outcomes from 122 individual studies, they reported that small group learning with technology, on average, produced more positive effects than individual learning with technology on students' achievements with effect sizes ranging from .15 to .31. Similar positive effects were also reported in a recent meta-analysis by Chen, Wang, Kirschner, and Tsai (2018). They compared individual learning with computers with collaborative learning with computers in one of the subgroup analyses and reported effect sizes ranging from 0.38 to 0.64. These results demonstrated that learning together is effective compared to learning alone even when learners work with computers.

When it comes to the supports for collaborative learning, we need to note that it can come in many different forms. It may involve providing tools such as online discussion boards, representational tools, and/or simulations as well as supports for pedagogical activities. The nature of these supports are likely to vary depending on whether learners collaborate face-to-face, synchronously or asynchronously. Vogel, Wecker, Kollar, and Fischer (2017) recently examined the effect of CSCL scripts on learning domain-specific knowledge and collaboration skills. Based on 45 outcomes from 24 studies, they reported that learning with CSCL scripts lead to a small positive effect on domain knowledge (d = 0.20) and a moderate to large positive effect on collaboration skills (d > 0.65) compared to unstructured CSCL. Such effects are promising, but their meta-analysis was restricted to CSCL scripts, one of the pedagogical approaches used CSCL. A more diverse array of pedagogical approaches and technological tools were examined in Chen et al. (2018) in what they defined as “learning environments or tools” and “supporting strategies”. They analyzed a number of outcome measures (e.g., knowledge achievement, skills, or perceptions) and reported positive effect sizes ranging from 0.23 to 0.89 based on results from 208 studies. They also reported that effect sizes varied significantly as a function of research design (e.g., experimental, quasi-experimental), sample size, and duration of study, although the moderator effects were significant mostly in the knowledge achievement measures.

These positive reports are promising, but the effectiveness of CSCL needs to be assessed more comprehensively and carefully. Chen et al. (2018) considered only a limited range of technologies and pedagogies, that is, seven learning environments or tools and four supporting strategies. Influential CSCL approaches such knowledge building, argumentation and problem-based learning were not included. Widely used CSCL tools and environments such as simulations, games, and/or mobile devices were not examined, They did not mention whether the collaboration was face-to-face or distributed. In addition, the analysis did not take into account that many of these technologies and pedagogical approaches may co-occur. Given that past research warned us that technology itself cannot guarantee learning and that media effectiveness may be a myth (Dillenbourg et al., 2009; Clark, 1994), it is critical is to understand the conditions under which CSCL can successfully elicit and facilitate the kinds of processes needed for students to achieve deep understanding of the learning domains. This requires a clearer assessment of the contributions made by different aspects of CSCL as well as an understanding of how they may interact and contribute to the overall effectiveness of CSCL.

1.3. Current study

This paper is built on our earlier work on CSCL research in which we have examined the nature of CSCL research practices with a focus on methodological and theoretical frameworks (e.g., Jeong, Hmelo-Silver, & Yu, 2014) and technological and pedagogical approaches surrounding CSCL (e.g., Jeong & Hmelo-Silver, 2012). The goal of the current meta-analysis is to understand the overall effectiveness of CSCL in STEM education while attending to some of the complexities that mediate the outcomes of CSCL. In order to achieve this goal, we first focused on the role of collaboration mode, technology, and pedagogy, the three pillars of CSCL. The nature of collaboration may vary a great deal depending on the mode of collaboration. Although face-to-face collaborative learning has been around for a long time, distributed collaboration is relatively new. Both synchronous and asynchronous communication technologies allow learners to collaborate anytime and anyplace, but they are also limited in the kind of information they channel through (Jeong & Hmelo-Silver, 2016). Technology supports for collaboration also extend beyond communication supports and include supports for information sharing, visualization of complex ideas and phenomena, and co-construction and community building (Jeong & Hmelo-Silver, 2016). Taking advantage of these affordances of technology critically depends on the pedagogy. It helps to make technology supported collaboration meaningful and relevant to learning, encouraging students to share and co-construct with partners and prompting them to examine the problem deeply and find out underlying mechanisms.

We additionally examine the role of learner's educational levels and learning domains. Collaborative learning is supported by technology at diverse educational levels ranging from as early as pre-K to graduate level and in a number of different domains (Gijlers, Weinberger, van Dijk, Bollen, & van Joolingen, 2013; Tsuei, 2012). It takes time and experience for students to develop the competencies to regulate their own behaviors and manage conflicts that can arise during group work. Different learning domains also focus on the development of different types of knowledge and skills, which can impose different demands and emphases on the design of instructional and technological interventions. Educational levels and learning domains were reported to be a significant moderator of learning outcomes in a number of meta-analyses on other educational interventions (Cheung & Slavin, 2012; Kyndt et al., 2013; Lou et al., 2001). This study will examine whether learning domains and educational levels can also significantly moderate outcomes in CSCL, although the examination of learning domains will be within STEM domains.

We assessed the overall effectiveness of CSCL and the effects of these moderators on three kinds of outcome measures: cognitive outcomes, processes, and affective measures. Earlier studies reported different effect sizes depending on the outcome variables, but did not always test whether the difference is significant. In this study, we additionally tested whether the differences in outcomes types are significant. We also examined how methodological features of the study affect the outcomes. Methodological features of the primary research can influence the outcomes of the meta-analysis in significant ways (Chen et al., 2018; Cheung & Slavin, 2012; Slavin, 2008) and we need to be aware of their influences in CSCL research as well. In this study, we focused on three methodological features of the study; they are research design, settings, and types of comparison condition. CSCL research has employed a diverse range of research designs (Jeong et al., 2014) and we compared the outcomes between randomized and quasi experimentation that were conducted in classrooms, laboratories, or other settings (e.g., online communities). Different conditions have been used to compare and assess the effects of CSCL applications. That is, some studies may determine CSCL effectiveness based on the comparison between CSCL and face-to-face collaboration condition, whereas others may use comparison between CSCL to individual learning conditions, either with or without computers. Such diversity in comparison conditions makes it difficult to gauge the overall effectiveness of CSCL. In this study, we examined whether outcomes differ as a function of comparison conditions (e.g., face-to-face collaboration, individual learning with or without technology supports).

A better understanding of how these variables moderate the effects of CSCL can help us to evaluate its impact more accurately and guide our future efforts in further enhancing CSCL environments and interventions. Our research questions can be summarized as:

(1) What is the overall effect of CSCL in STEM education on collaborative learning processes and outcomes?
(2) To what extent are CSCL outcomes moderated by: (a) Types of outcome measures, (b) methodological features, (c) collaboration modes, (d) technology used, (e) pedagogical approaches, (f) education levels of the learners, and (g) domains of learning.

2. Methods

2.1. Literature search and screening

A three-pronged approach was used to search for and identify CSCL studies. First, empirical STEM CSCL investigations were located using searches in Web of Science and ERIC between 2005 and 2014. In searching databases, the following two sets of keywords were used in varying combinations: (1) computer supported collaborative learning, computers, collaboration, small groups, collaborative learning, group work, and (2) science, biology, physics, chemistry, earth science, medical, health professions, math, and engineering education. In addition, we identified and hand-searched a set of representative journals that frequently publish CSCL research. They are: (a) International Journal of Computer-Supported Collaborative Learning, (b) Computers and Education, (c) Computers in Human Behavior, (d) Journal of the Learning Sciences, (e) International Journal of Artificial Intelligence, (f) Learning and Instruction, and (g) Journal of Computer-Assisted Learning. Studies were also added through other sources such as literature reviews and existing meta-analyses (e.g., Vogel et al., 2017). Combining these strategies provided ample coverage of the intersection of CSCL and STEM literature.

CSCL studies were then screened for their final inclusion in the meta-analysis. To qualify, studies first needed to be empirical research (see Jeong et al., 2014 for further details on the definition of CSCL empirical research). Second, studies needed to examine the effects of CSCL in an experimental study that was either a randomized or quasi-experimental design that compared a CSCL treatment with a comparison group. Third, studies needed to report outcome measures such as cognitive gains, processes, or non-cognitive outcomes such as affective or attitude changes. Lastly, studies needed to report sufficient statistical information to estimate effect sizes. Studies were excluded if they presented only descriptive statistics or p-values without appropriate statistics. Taken together, 143 studies were included for final analysis out of approximately 2,669 articles screened. The list of included papers are in Online Supplementary A.

2.2. Effect size

An effect size provides a standardized measure of treatment effects. It is typically computed with Cohen's d or Hedges' g (Cohen, 1988; Hedges & Olkin, 1985). Both Cohen's d and Hedges' g are used to estimate effect sizes and produce similar results, but Hedges' g corrects for smaller sample sizes (e.g., less than 20). To reduce biases from smaller samples, which is not uncommon in educational research, we used Hedges' g values in our analysis. A “rule of thumb” to interpret effect sizes is: (1) 0.20-0.49 is considered a small effect, (2) 0.50-0.79 a moderate effect, and (3) 0.80 or higher a large effect (Cohen, 1988). What counts as “large” or “small” effect sizes can vary depending on the areas of study, however. In educational research, effect sizes above 0.25 are considered to be large enough or “substantively important” to be educationally significant (Slavin, 1990; US Department of Education, 2016).

Finding-level outcomes were used as the unit of analysis. Studies often assess multiple outcomes in CSCL research, up to 44 outcomes in a single study (e.g., Gijlers et al., 2013). Although individual outcomes were coded, we collapsed them if they belonged to the same outcome types. For example, a study may have assessed and tested more than one outcome (e.g., how many concepts students draw and how many of them are annotated), but if they assessed the same outcome types, in this case students' learning processes, they were collapsed into one effect size. When there was more than one treatment condition in the experiments, we calculated effect sizes separately for each experimental condition (Durlak, 2009). For example, a study tested the effect of an agent-based instructional program with individual, jigsaw, and cooperative learning approaches (Moreno, 2009). Effect sizes were calculated for each comparison involving a control and a CSCL treatment, in this case, one for the comparison between individual and jigsaw and another for the comparison between individual and cooperative learning. In the end, the current meta-analysis is based on 316 outcomes from 143 CSCL studies. Effect sizes were extracted by two coders. The first coder extracted the effect sizes and relevant statistics, which were verified by another coder.

The Comprehensive Meta-Analysis (CMA) program (Version 2.2) developed by Borenstein, Rothstein, and Cohen (2000) was used to compute effect sizes and test the effects of moderator variables. The program computes effect sizes for various study conditions, taking into consideration pre-test scores if they exist. The random-effect model was used for synthesizing the effect sizes. Homogeneity analysis examined whether the variance exhibited by a set of effect sizes is more than the variance expected when only sampling error is involved. A significant Q value suggests that the distribution of effect sizes is not due to chance alone and there is a need to examine the outcomes further using moderator variables (Hedges & Olkin, 1985).

2.3. Moderator variables

Several different kinds of moderator variables were examined in this study. First, we examined whether the effects of CSCL varied across different CSCL outcomes. Outcomes were coded as one of the following three types: (1) cognitive, (2) affective, and (3) process outcomes. Cognitive outcomes were further coded as (1a) knowledge acquisition (e.g., understanding of concepts and principles), (1b) application of knowledge (e.g., generating a design solution,), (1c) skill (e.g., critical thinking skills), and (1d) miscellaneous cognitive outcomes (e.g., course grades, unspecified achievement tests). Affective outcomes were further coded as (2a) attitude and perceptions, (2b) motivation/interests, (2c) efficacy (e.g., confidence), and (2d) satisfaction outcomes. Process outcomes were further coded as (3a) individual (e.g., individual time on task) and (3b) collaborative processes outcomes (e.g., argumentation sequences).

Second, we examined whether the effects of CSCL varied as a function of methodological features of the study. Methodological features of the primary research affect the conclusion that can be drawn from meta-analysis (Cheung & Slavin, 2012; Slavin, 2008). The features we examined were: (1) research design (i.e., randomized versus quasi-experiments), (2) settings (labs, classrooms, or other), and (3) types of comparison groups. The outcomes of CSCL were compared to one of the five comparison conditions in CSCL research: (a) individual work without computers, (b) individual work with computers, (c) face-to-face collaboration without computers, (d) traditional classroom, and (e) existing versions of CSCL interventions. Traditional classroom comparisons refers to cases when an existing classroom practice was used to assess the effects of CSCL interventions. Existing CSCL conditiosn refers to cases when existing versions of CSCL were used to assess the effects of often-improved versions of CSCL. The nature of existing CSCL applications varied widely. For example, studies may have compared structured discussion forums with unstructured forums (e.g., Yang, Newby, & Bill, 2008) or a discussion forum with or without access to mobile devices (e.g., Lan, Tsai, Yang, & Hung, 2012). An existing collaborative game may be compared to a revised version of the game (e.g., Echeverría, Barrios, Nussbaum, Améstica, & Leclerc, 2012). An addition of specific tools (e.g., concept mapping tool) in a given CSCL environment may also be tested (e.g., Erkens, Jaspers, Prangsma, & Kanselaar, 2005; Gijlers & de Jong, 2013).

Lastly, effects of CSCL implementation contexts were examined by coding five moderator variables: mode of collaboration, technology, pedagogy, learner levels, and domains of learning (see Table 1). Mode of collaboration refers to whether learners are co-located and interacted face-to-face or are distributed and interacted either synchronously or asynchronously with technology. Technology refers to digital technologies used to support collaborative learning in both face-to-face and distributed settings and consists of 20 types of technologies grouped in six subtypes such as communication technologies (e.g., chat) and dynamic technologies that present information in dynamic forms (e.g., simulations). Pedagogy refers to the instructional contexts in which CSCL was implemented. Pedagogical contexts associated with CSCL range from teacher-structured approaches that emphasize teacher preparation and delivery of the learning materials (e.g., lectures), constructivist approaches that emphasize learner's constructive engagement (e.g., problem-based learning, inquiry learning) to approaches that emphasize social mediation of learning (e.g., scripted collaboration, argumentation). We note that there was often ambiguity with how interventions were implemented and/or described in studies. Although traditional pedagogy can mean lectures or lecture-based instruction for example, it can also mean an existing classroom instruction of unknown nature as studies often focused on the design of the tools and did not provide sufficient information about the instructional contexts in which the tool was used. Learner level refers to the educational levels of the learners and consists of six codes (e.g., elementary, secondary, and undergraduate levels). Learning domains refer to the subject area in which CSCL is implemented in and consists of five STEM domains (i.e, science, math, engineering, computer science, and health) and the education domain (e.g., educational psychology, instructional designs). The education domain was included to gauge the difference between STEM and non-STEM domains (see Table 1 for definitions and coding categories; subtypes in each category are in bold).

Coding categories used in prior work (Jeong & Hmelo-Silver, 2012) were refined and expanded to reflect development in technologies and pedagogies. Much of the current moderator variable coding was done as part of a synthesis that examines CSCL research as a whole (McKeown et al., 2017) except for the coding of the outcome types and comparison group types. Coding reliability of the additional coding was checked for 20% of the papers ($n = 30$). Two coders independently coded the comparison conditions and outcome types. The resulting Kappa values were 0.88 for comparison condition coding and 0.98 for outcome coding. Differences in coding were resolved through discussion among the coders.

2.4. Outlier analysis and publication bias

Outliers can have undue effects on the final effect size. In order to reduce their effects, outlier analyses were performed using standardized residual procedures (Hedges & Olkin, 1985). Outliers with standardized residuals larger than ± 2.00 were identified. They were carefully examined to see if there were any computational errors in the studies or if there were any features in these studies that made them different from other studies. Computational errors, if found, were corrected. Though the outliers are far removed from the majority of the studies, they may represent genuine variance associated with specific features of CSCL. CSCL involves a diverse set of technological tools, pedagogical approaches involving diverse groups of learners and domains. Excluding these studies may mask these differences and may lead to biased results. We thus decided to include these data in the analyses. However, in order to avoid the over-influence due to their extreme values, these effect sizes were reduced to values with ± 2.00 standard deviations (Tabachnick & Fidell, 1996).

Publication bias is a source of effect size overestimation. Studies reporting statistically significant results are more likely to be published and cited (Rothstein, 2008). We generated funnel plots for visual inspection of the extent of publication bias. A funnel plot is a scatter plot of the effect sizes against their respective standard errors. A publication bias results in skewed plots and uneven density of studies. We also used the trim-and-fill algorithm to estimate the number of effects that may be missing in the literature due

Table 1
Moderator variables related to CSCL implementation contexts.

Moderators	Descriptions	Coding Categories
Learner level	Educational levels of learners.	Pre-K Elementary Secondary Undergraduate Graduate Mixed or unspecified higher education Other adults (e.g., professional education)
Domain	Subject areas in which CSCL is implemented.	Science Math Engineering Computer Science Health Education[a]
Collaboration mode	Collaboration types.	Face-to-face co-located Distributed: synchronous Distributed asynchronous
Technology	Digital technologies used to support collaborative learning in face-to-face and distributed settings.	**Communication technologies:** Video Chat E-mails Discussion board (DB) Knowledge Forum (KF) **Dynamic technologies:** Simulations Games Immersive technologies **Sharing and co-construction tools:** Shared information resources Interactive whiteboard (IW) Shared workspaces Representational tools Participatory technology (e.g., wiki) Peer assessment systems Social media Group awareness tools **Systems or environments:** Integrated environments (IE) Intelligent systems (IS) **Hardware:** Mobile technologies Tabletop **Miscellaneous software and hardware**
Pedagogy	Instructional approaches in which CSCL is implemented.	**Teacher-structured approaches:** Traditional Distance education Blended courses Problem solving driven by drill and practice **Constructive engagement approaches:** Case-based Project or problem-based learning (PBL) Game-based Design-based Discovery Hands-on and active Inquiry/model-based **Socially-mediated supports:** Discussion Argumentation Scripted collaboration Knowledge-building (KB) Scaffolding Miscellaneous collaboration (e.g., jigsaw) **Miscellaneous pedagogies that do not fit into other categories**

[a] Education domain was included as a comparison.

to publication bias. In the case of significant publication bias, an adjusted effect size was computed.

3. Results

3.1. Overall effectiveness of CSCL

The overall effect size of CSCL including education domain was 0.49 with a standard error of 0.03. The 95% confidence Interval was 0.43–0.56. The overall effect size of STEM CSCL, excluding the education domain, was .51, making it a moderately sized effect; effect size difference in learning domains is examined in detail in section 3.2.7 (see Online Supplementary B for the distribution of Hedges' g value across the included studies; effect sizes are presented at the study level in Online Supplementary B).

3.2. Moderator analysis

Homogeneity analysis was significant, Q (315) = 1,866.71, $p < .001$, indicating a call for further tests of moderator variables. In examining the effects of moderators on CSCL effectiveness, note that multiple codes can be associated with a single outcome. For example, some CSCL applications may have examined the effects of single technology (e.g., discussion boards) or pedagogical strategy (e.g., scripted collaboration), while others may have used multiple technologies (e.g., discussion board along with simulations) or pedagogical strategies (e.g., scripted collaboration with inquiry learning). To clearly test the effects of moderator variables, we used the effect sizes that received a single code for a given moderator. Effect sizes with multiple code assignments were analyzed separately. In order to examine stable patterns of results, analyses are reported if there are more than five effect sizes for the given moderator variables or their combinations.

3.2.1. Outcome types

As can be seen in Fig. 1, the effect of CSCL was greatest on process outcomes ($g = .58$; $n = 34$), followed by knowledge outcomes ($g = .53$; $n = 201$). CSCL effects were smallest on affective outcomes ($g = .38$; $n = 81$). In the process outcome category, effect sizes were larger for individual process outcomes than collaborative process outcomes, suggesting that changing collaborative processes was harder than changing individual processes. It is likely because changing collaborative processes involves more than one person and coordination among a group. In the knowledge outcome category, the effects were bigger with knowledge acquisition than knowledge application. As for the affective outcomes, CSCL has larger effects on motivation and attitudes than satisfaction and self-efficacy (see Fig. 1). These differences, however, were not significant, $Q(2) = 4.99$, $p = .083$. We thus collapsed outcome types in the rest of the analysis to ensure enough cases for the moderator analysis, although we examined the outcomes at the coded level if needed.

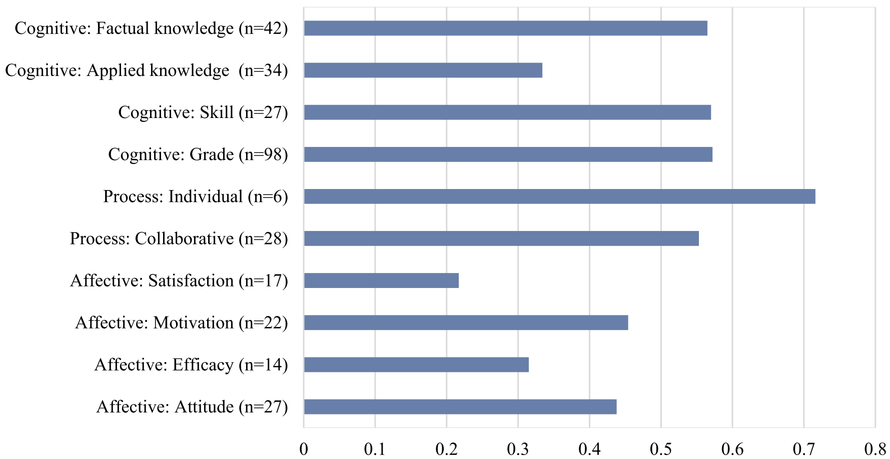

Fig. 1. Effect sizes across outcome types.

3.2.2. Methodological features of studies

Quasi-experimental studies produced significantly greater effects ($g = .58$; $n = 116$) compared with randomized experiments ($g = .44$; $n = 200$), $Q(1) = 3.77$, $p = .052$. Classroom studies also produced larger effects ($g = .54$; $n = 209$) than studies in other settings (e.g., workplaces, virtual communities; $g = .52$; $n = 7$) and laboratory settings ($g = .39$; $n = 100$), but the difference was not significant, $Q(2) = 4.39$, $p = .112$. Overall, studies with stricter control (e.g., randomized experiments, laboratory settings) produced smaller effect sizes than studies in real settings and/or looser comparison groups.

The effect size was greatest when the comparison was with traditional classrooms using little collaboration or technology ($g=.55$; $n=89$), followed by existing versions of CSCL ($g=.51$; $n=77$), and individual conditions either with ($g=.48$; $n=87$) or without computers ($g=.42$; $n=29$). The advantage of CSCL was the smallest when the comparison condition was face-to-face collaboration ($g=.41$; $n=34$). The choice of the comparison condition did not significantly influence the effect size, $Q(4) = 2.14$, $p=.710$, but the pattern of results suggests that CSCL produce greatest effects in comparison to traditional classrooms or CSCL settings in which students may already collaborate using computers but perhaps not very effectively (e.g., distance education with an inactive discussion forum). As for its comparison to individual learning conditions with or without technology, interestingly, greater (instead of smaller) effects were obtained when individual students were already using technology compared to when they were using none, suggesting that use of technology may not have been very effective in when students learn alone. Between collaboration and technology supports, collaboration seemed to matter more than technology. Still, there are small but definite benefits technology can provide as can be seen in the advantage of CSCL over face-to-face collaboration. Given that face-to-face collaboration is already proven to be quite effective, the strictest or most conservative test of CSCL effectiveness would be when its effects are compared to and tested with face-to-face collaborative conditions. The results of the study showed that CSCL was effective even with this most conservative comparison.

3.2.3. Collaboration mode

CSCL effects were similar across different modes of collaboration, $Q(3) = 1.87$, $p=.599$. The effects were almost identical in face-to-face collaboration ($g=.51$; $n=146$), synchronous collaboration ($g=.51$; $n=75$), and asynchronous collaboration ($g=.50$; $n=73$).

3.2.4. Technology

The differences across technology types were significant, $Q(17) = 46.18$, $p = .001$. Effect sizes of the technology types that have five or more outcomes are shown in Fig. 2. Simulations produced the greatest effects ($g=1.07$; $n=13$) followed by integrated environments ($g=.68$; $n=24$) and participatory technology ($g = 0.65$; $n = 9$). In CSCL, simulations were used to help learners to understand a variety of phenomena such as electrical wiring (Liu & Su, 2011) or probability (Gürbüz & Birgin, 2012). Students typically formed groups in face-to-face settings and explored simulations to construct and test models of the phenomena. Integrated environments incorporated multiple tools and ranged from course management systems designed to provide a one-stop online environment for lectures, resources, and discussion space (Cakiroglu, 2012) to environments designed with more specific pedagogical goals such as online problem-based learning (Şendağ & Odabaşı, 2009), argumentation (Bouyias & Demetriadis, 2012), or inquiry (Mäkitalo-Siegl, Kohnle, & Fischer, 2011; Slotta & Linn, 2009).

In our study, about one third of effect sizes (103 out of 316) were from studies that used multiple technological tools, and there were seven technology combinations with five or more cases (see Table 2). It appears that combining technologies does not necessarily increase the effect size of an individual technology and the effect of a technology varied depending on the technology which it was paired with. Video-conferencing appeared twice in Table 2, for example, but the outcome differed depending on the technology it was paired with; it produced a larger effect when it was used with shared workspaces ($g=.78$; $n = 6$), the biggest in Table 2, but produced a negative outcome when it was used with chat and emails ($g=-0.02$; $n = 7$). This negative effect may result from multiple tools that serve similar purposes, in this case, as communication channels. Note, however, that the results do not necessarily suggest we should refrain from combining multiple technological tools. As we noted earlier, integrated CSCL environments were quite effective, producing the second largest effects (Fig. 2). What matters more is likely to be how they are combined. It is possible that

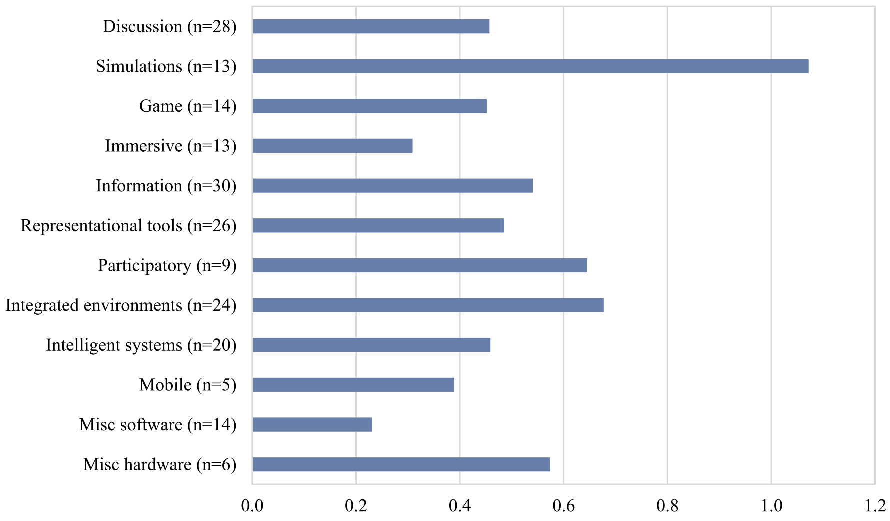

Fig. 2. Effect sizes across technological tools.

Table 2
Effect sizes for frequent technology combinations.

Technology Combinations (with n = 5 or more)	N	ES(g)
Video conferencing & Shared workspaces	6	0.78
Discussion board & Information resources	5	0.48
Chat & Representational tools	5	0.35
Knowledge Forum & Group awareness tools	9	0.30
Representational tool & Group awareness Tools	6	0.24
Games & Tabletop	6	0.15
E-mail & Chat & Video conferencing	7	−0.02
Overall	44 (103[a])	0.31(0.51[a])

[a] Values when effect sizes with n < 5 are considered.

when multiple technologies are well integrated, the learning curve for using the technology is easier with a common set of interfaces. On the other hand, when multiple technologies are not well integrated, learners may need to learn multiple interfaces, which would impose extra cognitive complexity for learners to manage. Integrated environments are also likely to integrate pedagogical and other instructional components, making CSCL a more seamless learning experience.

3.2.5. Pedagogy

The effects of CSCL varied significantly as a function of pedagogies in which they were embedded, $Q(18) = 29.74$, $p = .040$. CSCL produced the largest effects in the contexts of traditional pedagogy ($g = 1.01$; $n = 7$), followed by knowledge-building ($g = .67$; $n = 14$), and scaffolding ($g = .64$; $n = 10$). Effect sizes of the pedagogy types that have five or more outcomes are shown in Fig. 3. It is encouraging that CSCL produced positive outcomes in most of the pedagogical contexts including those that are not typically associated with collaborative learning. In fact, the effect was biggest in the traditional pedagogies that are often associated with lectures and/or individualistic learning. CSCL has the greatest potential in transforming existing practices. The effect of CSCL was negative in cased-based learning. Part of the negative outcomes appear to be due to mixed outcomes in some of the studies; CSCL intervention produced both positive and negative outcomes; it increased the amount of discourse significantly, but marginally decreased information-seeking behaviors and individual learning outcomes (Alonso, Manrique, & Viñes, 2009; Mäkitalo, Weinberger, Häkkinen, Järvelä, & Fischer, 2005).

3.2.6. Learner levels

The effects of CSCL differed significantly, $Q(5) = 11.56$, $p = .041$. As can be seen in Fig. 4, the effect was greatest with secondary students ($g = .66$; $n = 59$), followed closely by graduate students ($g = .64$; $n = 17$). Effects were smaller with elementary students ($g = .49$; $n = 65$), undergraduates ($g = .45$; $n = 94$), and other adults ($g = .11$; $n = 9$). CSCL was more effective with secondary level students than students at the elementary level. Such findings suggest that effects of CSCL would increase with the knowledge and experience of the learners, but undergraduate and other adult learners, in spite of their maturity, did not benefit as much as the

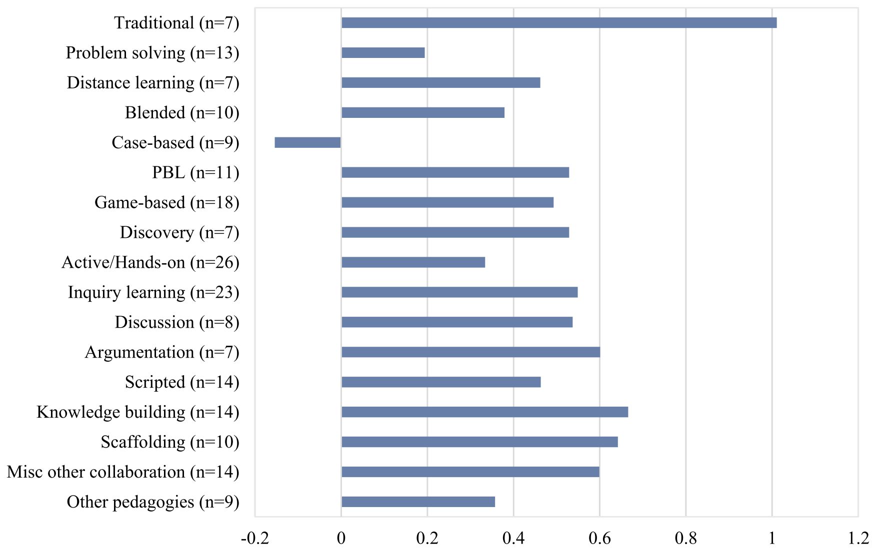

Fig. 3. Effect sizes across CSCL pedagogies.

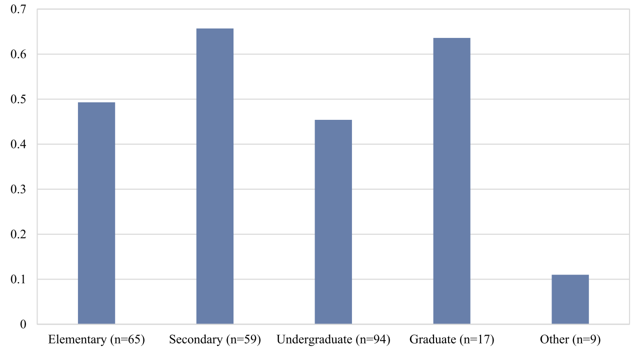

Fig. 4. Effect sizes across educational levels.

learners at the secondary level. Learner's competency to deal with collaboration and technology use would increase with educational levels, but so does the complexity of materials that they need to deal with. In addition, older learners are also placed in different socio-cultural environments with different class schedules and social demands. These contextual factors are all likely to play a part.

3.2.7. Domains

The effect of CSCL was greatest for science ($g = .67$; $n = 95$), followed by computer science ($g = .50$; $n = 76$), health science ($g = .41$; $n = 28$), and math ($g = .33$; $n = 57$). The effect size in education ($g = .39$; $n = 56$) was not as large as in science and computer science, but was similar to health science and somewhat larger than math. Effect sizes differed significantly as a function of learning domains, $Q(4) = 15.15$, p = .004 (see Fig. 5). Such results suggest that CSCL may differentially support learning across learning domains. The large effects for science domains such as physics and chemistry may be explained in that they deal with complex phenomena hidden from naked eyes. Thus technology, such as simulations, can help students visualize and explore these complex relationships.

3.3. Interactions among moderator variables

We explored some of the cross-relationships of the moderator variables to understand the workings of these variables at a deeper level. Specifically, we analyzed *how the effects of technology and pedagogies interacted with modes of collaboration, educational levels, and domains of learning*. There is no established guideline as to the minimum number of cases for this kind of analysis, but to obtain a more stable pattern of results we did this analysis when at least five effect sizes were reported for the given interaction and when there were at least two comparisons for given technology or pedagogy types (i.e., data in at least two cells for either the technology or pedagogy rows in Tables 3–5 in this section). Missing cells in the result tables indicates that there were not enough studies or effect sizes to examine the cross-relationships between the two moderator variables. Given the sparse and uneven distribution, the interaction results reported in this section need to be interpreted as exploratory only. They are presented to guide the explorations of the conditions under which CSCL effectiveness may vary.

3.3.1. Collaboration types moderated the effects of technology and pedagogy in CSCL

Five technology types were used in more than one collaboration modes (see top section of Table 3). The effects of technology varied depending on the mode of collaboration; discussion boards, information resources, and intelligent systems produced larger effects in distributed settings, whereas representational tools and integrated environments produced larger effects in face-to-face settings than in distributed settings. This appears to be in part due to the affordances of the particular technology. Discussion boards

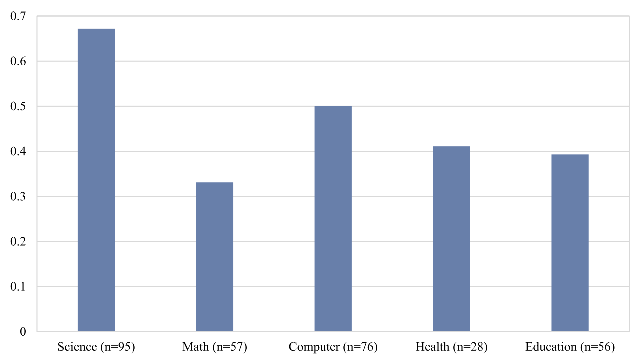

Fig. 5. Effect sizes across STEM learning domains.

are developed and used to support distributed collaboration where face-to-face contact is either impossible or infrequent. Discussion boards, when used in face-to-face settings, are also likely to have interfered with the natural dialogue in face-to-face settings. Similarly, information resources, while they can be and have been used in both face-to-face and distributed synchronous collaboration, might be better suited for asynchronous settings. It takes time to process information thoroughly. Asynchronous settings are more likely to allow learners to process information at their own pace. On the other hand, representational tools might have some advantage in face-to-face collaboration over distributed collaboration. This is likely to be due to the difficulty of understanding indexical and deictic information in distributed settings. Intelligent systems also produced quite different effects depending on the mode of collaboration, but this is likely to be confounded with the type of the intelligent system. Intelligent systems can play a different role depending on how they are designed; they may provide instructions to students' groups serving as a tutor in face-to-face settings or play the role of a simulated partners in synchronous distributed settings (Dzikovska, Steinhauser, Farrow, Moore, & Campbell, 2014). Further research is needed to understand the type and level of sophistication they provide for collaborative learning.

Table 3
Collaboration modes by technology (top section) and pedagogical approaches (bottom section).

	Face-to-face	Asynchronous	Synchronous
Discussion boards	0.33 ($n=6$)	0.64 ($n=12$)	0.40 ($n=9$)
Information resources	0.54 ($n=13$)	0.60 ($n=13$)	
Representational Tools	0.51 ($n=15$)	0.36 ($n=9$)	
Integrated environments	0.82 ($n=8$)	0.64 ($n=5$)	0.63 ($n=7$)
Intelligent systems	0.06 ($n=10$)		0.93 ($n=8$)
Inquiry learning	0.58 ($n=10$)	0.36 ($n=9$)	
Scripted	0.26 ($n=6$)		0.66 ($n=8$)
Misc. collaboration	0.52 ($n=6$)		0.70 ($n=5$)

As can be seen in the bottom half of Table 3, three pedagogical approaches were used in more than one mode of collaboration. Inquiry learning produced larger effects in face-to-face settings than in distributed settings. Inquiry learning often involves hands-on exploration and coordination of activities and may only be implemented in a limited fashion in distributed settings and may have contributed to the reduced effect size. On the other hand, scripted and miscellaneous collaboration produced larger effects in the distributed synchronous setting than the face-to-face setting. Scripted collaboration was initially developed in face-to-face settings (O'Donnell & King, 1999), but it appears to be more useful in distributed settings. Scripting collation is likely to be more needed in distributed settings where teacher presence is diminished. Modes of collaboration are likely to impose certain restrictions and/or demands on the kinds of pedagogical activities possible. The resulting effects suggest that CSCL activities and applications are likely to produce better outcomes when they meet the specific demands of different collaboration modes.

3.3.2. Educational levels moderated the effects of technology and pedagogy in CSCL

Four types of technologies were used in more than one educational level (see top section of Table 4). The effects of technology interacted with educational levels of the learners. Informational resources and intelligent systems were more effective with younger learners, whereas representational tools and integrated environments were more effective with older learners. There is no obvious explanation for this pattern of results yet, but complexity of the technology might play a role. The negative effect of intelligent systems at the secondary level was mainly due to a study that extended a computer-based tutoring system into a collaborative setting of mathematics learning both with and without scripts. This study could not establish their advantage over an individual setting with quantitative outcomes ($g = -1.01$), although qualitative analysis demonstrated positive impacts on interaction processes (Rummel, Mullins, & Spada, 2012). It is possible that it takes time for the positive changes in the interaction processes to produce solid improvement in the quantitative outcomes, but more research is needed to understand the relationships between them as well as how the effects of technology may interact with different levels of learners and other contexts of learning.

Table 4
Educational levels by technology (top section) and pedagogical approaches (bottom section).

	Elementary	Secondary	Undergraduates	Graduates
Information resources	0.48 ($n=12$)		0.32 ($n=11$)	
Representational Tools		0.40 ($n=15$)	0.56 ($n=7$)	
Integrated environments		0.49 ($n=7$)	0.54 ($n=6$)	1.15 ($n=5$)
Intelligent Systems	1.00 ($n=7$)	−0.31 ($n=6$)	0.57 ($n=7$)	
Active/Hands-on		0.91 ($n=5$)	0.02 ($n=14$)	
Misc. collaboration	0.79 ($n=6$)		0.37 ($n=6$)	

Two pedagogical approaches, active/hands-on and miscellaneous collaboration were examined at more than two educational levels (see bottom section of Table 4). Both approaches produced larger effects with younger learners than undergraduate learners. The difference of the learner level was quite striking with the active/hands-on approach: Active/hands-on approaches produced larger effects with secondary level learners, but almost no effect with undergraduate level learners. It appears that active/hands-on approach is particularly suited for secondary level learners.

3.3.3. Domains moderated the effects of technology and pedagogy in CSCL

The top section of Table 5 shows how the effects of technology may not be uniform across domains of learning. For example, information resources seem to work better in education and science than in computer science. Representational tools appear to support learning more effectively in science than in math. More data is needed, but these differences suggest that CSCL designers and practitioners need to consider the demands of the domain carefully. Education and science often focus on integrating a diverse range of information through discussion, whereas domains such as mathematics and computer science might emphasize the logical consistency and structures of the learning materials more. More research is needed to understand how technology affords the given demands of the learning domains and tasks in combination with pedagogical features of the learning environments.

Table 5
Domains of learning by technology (top section) and pedagogical approaches (bottom section).

	Science	Math	Computer	Health	Education
Discussion board	0.41 ($n=5$)				0.38 ($n=16$)
Information resources	0.58 ($n=15$)		0.29 ($n=6$)		0.69 ($n=5$)
Representational tools	0.62 ($n=10$)	0.28 ($n=10$)			
Integrated environments	0.58 ($n=8$)		0.86 ($n=9$)	0.52 ($n=7$)	
Intelligent systems	1.04 ($n=5$)	0.19 ($n=10$)			
Problem-solving		−0.60 ($n=5$)	0.52 ($n=5$)		
Active/Hands-on	0.93 ($n=6$)		0.17 ($n=10$)	0.11 ($n=9$)	
Inquiry learning	0.71 ($n=8$)	0.47 ($n=11$)			
Scripted	0.25 ($n=8$)				0.87 ($n=6$)

Four pedagogical approaches were used in multiple domains with quite different outcomes (see bottom section of Table 5). Active/hands-on and inquiry approaches were more effective in science domains than other STEM domains. On the other hand, a scripted collaboration approach was more effective in the education domain than in science. Problem solving was used both in mathematics and computer science; it was associated with a negative outcome in math, but a positive outcome in computer science. A few studies reported negative outcomes. van der Meijden and Veenman (2005), for example, compared face-to-face and computer-mediated communication groups and found that face-to-face groups performed better when solving mathematics problems, which seems to be partly because computer-mediated groups needed to engage in more regulatory activities as they could not see each other and lacked nonverbal signals. Not all pedagogical approaches are studied in all STEM domains, at least not in sufficient numbers, but the pattern of the results suggest that the effects of given CSCL pedagogies are likely to interact with the demands of the learning domains and tasks, although it may be due to the fact that certain pedagogical approaches are not sufficiently tried and elaborated in some domains.

3.4. Analysis of publication bias

To assess the publication bias on the findings of this meta-analysis, we generated a funnel plot for the outcomes (see Fig. 6) and then performed Egger's test. The result showed a significant non-zero intercept, 1.30 (CI: 0.75 to 1.85), $t(314) = 4.64$, $p = .00$, suggesting that the effect sizes are overestimated due to publication bias in this sample.

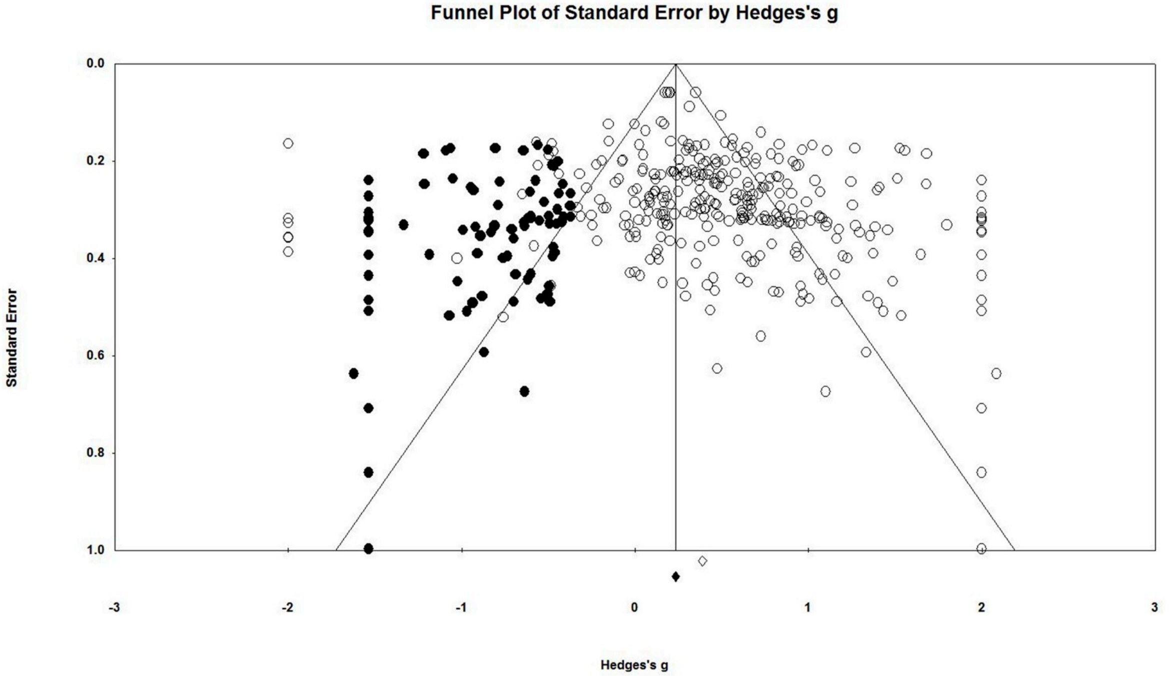

Fig. 6. Funnel plot of standard error by Hedges' *g*.

In order to understand the sources of publication bias, we first examined whether it varied depending on the outcome types. As can be seen in Table 6, Egger's test was significant mostly for product outcomes except for application of knowledge. Publication bias was not significant for process and affective outcomes. We also examined whether the results of the Egger's test varied as a function of the methodological features of the study. The Egger's test was significant for both quasi-experimental, $t(114) = 4.34$, $p < .001$ and randomized experiments, $t(198) = 3.80$, $p = .000$. In terms of setting, it was significant for classroom studies, $t(207) = 5.20$, $p < .001$ but not for laboratory studies, $t(98) = 0.22$, $p = .411$. As for the comparison conditions, Egger's test was significant for traditional classrooms, $t(88) = 3.39$, $p < .001$ but not in the other comparison conditions. These results suggest that publication bias may be a problem in some subgroups of CSCL research. Classroom studies that compared the effects of CSCL against traditional classrooms using product outcomes, especially knowledge, grades, and skills, were more prone to publication bias than other studies.

Table 6
Egger's test: Overall and by outcome types.

		Intercept	CI	df	t
Overall		1.30	0.75–1.85	314	4.64*
Product	Knowledge	2.11	0.19–4.04	40	2.22*
	Application	−0.03	−1.20–1.14	32	0.05
	Skill	2.63	−0.46–5.73	25	1.75*
	Misc.	1.47	0.50–2.46	96	2.96*
Process	Individual	−0.38	−6.89–6.13	4	0.16
	Collaborative	1.62	−0.42–3.67	26	1.63
Affective	Satisfaction	3.00	−2.34–8.34	15	1.20
	Motivation	1.40	−1.10–3.89	20	1.17
	Efficacy	0.96	−0.23–2.15	12	1.76
	Attitude	0.79	−1.38–2.97	25	0.75

* significant at $p < .05$.

The trim-and-fill method was used to adjust for the publication bias based on the funnel plot (Duval & Tweedie, 2000; Taylor & Tweedie, 1998). The basic procedure involves removing (trimming) studies causing plot asymmetry and augmenting the data so that the plot is more symmetric (filling). The result of the trim-and-fill analysis is presented in Table 7. Eighty-four studies were trimmed with an adjusted effect size of 0.23 with a random effect model. The 'corrected' effects of CSCL estimates need to be taken with caution. The trim and fill method accounts for publication bias, but funnel plot asymmetry can be caused by reasons other than publication bias. The trim-and-fill method can inappropriately adjust for publication bias when there is substantial between-study heterogeneity (Peters, Sutton, Jones, Abrams, & Rushton, 2007) and need to be interpreted with caution in studies like our study that

Table 7
Duval's trim and fill test.

	Studies trimmed	Point estimate	Lower limit	Upper limit	Q
Observed values		0.49	0.43	0.56	1867.23
Adjusted values	84	0.23	0.17	0.30	3169.85

synthesizes effect sizes over a very diverse set of CSCL practices involving different technology, pedagogy, and other contextual factors.

4. Discussion

4.1. CSCL is effective

Computers and other digital technology can provide a number of affordances for collaborative learning (Jeong & Hmelo-Silver, 2016), but they also bring in a number of challenges. The results of our study indicate that, in general, CSCL produces positive outcomes in STEM with the overall effect size of 0.51. The range of effect sizes we found for different outcomes types are comparable to the effect sizes reported in Chen et al. (2018) despite their coding the outcomes differently. The overall effect size we found in this study was moderate according to Cohen's guideline, but is quite substantive for educational research (U.S. Department of Education, 2016). CSCL experience had the largest effects on the students' learning processes but also showed effects on learning gains and affective outcomes such as motivation, self-efficacy, and attitude toward STEM domains. The overall effects of technology support for collaborative learning in STEM is positive, in spite of the complications that may arise in the process of coordinating collaborative activities around and with technology.

4.2. Toward a more sophisticated understanding of CSCL effectiveness

In order to better understand conditions under which CSCL is effective, we examined five moderator variables that are relevant to how CSCL was implemented: mode of collaboration, learner educational levels, domain of learning, technology types, and pedagogical strategies. The results indicated that CSCL outcomes varied significantly as a function of the technology and the pedagogy in which the technology is embedded, Similar findings related to technology and pedagogy were also reported in Chen et al. (2018), but direct comparisons are not possible due to differences in coding categories. They did not isolate the contribution of different tools and pedagogical supports. By including studies that are associated with single type of technology or pedagogy, for example, we were able to assess the contribution of different moderator variables more clearly. Consistent with meta-analyses of other educational interventions, our study found that educational level and learning domains were significant moderators of CSCL outcomes. This is in contrast to Chen et al. (2018) that reported that these two were not significant. However, no coding schemes or results were presented with these claims, and more details are needed to be able to understand these contrasting findings.

When we consider the effects of these moderators, it is important to note that their effects are likely to result from a complex set of interactions with other moderator variables. We examined some of the interactions in this study, but do not have sufficient data yet to comprehend the full extent of this interaction. We also note that there is likely ambiguity with how interventions are described, such that active/hands-on may not be clearly distinguished from inquiry, and inquiry and problem-based pedagogies are not always clearly distinct (e.g., Hmelo-Silver, Duncan, & Chinn, 2007). Still, the exploratory analysis reported in section 3.3 suggests that the effects of a particular technology or pedagogy are likely to vary as a function of the collaboration mode, learner levels, and domains of learning. Representational tools produced bigger effects in face-to-face collaboration than in distributed collaboration, for example. Similarly, active/hands-on pedagogies produced moderate overall effects, but were quite effective with secondary level students and/or in science education. The current study examined a quite expansive array of moderators, but still falls short of examining all moderators that are potentially important to the success of CSCL. Continued efforts are needed to further synthesize and elaborate the outcomes of CSCL. CSCL designers and researchers need to be mindful of the complex web of interrelationshps surrounding CSCL (Hew & Cheung, 2013).

4.3. Next steps

Meta-analysis does not have control over the quality of the primary studies included in the analysis. The principal claim of a meta-analysis is from the primary studies in which the data are collected. The current study attempted to control for the methodological quality of the study by selecting experimental and quasi-experimental studies as well as examining how the outcomes varied depending on a number of methodological dimensions. Still, the range of study quality was quite diverse. More stringent control over the study quality may be needed in the future to get a more accurate estimate of CSCL effects. The current meta-analysis is based on the published studies only. Although our corpus included a number of studies with negative outcomes, there is a tendency to underreport negative outcomes. Including grey literature such as unpublished theses and conference proceedings could likely reduce the publication bias in future synthesis efforts. At the same time, the asymmetry in the funnel plot and resulting assumption of publication bias in this study can also be due to the expansive nature of the current meta-analysis. Our meta-analysis attempted to integrate CSCL outcomes across a diverse set of technologies, pedagogies, educational levels, learning domains, and modes of collaboration. They are not evenly distributed in CSCL research, but rather clustered around a combination of technologies, pedagogies, and learner types (McKeown et al., 2017). The uneven distribution of these pockets of CSCL practices is likely to contribute to the plot asymmetry. Advances in appropriate statistical techniques are needed to deal with the diversity in composite studies in the corpus.

The current study synthesized CSCL outcomes across a diverse set of CSCL studies. Such synthesis efforts are useful, but we should not lose sight of the specifics of CSCL. We need to be aware of the fact that quantitative summarization of CSCL outcomes may fall short of revealing the complexities involved in implementing as well as evaluating CSCL effectiveness. Not all educationally meaningful variables can be manipulated and assessed in quantitative manners. Qualitative research and analyses help us to understand how a given CSCL application may succeed or fail within the specific context of learning beyond the small subset of quantifiable measures. Given that a big proportion of CSCL studies use some form of qualitative analysis (Jeong et al., 2014), efforts are needed to synthesize findings and lessons from qualitative studies as well. Efforts are currently underway to construct a more nuanced understanding of CSCL effectiveness incorporating both quantitative and qualitative outcomes (McKeown et al., 2017).

What does this all mean for teachers and practitioners? The rapid development of technology means teachers, as well as researchers, will be in a perpetual search for the next new technology that can improve and even transform teaching and learning. An intuitive strategy in the face of a large set of available technology has been to have a full range of tools available with a common interface so that teachers and students can pick and choose what they need. However, providing more is not necessarily better. The results of the current study showed that the effects of CSCL is dependent on a diverse range of variables and we need to pay attention to the contingencies that exist among these variables. This means that in addition to selecting appropriate CSCL interventions, careful attention needs to be paid to who is using the tools and in what contexts. In providing information resources to students, for example, teachers need to pay attention to whether the resources are appropriate for the learning topics, whether students are given sufficient time to process them as well as possess the ability to process them, and/or whether students can participate in meaningful and engaging collaborative activities with the resources. Understanding and orchestrating all these contingencies is challenging and by no means an easy task. Support and training needs to be provided to teachers and practitioners so that they can better manage how these tools are used and orchestrate the pedagogical interventions in their classrooms. We also need to cultivate the technological readiness of the students and build infrastructures of schools and other CSCL settings. Researchers as well as practitioners may be able to work on only parts of the complex ecology of CSCL at a time, but understanding how components fit together does matter. We hope that the current study will advance research and practice in CSCL as we mve towards an improved understanding of CSCL environments.

Acknowledgment

This work was supported by the US National Science Foundation [Grant No 1249492, 2012]; and the National Research Foundation of Korea [Grant No NRF-2016R1D1A1B03935697, 2016]. We thank MyeongHo Shin and Andrew Walker for their contribution and feedback on an earlier phase of the work. We also thank the efforts of those how have involve in earlier phase of the project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2019.100284.

References

Alonso, F., Manrique, D., & Viñes, J. M. (2009). A moderate constructivist e-learning instructional model evaluated on computer specialists. *Computers & Education, 53*(1), 57–65.

Arnseth, H. C., & Ludvigsen, S. (2006). Approaching institutional contexts: Systemic versus dialogic research in CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 167–185.

Borenstein, M., Rothstein, H., & Cohen, J. (2000). *Comprehensive meta-analysis: A computer program for research synthesis [computer software]*. Englewood Cliffs, NJ: Biostat.

Bouyias, Y., & Demetriadis, S. (2012). Peer-monitoring vs. micro-script fading for enhancing knowledge acquisition when learning in computer-supported argumentation environments. *Computers & Education, 59*(2), 236–249.

Cakiroglu, U. (2012). Comparison of novice programmers' performances: Blended versus face-to-face. *Turkish Online Journal of Distance Education, 13*(3), 135–151.

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C. C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research, 88*(6), 799–843.

Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198–215.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73–105.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *Handbook of research synthesis and meta- analysis.* New York: Russell Sage Foundation.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In N. Balacheff, S. Ludvigsen, T. Jong, A. Lazonder, & S. Barnes (Eds.). *Technology-enhanced learning* (pp. 3–19). Springer Netherlands.

Doise, W., Mugny, G., & Perret-Clermont, A. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*(3), 367–383.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method. *Biometrics, 56*(June), 455–463.

Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep natural language understanding and automatic feedback generation for

intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education, 24*(3), 284–332.
Echeverría, A., Barrios, E., Nussbaum, M., Améstica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education, 59*(2), 806–816.
Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*, 399–483.
Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior, 21*(3), 463–486.
Gijlers, H., & de Jong, T. (2013). Using concept maps to facilitate collaborative simulation-based inquiry learning. *The Journal of the Learning Sciences, 22*(3), 340–374.
Gijlers, H., Weinberger, A., van Dijk, A. M., Bollen, L., & van Joolingen, W. (2013). Collaborative drawing on a shared digital canvas in elementary science education: The effects of script and task awareness support. *International Journal of Computer-Supported Collaborative Learning, 8*(4), 427–453.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.
Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Higher Education, 5*, 351–379.
Greeno, J. G. (2006). Authorative, accountable positioning and connected, general knowing: Progressive themes in understanding transfer. *The Journal of the Learning Sciences, 15*(4), 537–547.
Gürbüz, R., & Birgin, O. (2012). The effect of computer-assisted teaching on remedying misconceptions: The case of the subject “probability”. *Computers & Education, 58*(3), 931–941.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.
Hew, K. F., & Cheung, W. S. (2013). Use of Web 2.0 technologies in K-12 and higher education: The search for evidence-based practice. *Educational Research Review, 9*, 47–64.
Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and how do students learn? *Educational Psychology Review, 16*, 235–266.
Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107.
Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.
Jeong, H., & Hmelo-Silver, C. E. (2012). Technology supports in CSCL. In J. van Aalst, K. Thompson, & M. J. Jacobson (Vol. Eds.), P. Reimann (Vol. Ed.), *Proceedings of International Conference of the Learning Sciences (ICLS) 2012: Vol. 1*, (pp. 339–346). Sydney, Australia: International Society of the Learning Sciences.
Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist, 51*, 247–265.
Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *International Journal of Computer-Supported Collaborative Learning, 9*(3), 305–334.
Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379.
Kafai, Y. B., Fields, D. A., & Burke, W. Q. (2010). Entering the clubhouse: Case studies of young programmers joining the Scratch community. *Journal of Organizational and End User Computing, 22*, 21–35.
Kershner, R., Mercer, N., Warwick, P., & Staarman, J. K. (2010). Can the interactive whiteboard support young children's collaborative communication and thinking in classroom science activities? *International Journal of Computer-Supported Collaborative Learning, 5*, 359–383.
Kirschner, P., & Erkens, G. (2006). Cognitive tools and mindtools for collaborative learning. *Journal of Educational Computing Research, 35*(2), 199–209.
Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*, 159–185.
Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review, 10*, 133–149.
Lan, Y. F., Tsai, P. W., Yang, S. H., & Hung, C. L. (2012). Comparing the social knowledge construction behavioral patterns of problem-based online asynchronous discussion in e/m-learning environments. *Computers & Education, 59*(4), 1122–1135.
Liu, H. C., & Su, I. (2011). Learning residential electrical wiring through computer simulation: The impact of computer-based learning environments on student achievement and cognitive load. *British Journal of Educational Technology, 42*(4), 598–607.
Lou, Y., Abrami, P. C., & D'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449–521.
Mäkitalo-Siegl, K., Kohnle, C., & Fischer, F. (2011). Computer-supported collaborative inquiry learning and classroom scripts: Effects on help-seeking processes and learning outcomes. *Learning and Instruction, 21*(2), 257–266.
Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: Fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior, 21*(4), 603–622.
Mateu, J., & Alaman, X. (2013). CUBICA: An example of mixed reality. *Journal of Universal Computer Science, 19*(17), 2598–2616.
McKeown, J. M., Hartley, K. A., Faulkner, R. T., Hmelo-Silver, C. E., Jeong, H., & Emmanuel, N. (2017). A meta-synthesis of CSCL literature in STEM education. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Vol. Eds.), *Proceedings of Computer-Supported Collaborative Learning (CSCL) 2017: Vol. 1*, (pp. 439–486). Philadelphia, United States: International Society of the Learning Sciences.
van der Meijden, H., & Veenman, S. (2005). Face-to-face versus computer-mediated communication in a primary school setting. *Computers in Human Behavior, 21*(5), 831–859.
Moreno, R. (2009). Constructing knowledge with an agent-based instructional program: A comparison of cooperative and individual meaning making. *Learning and Instruction, 19*(5), 433–444.
O'Donnell, A. M., & King, A. (1999). *Cognitive perspectives on peer learning.* Mahwah, NJ: LEA.
Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562.
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*(2), 240–257.
Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology, 4*(1), 61–81.
Rummel, N., Mullins, D., & Spada, H. (2012). Scripted collaborative learning with the cognitive tutor algebra. *International Journal of Computer-Supported Collaborative Learning, 7*(2), 307–339.
Şendağ, S., & Odabaşı, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education, 53*(1), 132–141.
Serrano-Cámara, L. M., Paredes-Velasco, M., Alcover, C. M., & Velazquez-Iturbide, J.Á. (2014). An evaluation of students' motivation in computer-supported collaborative learning of programming concepts. *Computers in Human Behavior, 31*, 499–508.
Sfard, A. (1998). On two metaphors of learning and the dangers of choosing just one. *Educational Researcher,* 4–13.
Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 93*(4), 429–445.
Slavin, R. E. (1990). IBM's writing to read: Is it right for reading? *The Phi Delta Kappan, 72*(3), 214–216.
Slavin, R. E. (2008). Perspectives on evidence-based research in education–what works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*, 5–14.
Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom.* New York: Teachers College Press.
Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51.
Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press.
Stahl, G., Koschmann, T., & Suthers, D. D. (2014). Computer-supported collaborative learning. In R. K. Sawyer (Ed.). *Cambridge handbook of the learning sciences* (pp. 479–500). (2nd ed.). NY: Cambridge University Press.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 315–337.
Tabachnick, B. G., & Fidell, L. S. (1996). Analysis of covariance. *Using multivariate statistics, 8*(1), 321–374.
Taylor, S., & Tweedie, R. (1998). *Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analyses.* Fort Collins, CO: Colorado State University.
Tsuei, M. (2012). Using synchronous peer tutoring system to promote elementary students' learning in mathematics. *Computers and Education, 58*(4), 1171–1182.
U.S. Department of Education (2016, September). *What works clearing house procedures and standards handbook (Version 3.0)*. Retrieved from http://whatworks.ed.gov.
Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*(3), 477–511.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.
Yang, Y. T. C., Newby, T., & Bill, R. (2008). Facilitating interactions through structured web-based bulletin boards: A quasi-experimental study on promoting learners' critical thinking skills. *Computers & Education, 50*(4), 1572–1585.
Zhang, J., & Norman, D. A. (1994). Representation in distributed cognitive tasks. *Cognitive Science, 18*(1), 87–122.