Journal of Behavioral Education
https://doi.org/10.1007/s10864-023-09513-9

ORIGINAL PAPER

A Systematic Review and Meta-Analysis of Productive Peer Talk Moves

Liru Hu[1] · Gaowei Chen[1]

Accepted: 19 February 2023

Abstract
Productive peer interactions are often characterized by productive peer talk moves. This study aims to synthesize an empirical list of productive peer talk moves from existing studies, quantify the efficacy of talk moves in promoting peer interaction and collaboration outcomes, and understand the preconditions of talk moves benefits in authentic settings. A total of 24 empirical studies were included in the systematic review, where 17 of the experimental studies ($k = 39$, $n = 2636$) were analyzed in meta-synthesis. The study offers three main contributions: (1) an ordered list of 24 productive peer talk moves extracted from a range of empirical studies; (2) the aggregate sizes of the positive effects that productive peer talk moves have on interaction quality (Hedges' $g = 1.27$), domain-specific knowledge ($g = 0.96$), domain-general knowledge ($g = 1.02$), and solution quality ($g = 0.70$); and (3) common explanations for the malfunction of productive peer talk moves in existing interventions. This review confirms the robust positive effects of productive peer talk moves on peer interaction, learning, and problem-solving, and may inform future research on the analysis of peer interaction or the design of peer talk scaffolds.

Keywords Productive peer talk · Talk move · Collaborative script · Collaborative discourse · Review

✉ Liru Hu
liruhu.hku@gmail.com

Gaowei Chen
gwchen@hku.hk

[1] Faculty of Education, The University of Hong Kong, Room 525, Meng Wah Complex, Pokfulam Road, Hong Kong 999077, Hong Kong

Published online: 12 March 2023

Introduction

Collaboration, where co-equal parties engage in shared decision making towards a common goal (Friend & Cook, 1992), has been widely acknowledged as one important twenty-first-century skill (Griffin et al., 2012). It has the potential to bring social, cognitive, and emotional gains for students of various educational levels (Miyake & Kirschner, 2014). Therefore, it has become a common educational practice in classrooms of primary schools, secondary schools, and universities.

The benefits of collaboration only happen when students engage in productive peer talk where they elaborate and justify their own ideas and engage with others' ideas. Productive peer talk can help students deepen individual thinking through activities such as elaboration, justification, and reflection. It can also promote students to think with others through activities such as evaluation, building on each other, and pressing for reasoning (Gillies, 2019). It helps reduce superficial collaborations where students just pool their ideas without constructive criticism or cannot effectively make consensus (Littleton & Mercer, 2013). Yet, students, no matter kids or adults, seldom spontaneously engage in productive peer talk when they are just seated together. Research has shown that explicit teaching of how to talk with their peers is necessary to ensure the benefits of collaboration (Blatchford et al., 2003; Gillies, 2003; Gillies & Haynes, 2011).

Some researchers have identified a set of fine-grained features of group talk (i.e., talk moves) that are academically productive, such as "elaborate on information," "justify one's idea," "challenge other's idea," and "press for reasoning". These talk moves aim to help peers initiate and sustain productive interactions in problem-solving and facilitate high-level cognitive processing. Different talk moves perform different local social and cognitive functions. For example, the "press for reasoning" talk move aims to promote thinking with others and the cohesion of a learning community. Students are taught these beneficial practice to structure their communication with group members (Gillies, 2019; King, 1997; Noroozi et al., 2013a, 2013b; Webb et al., 2014). For example, an early study by King (1990) developed a guided reciprocal peer-questioning procedure to teach college students how to ask peers strategic questions, such as "what does … mean", "explain why…", and "how are … and … similar?" King (1997) also applied this procedure to guide the discussion of primary school students. These students could access to papery prompt cards of these question starters as external scaffolds at first which then gradually faded away. Students were expected to internalize the external talk scaffolds through practice and the timely fading of these scaffolds (King, 2007). King has demonstrated that her guided reciprocal peer-questioning procedure could improve peer interaction and academic outcomes for both college and primary school students.

Different scholars have referred to such fine-grained behavioral supports for academically productive peer talk as different names, such as micro-collaboration scripts (Gelmini-Hornsby et al., 2011; Noroozi et al., 2013a, 2013b; Stegmann et al., 2012), questioning prompts (Ge & Land, 2004; King, 2008), sentence openers/

starters (Cook, 2008; Gogoulou et al., 2008; Teo & Daniel, 2007), problematizing scaffolds (Molenaar et al., 2011, 2014), and communicative acts (Gogoulou et al., 2008; Hennessy et al., 2016). In this review, we call all of them as "productive peer talk moves" for convenience and transform them as verb-object and content-free phrases. For example, question stems such as "why do you think…?", "can you explain…?", and "what is your evidence for…?" are all various versions of "press (i.e., verb) for reasoning (i.e., object)". Students can use it to generate their own content-specific questions (e.g., can you explain why you think the first is better?).

Theoretical Background

Productive peer talk moves encourage learners to verbalize their viewpoints and underlying reasoning, as well as verify, evaluate, and build on the contributions of peers. In cognitively demanding social contexts, students are likely to experience socio-cognitive conflicts which refer to disturbances of one's cognitive systems resulting from others' different conceptions (Mugny & Doise, 1978). Such disturbances motivate individuals to re-examine their own ideas and seek additional information to resolve the conflicts and achieve equilibrium. Piaget (1932) viewed social interaction as a method to disrupt children's egocentrism by exposing them to multiple perspectives. This was called cognitive dissonance (Piaget, 1932), and it was later formalized as a socio-cognitive conflict theory by neo-Piagetians (e.g., Mugny & Doise, 1978). Studies confirmed that socio-cognitive conflict is a strong predictor of group performance (Howe et al., 1992).

Productive peer talk moves have dual functions: (1) elicit high-level cognitive activities and (2) facilitate high-level social interdependence. Regarding the first function, generally, students seldom spontaneously engage in thoughtful questioning; they tend to focus on answers rather than the problem-solving process (Byun et al., 2014). Productive peer talk moves are designed to elicit high-level cognitive processing (e.g., explaining, questioning, arguing, monitoring, evaluating, reflecting, and summarizing). They helped induce beneficial behavior patterns where students constructively engage with each other's viewpoints and involve students in cognitively demanding situations (Bouyias & Demetriadis, 2012; King, 1991; Noroozi et al., 2013a, 2013b; Popov et al., 2019). Productive peer talk moves also limited low-level responses like repeating others' viewpoints (King, 1990), delayed students' primary reactions to others' contributions (e.g., immediately expressing their own opinions rather than verifying others' first) (Kirschner et al., 2008), reduced detrimental behaviors (i.e. ignoring peer questions or errors) (Webb, 1982), and prevented students from relapsing into novice-level strategies (e.g., getting satisfied with one solution and failing to check room for improvement or consider alternatives in case of problems with their single solution) (Byun et al., 2014; Wecker & Fischer, 2010).

Regarding the second function of productive peer talk moves, the built-in reciprocal nature of productive talk moves was found to promote peer interdependence and facilitate high-quality questioning and responses (King, 1990). Productive peer talk moves encouraged students to verbalize their covert cognitive processes, which helped expose and resolve peers' cognitive discrepancies and facilitated shared understanding within a group (Gelmini-Hornsby et al., 2011; King, 1991; Kirschner et al., 2008). In addition, students had opportunities to model sound cognitive strategies in the process of exposing themselves to the overt cognitive and metacognitive behaviors of others (King, 1991).

The high-level cognitive and social processes fostered by productive peer talk moves might manifest in forms of overt high-level collaborative discourse (King, 1991). Such overt high-level discourse was assumed to correspond to underlying high-level collaborative knowledge construction (e.g., knowledge assimilation and integration), which was in turn expected to enhance learning and solution quality (King, 1994).

The Present Study

Although studies have identified a set of peer talk moves that facilitate student engagement in productive group work, there has been variability in the wording and presentation of the talk moves. To date, consensus on a list of evidence-based productive peer talk moves has not been reached though some essential efforts for such consensus have been made (e.g., Hennessy et al., 2016). In addition, the seemingly validated benefits of talk moves require theoretical and empirical support (Sfard, 2020). Although empirical studies have demonstrated the benefits of peer talk interventions (e.g., Bouyias & Demetriadis, 2012; King, 1994; Popov et al., 2019; Webb, 1982), the effect size of productive peer talk moves on enhancing collaborative discourse, domain learning, and problem-solving remains unclear. The present study aims to synthesize an empirically grounded peer talk repertoire that can be readily adopted in practice to scaffold productive peer interaction. It is also going to aggregate the exact impact sizes of productive peer talk moves on collaborative interaction and outcomes based on existing quantitative findings.

Recently, two relevant meta-analyses of computer-supported collaborative learning (CSCL) scripts (Radkowitsch et al., 2020; Vogel et al., 2017) examined the empirical studies that analyzed the effectiveness of CSCL scripts and found robust positive effects of such scripts on domain learning and collaboration skills. CSCL scripts include both micro-collaboration scripts (e.g., scaffolding argumentation or question asking) and macro-collaboration script (e.g., prompting problem-solving steps, assigning members different roles). In addition, they also include content-related scripts that provide domain-specific prompts. However, these two meta-analyses did not systematically examine the effect of productive peer talk moves, a type of micro-collaboration scripts. Neither did they consider situations without computers such as providing students with papery prompt cards. These are the aspects that the current study aims to build on by focusing on the effectiveness of fine-grained peer talk moves in all levels of educational contexts with or without the computer support.

Collaboration research is primarily concerned about both proximal (e.g., intersubjectivity, knowledge construction) and distal outcomes (e.g., better group solution or better individual understanding of a concept) (Enyedy & Stevents, 2014). Productive peer talk moves act as a structure for peer interaction and may easily induce change within collaboration. However, productive peer talk moves may sometimes fail to generate expected distal outcomes in authentic interventions (e.g., Hu, 2020; Stegmann et al., 2007). Proximal outcomes are also not ensured when students are presented with productive peer talk moves without teacher guidance. Therefore, this study also aims to identify possible conditions to observe true effects of productive peer talk moves and contribute to our understandings when negative outcomes result from productive peer talk moves.

In brief, this review aims to answer the following questions:

1. In existing studies, what are the frequently used productive peer talk moves?
2. What are the specific aggregate effect sizes of productive peer talk moves that promote collaborative interaction and outcomes?
3. Why do productive peer talk moves sometimes fail to generate expected outcomes?

Method

Data Collection and Inclusion Criteria

This review followed the procedures of the Preferred Reporting of Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009). To ensure a rigorous and comprehensive study, we first reviewed classical studies to extract a pool of keywords that will systematically locate studies of talk moves. The search was restricted to studies of talk moves in peer interactions by adding search terms such as collaborat*, group, peer, and collective. We used the wildcard character * to search various forms of a keyword. For example, "collaborat*" includes various forms of "collaborate" like collaboration, collaborating, and collaborative. Specifically, we used the following search string to search titles, subjects, and abstracts in the EBSCO, Web of Science ProQuest, and IEEE Xplore databases:

> (Scaffold* OR "questioning frames" OR "sentence opener" OR "sentence starter" OR "prompt" OR "talk move" OR "talk tool" OR "question starter" OR "script*" OR "strategic questioning" OR "guided reciprocal peer questioning" OR "speech act" OR "dialogue act" OR "communicative act" OR "structured" OR "structuring" OR "productive talk" OR "exploratory talk" OR support*) AND (talk OR conversation OR discussion OR dialogue OR interaction OR argument* OR discourse) AND (collaborat* OR group OR peer OR collective)

There was no limit on the range of publication years because we aim to cover all available studies. The literature review occurred over two rounds, taking place in August 2020 (first round) and January 2021 (second round). Studies were included if they: (1) were peer-reviewed full-length journal articles, (2) examined talk moves in peer interactions, (3) contained empirical evidence of the efficacy of talk moves, and (4) were written in English. We restricted studies in peer-reviewed journals and did not further screen for methodological rigor. From the articles, we selected eligible studies for a quantitative meta-analysis. Studies were included in the meta-analysis if they satisfied the following additional criteria: (1) compared peer interaction or group outcomes of productive peer talk moves to that of a control group, (2) measured peer interaction or group outcomes using objective quantitative criteria (i.e., domain knowledge tests, reasoning ability tests, or quantitative measures of interaction) rather than self-reported measures (i.e. surveys or interview data on students' perceived interaction experiences), and (3) contained sufficient statistical information to compute effect sizes.

The initial searches returned 770 primary studies (see Fig. 1). After title screening, 189 were retained for abstract screening and then 128 for further full-text review. Both abstract screening and full-text review were conducted following the inclusion criteria to identify suitable studies. The final sample for the systematic review contained 24 studies, of which 17 were included in the meta-analysis.

Study Selection and Data Extraction

One trained coder and the first author conducted the search and data coding processes, where they independently screened the titles and abstracts of all the

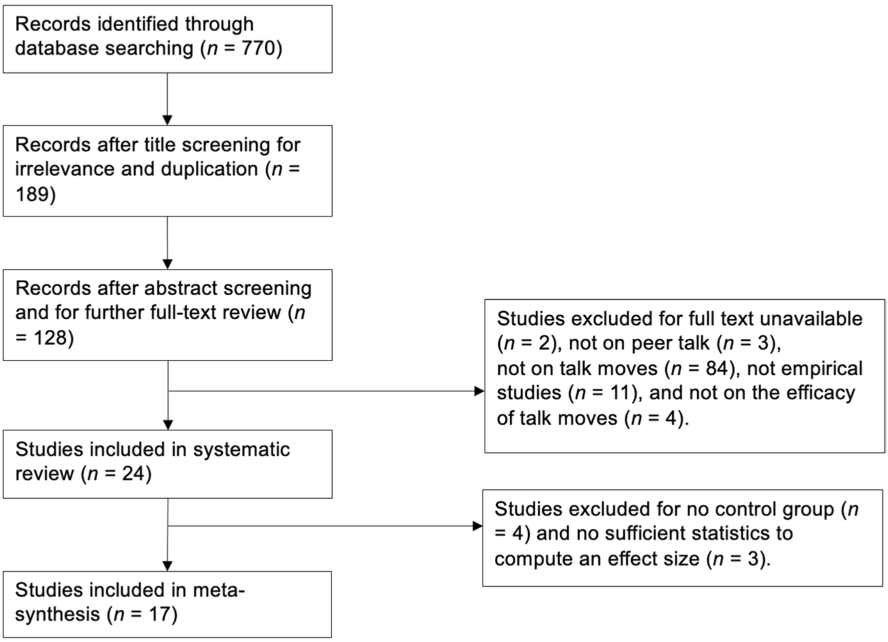

Fig. 1 Flow diagram of study selection

Table 1 Codes for characterizing the studies

Category	Code	Description
Study	Study	Information about the authors and year of publication
	Location	Country/region in which the institution of the first author was located
	Outcome	Dependent variables investigated, including interaction quality, knowledge learning, and solution quality
	Background	Theoretical background of the study, such as collaborative learning and collaborative problem-solving
Participant	Grade	Educational level of the participants
	Sample size	Total number of participants
	Group size	Number of members of each group
Intervention	Medium	Communication medium of peer interaction, i.e., face-to-face or online
	Duration	Length of the intervention period
	Intervention	Talk scaffolds designed in the study, such as sentence openers, questioning prompts and collaborative scripts
	Talk moves	Consistently formatted talk moves recoded from talk scaffolds in the study
Quality	Certainty of evidence	The degree of confidence about that the observed outcomes are the result of the intervention

citations in the articles that were identified in the initial search. The coder and author also checked the full texts of the studies that met the inclusion criteria. The inter-coder reliability, as measured by Cohen's κ, on the inclusion of studies was high ($\kappa > 0.80$; Landis & Koch, 1977). All disagreements at each stage were resolved through discussion. A set of codes was developed to characterize the sampled studies based on the research questions (see Table 1).

A quality appraisal was conducted following a two-step strategy adopted in previous reviews to categorize the certainty of evidence of each included study as either conclusive or inconclusive (Sigafoos et al., 2009, 2019). Specifically, we evaluated the research design of the study at the first step. Those studies that did not adopt a recognized experimental design (e.g., randomized controlled trial) would be labeled as inconclusive evidence. At the second step, we further evaluated whether the experimental studies qualify the following criteria: (a) have convincing demonstration of the intervention effect (e.g., statistically significant results), (b) have sufficiently reliable measures (Cohen's $\kappa > 0.70$ or Cronbach's $\alpha > 0.70$), (c) have sufficient details for replication (e.g., operational definitions of dependent and independent variables, detailed description of experiment procedures). We chose 0.70 as the threshold of adequate inter-coder agreement because this value could indicate satisfactory agreement (Cooper, 2018) and could be tolerated in the context of discourse labeling.

Analysis Methods

To establish a repertoire of frequently used productive peer talk moves (RQ1), we reworded or categorized the talk scaffolds into a consistent set of talk moves, which was necessary because different scholars would use variations in wording for similar talk moves. For example, we coded "ask for explanations", "why is … important?", "why do you think of it as a problem?", and "why…" as examples of a "press for reasoning" talk move. The inter-coder reliability was close to unity for this recoding process ($\kappa = 0.95$). The two coders discussed all disagreements and came to a consensus after reexamining the studies together.

To quantify the efficacy of productive talk moves (RQ2), we adopted a meta-analytic approach that aggregated the available quantitative findings. There was a limited number of primary studies, and in order to conduct an effective meta-analysis, we focused on four outcomes that indicated the efficacy of productive peer talk moves: interaction quality, domain-specific knowledge acquisition, domain-general knowledge acquisition (e.g., argumentation), and solution quality. The quality of an interaction can be measured using various criteria, but this review focused on learner's engagement in high-level cognitive thinking. In the sampled studies, learner's engagement was usually measured through a coding and counting approach to summarize the occurrence of certain discourse units involving high-level cognition. The acquisition of domain-specific or domain-general knowledge was mainly tested through pre- and post-objective tests, and solution quality was measured by assessing the quality of joint or individual outputs, typically in the context of collaborative problem-solving.

Hedges' g was chosen to measure the standard effect size because it represents a sample-adjusted standardized mean difference (Hedges & Olkin, 1985). Due to the diversity of the measures under study, a random effects model was used (Field & Gillett, 2010). The Comprehensive Meta-Analysis (CMA, version 3.3.070) software package was used to help aggregate statistics in various formats and to conduct the heterogeneity analyses. We used online calculators to generate effect sizes (Lenhard & Lenhard, 2016; Wilson, n.d.) when the variety of available statistics overwhelmed the CMA software.

We used R (a software environment for statistical computing and graphics) to produce high-quality funnel plots to detect publication bias. The visual symmetry of funnel plots intuitively suggests the absence of publication bias (Dual & Tweedie, 2000). Statistical tests—including Egger's regression test (Egger et al., 1997) (*regtest* function in R), Begg's rank correlation test (Begg & Mazumdar, 1994) (*ranktest* function in R), and Rosenthal's failsafe N test (*fsn* function in R) (Rosenthal, 1995)—were also conducted to quantitatively diagnose the asymmetry of the funnel plots. Nonsignificant regression and rank correlation tests indicate the absence of publication bias. The failsafe N test determines the number of missing studies averaging a z-value of zero that should be added to nullify the found overall effect size. Publication bias is unlikely to exist when the failsafe N exceeds the suggested threshold (the quintuple of pairwise comparisons plus 10) (Ellis, 2010). Further, we included a narrative synthesis of studies for dependent measures that could not be statistically pooled or that were threatened by significant heterogeneity.

Finally, a thematic analysis (Braun & Clarke, 2022) was conducted using all the studies in the sample to gain insights into possible reasons for the failure of productive peer talk moves in benefiting peer interactions and group outcomes (RQ3). Specifically, we first extracted all relevant text from included articles, then came up with codes that captured the main points and common meanings of the text, then created broader themes based on emergent patterns of the codes, and lastly refined and structured the themes to answer our research question.

Results

Study Features

Table 2 presents the major features of the included studies. They were fairly uniformly distributed by publication period, with eight published before 2000, seven published between 2000 and 2010, and nine published after 2010. Most of the studies had been conducted in the USA ($n = 10$) or Europe ($n = 12$; of which the Netherlands: $n = 5$, Germany: $n = 3$). Regarding the theoretical background, only six studies focused on collaborative problem-solving. Most ($n = 18$) adopted a collaborative learning perspective, and seven of these examined collaborative argumentation. The number of studies that investigated online peer interaction ($n = 14$, of which nine were asynchronous and five were synchronous) was slightly larger than the number that investigated face-to-face interaction ($n = 10$).

Table 2 Major characteristics of included studies

Study	Background	Grade	Sample size	Group size	Medium	Duration	Outcome	Hedges' g	Certainty of evidence
Avcı (2020)	CL	University	77	8 to 11	Online-A	8 weeks	Interaction quality	1.24	Conclusive
							Interaction intensity	−0.10	Inconclusive
Bouyias and Demetriadis (2012)	CA	University	34	2	Online-S	2 days	Domain-specific knowledge	1.02	Conclusive
							Domain-general knowledge	0.24	Inconclusive
Byun et al. (2014)	CPS	University	205	3	F2F	3 weeks	Interaction quality	0.67	Inconclusive
Gelmini-Hornsby et al. (2011)	CPS	Primary (lower)	36	2	F2F	Unclear	Interaction quality	2.79	Conclusive
							Solution quality	0.59	Conclusive
Gogoulou et al. (2008)	CL	University	80	2 or 3	Online-S	16.5 h	Interaction coherence	2.12	Inconclusive
Guzdial and Turns (2000)	CL	University	1300	> 6	Online-A	2 years	Interaction intensity	0.18	Inconclusive
King (1990) (1)	CL	University	26	2 or 3	F2F	Approx. 5 h	Interaction quality	1.52	Conclusive
							Domain-specific knowledge	1.92	Conclusive
King (1990) (2)	CL	University	39	3	F2F	Unclear	Interaction quality	1.01	Conclusive
							Domain-specific knowledge	2.19	Conclusive
King (1991)	CPS	Primary (upper)	46	2	F2F	3 weeks	Interaction quality	0.93	Conclusive
							Solution quality	1.50	Conclusive
							Problem-solving ability	1.54	Conclusive

Table 2 (continued)

Study	Background	Grade	Sample size	Group size	Medium	Duration	Outcome	Hedges' g	Certainty of evidence
King (1994)	CL	Primary (upper)	58	2	F2F	3 weeks	Interaction quality	2.06	Conclusive
							Interaction quality (transfer)	0.90	Conclusive
							Domain-specific knowledge	1.05	Inconclusive
							Knowledge retention (domain-specific)	2.03	Inconclusive
							Knowledge transfer (domain-specific)	0.74	Inconclusive
Kirschner et al. (2008)	CPS	University	147	3	F2F & Online-A	unclear	Interaction quality	1.63	Conclusive
Kollar et al. (2007)	CA	Secondary	90	2	F2F	2 h	Domain-specific knowledge	−0.10	Inconclusive
							Domain-general knowledge	0.62	Conclusive
Noroozi et al., (2013a, 2013b)	CA	University	120	2	Online-A	3.5 h	Solution quality	1.25	Conclusive
Noroozi et al., (2013a, 2013b)	CA	University	60	2	Online-A	3.5 h	Interaction quality	1.26	Conclusive
							Domain-specific knowledge	0.78	Conclusive
							Domain-general knowledge	1.18	Conclusive
							Interaction intensity	0.79	Conclusive
Popov et al. (2019)	CL	University	74	2	Online-S	3 days	Interaction quality	0.34	Inconclusive
							Solution quality	−0.08	Inconclusive

Table 2 (continued)

Study	Background	Grade	Sample size	Group size	Medium	Duration	Outcome	Hedges' g	Certainty of evidence
Saab et al. (2007)	CL	Secondary	76	2	Online-S	1.5 h	Interaction quality	0.72	Conclusive
							Participation equality	0.42	Inconclusive
Stegmann et al. (2007)	CA	University	120	3	Online-A	2.75 h	Interaction quality	1.37	Inconclusive
							Domain-specific knowledge	0.63	Inconclusive
							Domain-general knowledge	1.13	Conclusive
Stegmann et al. (2012)	CA	University	48	3	Online-A	3 h	Interaction quality	1.23	Conclusive
							Domain-specific knowledge	0.58	Inconclusive
							Domain-general knowledge	2.77	Conclusive
Mcmanus and Aiken (1996)	CL	University	27	3	Online-A	10 weeks	Domain-specific knowledge	NULL	
Baker and Lund (1997)	CPS	Secondary	16	2	Online-S	3 h	Solution quality & interaction quality	NULL	
Palincsar et al. (1993)	CPS	Primary (upper)	around 234	4	F2F	2 years	Domain-specific knowledge	NULL	
Soller (2001)	CL	Company staff	8	2 or 3	Online-S	1–1.5 h	Interaction quality	NULL	
Webb et al. (2014)	CL	Primary (lower)	111	dyad	F2F	multiple days	Domain-specific knowledge	NULL	
Webb (1982)	CL	Secondary	77	4	F2F	2 weeks	Domain-specific knowledge	NULL	
Webb et al. (1986)	CL	Primary (upper) and secondary	30	2	F2F	4–5 h	Domain-specific knowledge	NULL	

CL collaborative learning; *CPS* collaborative problem-solving; *CA* collaborative argumentation; *Online-A/S* online asynchronous/synchronous; *F2F* face-to-face

Regarding the intervention design, half of the reviewed studies focused on university undergraduates or graduates (52%), followed by primary school students (24%) and secondary school students (20%). Two studies tried to teach lower primary school students (approximately 7 years old) how to talk with peers (Gelmini-Hornsby et al., 2011; Webb et al., 2014). The included studies typically grouped peers in dyads (56%) or triads (30%). The number of participants in each study varied, with a median of 74, a mean of 126, a standard deviation of 251, and a range from 8 to 1,300.

Most of the interventions (46%) only lasted for 1–5 h. These short interventions typically contained brief learning and practice sessions introducing the talk scaffolds and a follow-up formal discussion session. A few other studies involved interventions lasting for 2–3 days (13%) or 2–3 weeks (17%). These longer interventions typically included a teacher training session, a pilot session to help adapt materials, a pre-test session for controlling group composition, or multiple discussion sessions. There were also a few long-term interventions, which occurred over 2–3 months (8%, $n=2$) (Avcı, 2020; Mcmanus & Aiken, 1996) or around 2 years (8%, $n=2$) (Guzdial & Turns, 2000; Palincsar et al., 1993). These interventions usually contained multiple design cycles or a set of experiments and numerous observations of group discussions.

Most of the studies (10 out of 13) that measured interaction quality provided conclusive evidence. Three studies were rated as inconclusive due to the reported low intercoder reliability (Byun et al., 2014, Cronbach's $\alpha=0.632$; Stegmann et al., 2007, Cohen's κ ranged from 0.50 to 0.69) and insufficient demonstration of the intervention effect (Popov et al., 2019, the effects are not statistically significant concerning discourse of planning, contributing, seeking input, and reflection/monitoring). Pertaining to the impact on knowledge learning, most evidence (4 out of 5 studies) on domain-general knowledge was conclusive with only one study rated as inconclusive due to its statistically insignificant result (Bouyias & Demetriadis, 2012). In contrast, only half of the studies (4 out of 8) measuring domain-specific knowledge provided conclusive evidence. The other half were classified as providing inconclusive evidence due to the reported low inter-rater reliability (King, 1994, Cronbach's α ranged from 0.52 to 0.78; Kollar et al., 2007, Cronbach's α ranged between 0.53 and 0.66) and statistically insignificant results (Stegmann et al., 2007, 2012). Most evidence (3 out of 4 studies) pertaining to the impact on solution quality was rated as conclusive with one study rated as inconclusive due to its statistically insignificant result (Popov et al., 2019). In brief, except for domain-specific knowledge, the evidence for other target outcomes was mostly conclusive.

RQ1: Frequently Used Productive Peer talk Moves

To summarize a list of frequently used productive peer talk moves in literature (RQ1), we grouped all of the peer talk scaffolds, variously worded in the sampled studies, into 28 unique productive peer talk moves. To simplify the frequency distribution patterns of these talk moves, we first categorized some talk moves as one type. For example, "disagree" ($n=14$) and "agree" ($n=9$) are specific evaluative positions. Therefore, we integrated them into the broader peer talk move category "evaluate" ($n = 7$). Similarly, we integrated "self-reflect" ($n = 10$) and "group reflect" ($n = 3$) as a broader talk move "reflect". Second, we combined several talk moves because they formed a conversational exchange and thus shared an essential socio-cognitive component. For example, "press for reasoning" and "explain oneself" were often expressed in a question–response format, and both highlighted the core component "explain". We integrated such dual forms of talk moves into single categories to highlight the major socio-cognitive operations identified in the studies.

After all the transformations, "evaluate" emerged as the most frequently promoted socio-cognitive component in productive peer interaction, followed by other frequently mentioned components including "express new idea", "explain", "elaborate", "reflect" and "share information". Table 3 presents the details of peer talk moves organized around these core socio-cognitive components. For example, the most frequently suggested component, "evaluate", occurred in the form of talk moves such as "disagree/agree", "evaluate", and "invite evaluation". These talk moves not only require students to engage in high-level cognitive thinking but also promote transactive discussion. The invitational forms of such socio-cognitive components help scaffold students' agency in eliciting expected cognitive activities in the group and promote group cohesion by involving all the group members.

RQ2: Effect Sizes of Productive Peer Talk Moves

We extracted 39 pairwise comparisons from the 17 studies included in the meta-analysis, because some of the articles measured multiple outcomes of productive talk moves. The total number of participants for the meta-analysis was 2636, with a range from 26 to 1300 and a median of 75 per comparison. The most frequently examined outcome was interaction quality ($k = 13$), followed by the acquisition of domain-specific knowledge ($k = 8$), the acquisition of domain-general knowledge ($k = 5$), and solution quality ($k = 4$). The very small number of original comparisons limited the reliability of the aggregate effect sizes for some of the outcome constructs in this review.

Publication Bias Analysis

The funnel plot for the interaction quality measures ($k = 13$) appeared to be visually symmetrical (see Fig. 2a). Furthermore, neither the regression test ($t = 0.28$, $df = 11$, $p = 0.79$) nor the rank correlation test (Kendall's $\tau = 0.21$, $p = 0.37$) showed significant asymmetry. The fail-safe N test showed that 616 missing publications would have been needed to make the significant combined effect size statistically nonsignificant; it is improbable that so many studies would have gone undetected. The fail-safe number also exceeded $5 \times k + 10$ (i.e., $5 \times 13 + 10 = 75 < 616$). The overall effect size for interaction quality was therefore robust and unlikely to have been inflated by a high level of publication bias.

Table 3 Peer talk moves of high frequency in included studies

Core socio-cognitive component (Relative frequency)[a]	Description	Talk move	Example of conversational form
Evaluate (20%)	Judge or comment on existing viewpoints by expressing agreement or disagreement or providing critical feedback	1 Disagree	You are wrong
			I do not agree with the following aspect of your position…
		2 Agree	I agree with you
			I think you are right to ask this question
		3 Evaluate	The limitation of your claim is …
			I think that your point …, but…
		4 Invite evaluation	Do you agree or disagree with this statement …? Support your answer
			Do you agree?
Express new idea (14%)	Put forward new opinions on how to solve group tasks, such as new strategies or solutions	5 Express new idea	I think…
			My opinion is…
		6 Invite expression	Are there any alternatives?
			What do you think?
Explain (14%)	Explain or justify the reasoning process behind a viewpoint or solution	7 Explain oneself	Here's my reasoning against your claim…
			Let me explain it this way…
		8 Press for reasoning	Why do you think of it as a problem?
			Explain how your proposed solution would work
Elaborate (13%)	Clarify a viewpoint through illustration, giving examples or adding more details	9 Elaborate oneself	I could show you…
			I could give you an example…
		10 Press for elaboration	What is a new example of …?
			I do not understand the following aspect of your position… Could you please elaborate on…
			Can you tell me more?

Table 3 (continued)

Core socio-cognitive component (Relative frequency)[a]	Description	Talk move	Example of conversational form
Reflect (9%)	Take a new look at or introspective view of one's own behavior and mental states	11 Self-reflect	I need to understand…
			I am not sure about what I am saying
		12 Group-reflect	Our understandings of the handout are still different!
			We only have five minutes left
			… didn't speak too much just now
		13 Invite reflection	What would we do differently next time?
			Are we getting closer to our goal?
Share information (7%)	Share relevant information such as conditions, personal experience, or authoritative knowledge that may benefit the group	14 Share information	Here is more information…
			I have done similar tasks before…
			We could try the method we learned before…
		15 Invite information	What is the problem to be solved in this task?
			What information do we have?
Compare (4%)	Identify the similarities or differences between two or more related viewpoints according to a certain criterion	16 Compare	What you said was the same as what I mean
			You two have given the same answer, but I think … is simpler
			What you just said is not consistent with what you said before
		17 Invite comparison	Please indicate the differences
			How are … and … similar?

Table 3 (continued)

Core socio-cognitive component (Relative frequency)[a]	Description	Talk move	Example of conversational form
Summarize (4%)	Draw conclusions about the group discussion, such as establishing the link between various viewpoints and identifying core points, conclusions, or implications	18 Summarize	I find that there are some patterns in our solutions
			Our viewpoints could be further integrated by…
			To adjust/combine our solutions, I suggest that…
Propose (4%)	Put forward collective action plans that may benefit problem-solving or help coordinate group work	19 Propose	No hurry! Let's take our time
			Let's move on to the next question
		20 Invite proposal	Which task should we start with?
			What should we do next?
Add on (4%)	Develop or extend existing viewpoints to form a new viewpoint	21 Add on	Next, we could…
			Here's a further thought offered in the spirit of your position …
Encourage (3%)	Compliment others on their good performance or encourage others to participate by allaying their concerns or fears	22 Encourage	Good point
			There you go. See, you can do it!
			Right or wrong, just say whatever you think!
Speculate (3%)	Conduct thought experiments or express possibilities based on existing information, viewpoints, or activities	23 Invite speculation	If it were you, what would you do?
			What would happen if …?
Revoice (2%)	Paraphrase others' viewpoints to verify whether one understands them correctly or highlight the importance of a viewpoint to draw group attention	24 Revoice	Do you mean…?
			Here is what I understand your thoughts are…

[a]Relative frequency of one socio-cognitive component refers to its occurrences in included studies divided by the total occurrences of all socio-cognitive components in this table

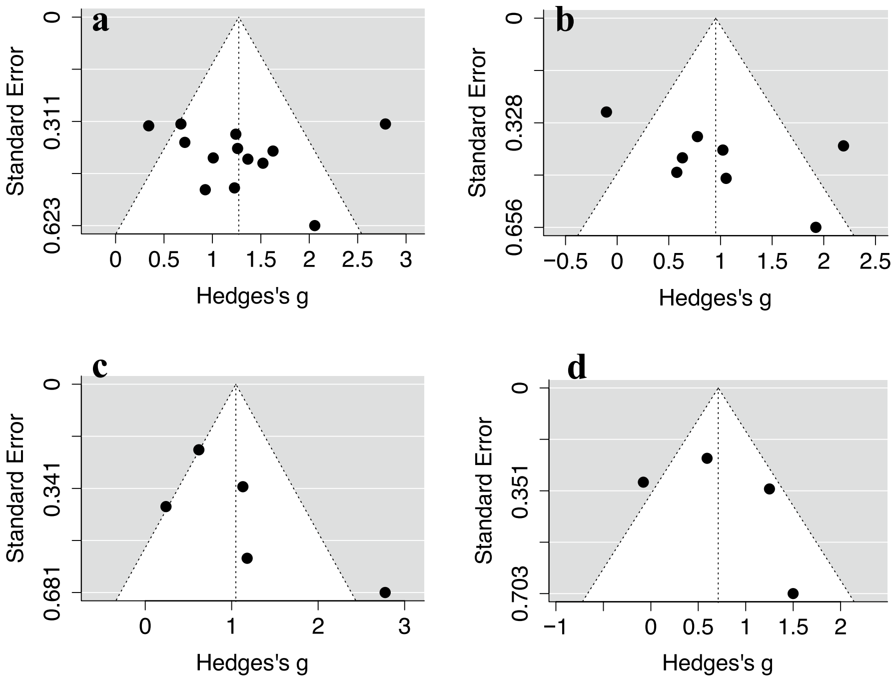

Fig. 2 Funnel plots for outcome measures. **a** Interaction quality; **b** Domain-specific knowledge; **c** Domain-general knowledge; and **d** Solution quality

In terms of publication bias in the effect sizes for domain-specific knowledge ($k=8$), the funnel plot appeared to be symmetrical (see Fig. 2b). The regression test ($t=1.77$, $df=6$, $p=0.13$) and the rank correlation test (Kendall's $\tau=0.21$, $p=0.55$) both indicated no threat of publication bias. The fail-safe number ($N=104$) also exceeded the critical value of 45 (i.e., $5\times8+10$), indicating the absence of threatening levels of publication bias. Therefore, publication bias did not significantly undermine the validity of the average effect size of domain-specific knowledge.

It was difficult to evaluate the symmetry of the funnel plots for domain-general knowledge ($k=5$) due to the limited number of comparisons (see Fig. 2c). Both the regression test ($t=1.50$, $df=3$, $p=0.23$) and the rank correlation test (Kendall's $\tau=0.40$, $p=0.48$) suggested the absence of significant publication bias. The fail-safe number ($N=58$) was also above the critical value (i.e., $5\times5+10$). Therefore, there did not appear to be any significant publication bias undermining the reliability of the aggregate effect size of domain-general knowledge.

The symmetry of the funnel plot for measuring solution quality ($k=4$) was also difficult to evaluate (see Fig. 2d). Neither the regression test ($t=0.66$, $df=2$, $p=0.58$) nor the rank correlation (Kendall's $\tau=0.33$, $p=0.75$) indicated significant publication bias. However, the fail-safe number ($N=20$) was below the critical value (i.e., $5\times4+10$). Given these inconsistent results, publication bias may have been a problem. Therefore, caution should be exercised when interpreting the aggregate effect size of solution quality.

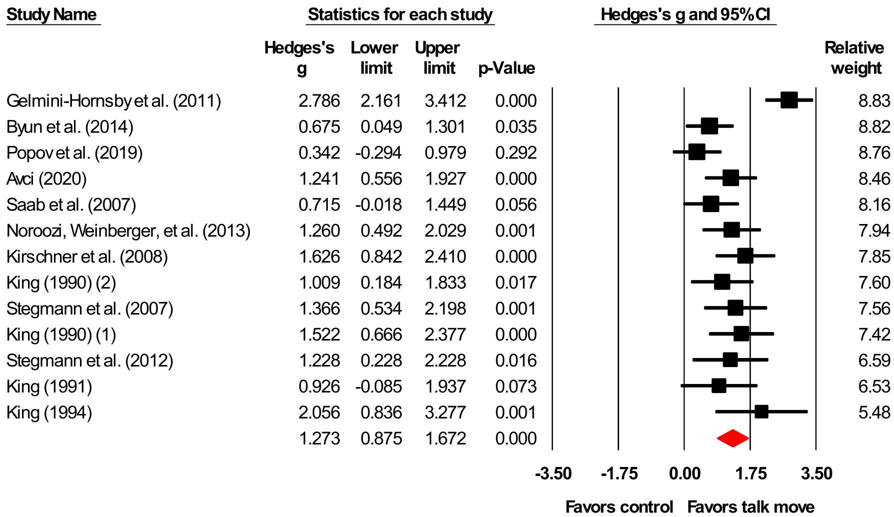

Fig. 3 Forest plot of the effect sizes of productive talk moves on interaction quality

Impact of Productive Talk Moves on Interaction Quality

The overall estimate showed that providing students with productive peer talk moves had a significant strong average effect on interaction quality (Hedges' $g = 1.27$, $p < .001$, 95% CI [0.88, 1.67]) (see Fig. 3). All the 13 comparisons reported positive effects of productive talk moves on peer interaction. However, there was significant heterogeneity among these comparisons (Q (12) = 40.00, $I^2 = 70.00$, $p < .001$). Therefore, the aggregate effect should be interpreted with caution.

As shown in Fig. 3, Gelmini-Hornsby et al. (2011) reported the strongest effect, and their results drove the high level of heterogeneity (Hedges' $g = 2.79$). The researchers provided lower-level primary school students (6–7 years old) with question prompts to guide their collaborative storytelling and found that these students asked significantly more thinking questions (e.g., how does … feel?, what does… think?, what does … want? and why?) than the controls did. King (1994) also reported a very strong effect on interaction quality (Hedges' $g = 2.06$). She compared the dyadic interactions of students who had been taught to question and explain with those of students who had only been taught to explain and found that the two groups asked totally different types of questions. Dyads with specific question prompts asked substantially more integration questions (e.g., explain why…, how are … and … similar?, how does … affect …?, and how does… tie in with... that we learned before?), whereas those without the questioning support tended to ask more factual questions.

Popov et al. (2019) reported the smallest effect size on interaction quality (Hedges' $g = 0.34$). They designed an interculturally enriched collaboration script to foster online cross-culture collaboration among university dyads. They also provided the scripted group with interaction prompts and sentence openers (e.g.,

Study Name	Hedges's g	Lower limit	Upper limit	p-Value	Relative weight
Kollar et al. (2007)	-0.104	-0.681	0.473	0.724	14.99
Noroozi, Weinberger, et al. (2013)	0.777	0.053	1.501	0.035	13.74
King (1990) (2)	2.191	1.405	2.976	0.000	13.21
Bouyias and Demetriadis (2012)	1.023	0.212	1.835	0.013	12.98
Stegmann et al. (2007)	0.632	-0.227	1.491	0.149	12.57
Stegmann et al. (2012)	0.577	-0.371	1.525	0.233	11.81
King (1994)	1.055	0.069	2.040	0.036	11.50
King (1990) (1)	1.923	0.636	3.209	0.003	9.20
	0.959	0.398	1.520	0.001	

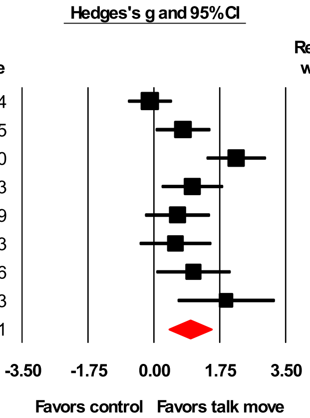

Fig. 4 Forest plot of the effect sizes of productive talk moves on domain-specific knowledge

I see what you mean, but…, considering… we may note that…, and to sum up…) to scaffold their communication. The results indicated that dyads with such support produced more contributing utterances (e.g., feedback giving, challenge, and explain) than the controls did, but they did not outperform the control group in the reflection/monitoring category.

Impact of Productive Talk Moves on Domain-specific Knowledge

The average effect on domain-specific knowledge was substantial and significant (Hedges' $g = 0.96$, $p < .01$, 95% CI [0.40, 1.52]) (see Fig. 4). However, due to the significant heterogeneity among these comparisons (Q (7) = 25.28, $I^2 = 72.31$, $p < .01$), the reliability of this aggregate effect is low.

As indicated by the forest plot (Fig. 4), only Kollar et al. (2007) reported a negative effect of productive talk moves on domain-specific knowledge. They compared how well two levels of external collaborative scripts (high vs. low structured) supported dyadic collaborative argumentation. The low structured condition only asked students to discuss two hypotheses about the phenomenon of frog deformities, whereas the high structured condition also introduced the core components of a sound argument and how to sustain collaborative argumentation following an argument sequence. The researchers also provided the high structured group with sentence openers to elicit high-quality arguments (e.g., it was found that …). However, the results showed that such high-level external collaborative scripts including argumentation principles and sentence openers weakened the students' learning of the scientific method, compared with the control group. Likewise, Stegmann's studies (Stegmann et al., 2007, 2012) examined the impact of argumentative scripts on domain-specific knowledge learning and found consistent nonsignificant relationships.

An early study by King (1990) strongly influenced the large aggregate effect on domain-specific knowledge in the meta-analysis and largely explained the heterogeneity of the results. In the first experiment (i.e., King, 1990 [1]), she compared

Study Name	Hedges's g	Lower limit	Upper limit	p-Value	Relative weight
Kollar et al. (2007)	0.620	0.200	1.040	0.004	27.71
Stegmann et al. (2007)	1.129	0.473	1.786	0.001	23.30
Bouyias and Demetriadis (2012)	0.239	-0.545	1.024	0.550	20.90
Noroozi, Weinberger, et al. (2013)	1.179	0.063	2.295	0.038	15.45
Stegmann et al. (2012)	2.775	1.440	4.110	0.000	12.64
	1.018	0.407	1.629	0.001	

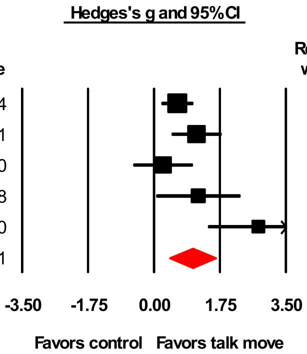

Fig. 5 Forest plot of the effect sizes of productive talk moves on domain-general knowledge

the domain comprehension of university students who had learned and practiced guided reciprocal peer questioning strategies and their peers who had not received this training. The results indicated a large positive effect of the training on domain comprehension (Hedges' $g = 1.92$). In the second experiment, King (1990 [2]) compared the comprehension of dyads supported by guided reciprocal peer questioning strategies with that of dyads who had been introduced to abstract questioning strategies but not given specific question stems (e.g., what is a new example of …?, explain why …, and what conclusions can you draw about …?). The result indicated a strong effect of the question stems on domain comprehension (Hedges' $g = 2.19$).

Impact of Productive Talk Moves on Domain-general Knowledge

The aggregate effect on domain-general knowledge based on the five available comparisons was significant and large (Hedges' $g = 1.02$, $p < .01$, 95% CI [0.41, 1.63]) (see Fig. 5). The heterogeneity analysis indicated significant variations among these comparisons (Q (4) = 12.46, $I^2 = 67.89$, $p < .05$). Therefore, the reliability of the calculated average effect on domain-general knowledge was limited, due to the small number of comparisons and significant heterogeneity.

As indicated in Fig. 5, all the comparisons indicated that productive talk moves had a positive influence on domain-general knowledge. Stegmann et al. (2012) reported the largest effect, which also drove the heterogeneity in these results (Hedges' $g = 2.78$). They found that university triads given collaboration scripts on argumentation (i.e., claim, ground, and qualification) showed substantially better knowledge of argumentation than triads without such scripts. The researchers found a similarly strong effect of these collaboration scripts on argumentative knowledge in another study (Stegmann et al., 2007) (Hedges' $g = 1.13$). There was a close correspondence between the argumentative scripts and the posttest measuring argumentative knowledge in these two studies. For example, the posttest required students to recall core components of an argument and construct convincing arguments, skills that had been taught to or practiced by the intervention group but not the control group.

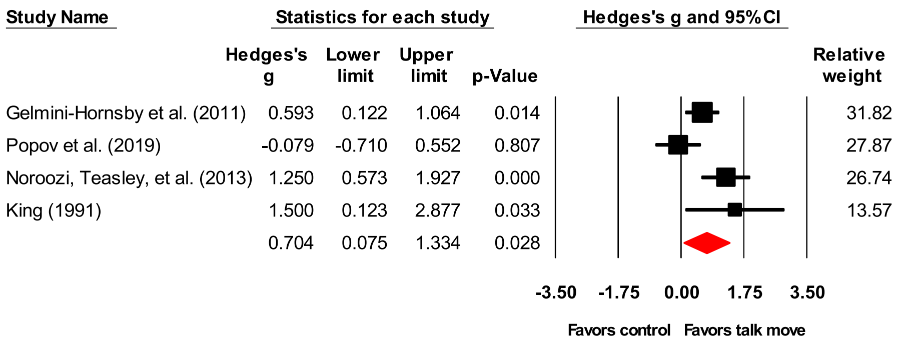

Fig. 6 Forest plot of the effect sizes of productive talk moves on solution quality

Bouyias and Demetriadis (2012) also investigated the impact of argumentative scripts on the learning of argumentative knowledge in a sample of university dyads. However, they only found a small and nonsignificant effect (Hedges' $g = 0.24$). Their posttest of argumentative knowledge included argumentation theory, recognition of constituent parts of an argument, and construction of convincing arguments, a more complex set of skills than the one used by Stegmann et al., (2007, 2012). Bouyias and Demetriadis (2012) found that students with argumentative scripts outperformed controls only in terms of argumentation construction ($p = 0.033$, Cohen's $d = 1.16$).

Impact of Productive Talk Moves on Solution Quality

The aggregate effect of productive talk moves on solution quality was medium-sized and significant (Hedges' $g = 0.70$, $p < .05$, 95% CI [0.08, 1.33]) (see Fig. 6). The heterogeneity analysis revealed significant variation in the reviewed studies ($Q(3) = 9.62$, $I^2 = 68.80$, $p < .05$). The limited number of comparisons may affect the reliability of the aggregate effect.

Among the four comparisons, only Popov et al. (2019) reported a negative effect on solution quality (Hedges' $g = -0.08$). They did not find significant differences in the quality of group outputs between university dyads with interculturally enriched collaboration scripts and those without.

Some studies investigated the efficacy of productive peer talk moves using other outcome measures. We do not report the aggregate effect sizes for these outcomes due to the limited number of these comparisons. For example, some studies (Avcı, 2020; Guzdial & Turns, 2000; Noroozi et al., 2013a, 2013b) investigated whether productive peer talk moves could improve interaction intensity, but only Noroozi et al., (2013a, 2013b) reported a significant effect (reporting that groups with transactive discussion scripts authored more messages than unscripted groups did) (Hedges' $g = 0.79$, $p < .05$). Saab et al. (2007) investigated whether providing talk moves promoted participation equality but failed to detect a significant effect. King (1994) investigated whether the effects of productive peer talk moves on domain-specific knowledge acquisition were retained and transferred in a population of

Table 4 Moderators of interaction quality and domain-specific knowledge

Outcome	Possible moderator	Subgroup	k	Hedges' g	95% CI	p
Interaction quality	Grade	Primary	3	1.97	[0.79, 3.15]	< .01
		Secondary	1	0.72	[−0.02, 1.45]	0.056
		University	9	1.09	[0.79, 1.38]	< .001
	Communication medium	F2F	7	1.51	[0.87, 2.16]	< .001
		Online	6	0.97	[0.62, 1.33]	< .001
	Group size	2	7	1.36	[0.64, 2.09]	< .001
		3	5	1.12	[0.77, 1.47]	< .001
		> 6	1	1.24	[0.56, 1.93]	< .001
Domain-specific knowledge	Grade	Primary	1	1.06	[0.07, 2.04]	< .05
		Secondary	1	-0.10	[−0.68, 0.47]	0.72
		University	6	1.16	[0.60, 1.71]	< .001
	Communication medium	F2F	4	1.23	[−0.001, 2.45]	0.05
		Online	4	0.77	[0.36, 1.18]	< .001
	Group size	2	5	0.82	[0.18, 1.46]	< .05
		3	3	1.15	[0.07, 2.24]	< .05

k = number of pairwise comparisons; *F2F* = face to face

primary dyads. She found a significant retention effect but no transfer effect. In addition, King (1994) suggested that the interaction quality of the intervention group in the transfer task was higher than the control group but lower than their own performance in the posttest task where they were provided with additional prompts. King (1991) also provided evidence of the positive effect of productive peer talk moves on the problem-solving ability of primary school students.

Moderator Analysis

A moderator analysis was conducted to identify possible explanations for the high level of heterogeneity in the studies' results. Due to the limited number of available comparisons, we only examined three moderators (i.e., grade, communication medium, and group size), and two outcomes (i.e., interaction quality and domain-specific knowledge). Given the limited range of group size, we treated group size as a categorical variable (i.e., dyad, triad, and others).

The results revealed that grade (Q_{bet} (2) = 3.13, p = 0.21), communication medium (Q_{bet} (1) = 2.08, p = 0.15), and group size (Q_{bet} (2) = 0.39, p = 0.82) were not significant moderators of interaction quality. Neither group size (Q_{bet} (1) = 0.28, p = 0.60) nor communication medium (Q_{bet} (1) = 0.48, p = 0.49) was a significant moderator of domain-specific knowledge acquisition. Only grade (Q_{bet} (2) = 10.35, p < .01) was found to be a significant moderator of domain-specific knowledge acquisition. However, the primary school and secondary school subgroups only contained one comparison. The reliability of this moderator analysis was therefore strictly limited (see Table 4).

RQ3: Possible Explanations of the Malfunction of Productive Peer Talk Moves

Some studies failed to demonstrate the benefits of productive peer talk moves, especially for knowledge acquisition. The studies identified some major plausible explanations for these failures. The most frequently mentioned explanation was the limited duration of the interventions. Students needed time to become familiar with the productive peer talk moves, as these external scaffolds must be internalized to realize their full and sustained benefits (Kollar et al., 2007; Popov et al., 2019; Saab et al., 2007). In studies with short interventions, students needed to spend extra cognitive resources to use the productive talk moves (Baker & Lund, 1997), which might have limited the cognitive resources they allocated to domain-specific knowledge learning (Noroozi et al., 2013a, 2013b; Stegmann et al., 2007, 2012).

Second, the low-fidelity implementation of productive peer talk moves may also explain the failure to detect the expected outcomes (Popov et al., 2019; Webb, 1982). Students might ignore offered talk scaffolds, use them superficially, or even not know how or when to effectively adopt the scaffolds (King, 1994). For example, Stegmann et al. (2007) found that around 40% of the online argumentative scripts they gave to university triads were ignored or used in unintended ways.

Third, the outcome assessment methods used in some studies may not have accurately detected the impact of productive peer talk moves. For example, the assessment process may not have been sensitive to students' social or cognitive growth (Popov et al., 2019); alternatively, the tasks may have been too difficult to show student progress in knowledge acquisition (Bouyias & Demetriadis, 2012). Furthermore, immediate measures may not have captured the long-term positive influence of productive peer talk moves on peer collaboration competence (Stegmann et al., 2007).

Finally, some scholars referred to the over-scripting effect to explain their non-significant findings (Kollar et al., 2007). "Over-scripting" refers to imposing excessively structured and rigid scripts on learners, which undermine their agency and the natural interaction process (Dillenbourg, 2002). Furthermore, high-level external scripts were found to be detrimental to domain-specific knowledge acquisition for learners with high-level internal scripts (Kollar et al., 2007). Therefore, some studies recommended providing external scripts with adaptable levels of structuredness based on students' knowledge, communicative needs, and styles (Gogoulou et al., 2008; Kirschner et al., 2008; Kollar et al., 2007).

Discussion

RQ1: Frequently Used Productive Peer Talk Moves

Here, 24 frequently used productive peer talk moves and 13 core socio-cognitive components (evaluate, express new idea, explain, elaborate, reflect, share information, compare, summarize, propose, add on, encourage, speculate, and revoice) are identified and summarized from recent literature. The productive peer talk moves and socio-cognitive components are consistent with those identified in qualitative studies of effective collaborative learning activities (e.g., explaining, asking thought-provoking questions, elaborating, argumentation) (Gillies, 2019; King, 2007; Kobbe et al., 2007). The synthesis of empirical studies presented here extends the field by establishing a more concrete and complete repertoire of empirically grounded productive peer talk moves. It reveals that talk intervention programs have examined not only the frequently discussed high-level cognitive activities (e.g., explaining and evaluating) but also some less cognitively demanding activities (e.g., sharing information and revoicing) and socially beneficial activities (e.g., encouraging). In addition, an aggregate repertoire of peer talk moves, in order of frequency, is provided and may help the designers of future intervention programs in selecting and adjusting peer talk moves according to contextual needs.

The set of productive peer talk moves collected in this study is consistent with the list of essential talk principles that are the basis of various talk strategies, such as exploratory talk (e.g., justifying one's own ideas and engaging critically and constructively with each other's ideas) (Littleton & Mercer, 2013; Mercer et al., 1999), collaborative reasoning (e.g., trying to look at both sides of an issue, making sure everyone has a chance to participate, and responding to the idea rather than the person) (Clark et al., 2003; Reznitskaya et al., 2009), philosophical talk (e.g., listening to one another with respect, building on one another's ideas, challenging one another to supply reasons for otherwise unsupported opinions) (Gorard et al., 2015; Lipman, 2003; Topping & Trickey, 2013), and knowledge building talk (e.g., improvable ideas, idea diversity, epistemic agency, rise above) (Scardamalia, 2002; van Aalst, 2009). The list supports the validity of the extracted peer talk moves. Compared with abstract talk rules, a concrete and specific set of peer talk moves is easier to learn—especially for young students—because it recommends specific desired behaviors that are necessary for effective group outcomes (Cohen, 1994). Meanwhile, the formatted peer talk moves collected in this study are more flexible and concise than sentence openers or questioning frames and, thus, reduce the risk of overwhelming learners or limiting the autonomy of their voices (Gogoulou et al., 2008).

RQ2: Effect Sizes of Productive Peer Talk Moves

The meta-analysis reveals large-sized effects that productive peer talk moves have on interaction quality (Hedges' $g=1.27$), domain-specific knowledge acquisition (Hedges' $g=0.96$), and domain-general knowledge acquisition (Hedges' $g=1.02$). It also demonstrates a medium-to-large size of positive effect on solution quality (Hedges' $g=0.70$). As to methodological quality, all included studies adopted experimental designs. Evidence for all outcomes but domain-specific knowledge was mostly conclusive. Therefore, despite significant heterogeneity in all the measures, the meta-analysis suggests robust positive effects of productive peer talk moves on collaborative discourse, domain-general knowledge learning, and problem-solving.

The meta-analysis reveals that the selected provided talk moves have a very large effect on peer interaction quality, likely due to the close correspondence between the provided talk moves and the measures of interaction quality. That is, most of the reviewed studies evaluated the quality of collaborative discourse by coding and counting the occurrence of expected talk moves or move sequences (e.g., Gelmini-Hornsby et al., 2011; King, 1994; Stegmann et al., 2012). The measure of interaction quality could therefore be characterized as a treatment check rather than a measure of a process outcome. Some scholars have responded to this criticism by emphasizing that students may not use the scaffolds as intended. Thus, the quality of their overt collaborative discourse is not self-evident (Stegmann et al., 2007).

The work presented here suggests that it is vital to treat peer interactions as an essential outcome rather than merely a treatment check. However, measures of interaction quality should extend beyond the implementation (i.e., the occurrence of the encouraged talk moves or move sequences in collaborative discourse). In-depth examinations should be conducted to determine whether the elicited talk moves fit the interaction context (Gogoulou et al., 2008) and whether a shared understanding is built in the process (Kirschner et al., 2008). Interaction quality should also be viewed as a core outcome, according to Bakhtinian dialogic theory (1981), which views dialogue as an educational goal. Future studies may consider whether and how offered talk moves help peers to interact equitably and open-mindedly and whether new knowledge naturally emerges from this process. The above is in line with the current research on collaboration-as-learning (e.g., Enyedy & Stevents, 2014) and could offer insight into how offered talk moves affect group dynamics.

RQ3: Possible Explanations of the Malfunction of Productive Peer talk Moves

Here, we identify frequently mentioned explanations for the malfunction of productive peer talk moves—the most common being short intervention duration. Most of the sampled studies conducted short interventions (one week or less). In the meta-analysis of studies of computer-supported collaboration scripts, Radkowitsch et al. (2020) found a similar trend. The most common explanation of the impact of intervention duration were based on cognitive load theory (Sweller, 1988), which assumes that individuals have a limited cognitive capacity. Therefore, there is a theoretical tradeoff between domain-general and domain-specific knowledge learning (Noroozi et al., 2013a, 2013b). In addition, we speculate that presenting students with the productive peer talk moves may not guarantee the intended usage of them, especially for young students. It might be necessary to help students clarify the functions of various peer talk moves and demonstrate the specific usage in peer interaction (King, 1997), which requires a longer invention duration.

Some scholars attributed the failure of productive peer talk (i.e. lack of positive outcomes) to the over-scripting effect (Kollar et al., 2007). According to the script theory of guidance for computer-supported collaborative learning (Fischer et al., 2013), optimal external scripting should fit the highest hierarchical level of internal collaboration scripts and avoid repeating subordinate components that are already available to learners. Therefore, the timely fading or adjustment of the offered productive peer talk moves is necessary to reduce restraints on natural interaction and avoid overlaying scripts (Kollar et al., 2007). However, meta-analyses showed that over-scripting did not produce an aggregate effect on decreasing student motivation or on undermining domain learning (Radkowitsch et al., 2020; Vogel et al., 2017). Bouyias and Demetriadis (2012) also found that the fading of scripts had insignificant impact on domain learning or argumentation quality. However, Gogoulou et al. (2008) found that students preferred communicative acts to sentence openers as the former provided flexibility in characterizing their messages. Therefore, it remains unclear how the structure of talk scaffolds affects student motivations and group performance, or whether a highly structured scaffold would jeopardize student interaction and learning.

Implications, Limitations, and Suggestions for Future Research

This study offers three major conclusions or contributions. First, the analysis presented provides an ordered repertoire of productive peer talk moves extracted from a range of empirical studies. Second, our findings confirm and quantify the robust positive effects of productive peer talk moves on collaborative interaction and outcomes. Third, we identify several common explanations for the malfunction of productive peer talk moves in authentic peer talk intervention programs. To the best of our knowledge, this is the first review to quantitatively synthesize an ordered list of empirically grounded productive peer talk moves. Though language has been widely recognized as essential to learning (Bakhtin, 1981; Piaget, 1932), there have been very limited number of studies evaluating the effect of productive peer talk moves in collaborative learning. This review calls for more studies in the area of productive peer talk moves. The extracted productive peer talk move repertoire may provide a reference for future intervention programs on peer talk across primary schools, secondary schools and universities. The repertoire is also insightful for constructing coding frameworks in analyzing the productivity of peer talk. This study also provides the very first quantitative synthesis of studies showing the robust positive effects of productive peer talk moves across educational levels.

However, the study is limited in some ways. First, the extracted productive peer talk move repertoire is based on the frequency of isolated talk moves and does not capture the sequential structures. For example, several studies (Kollar et al., 2007; Stegmann et al., 2007, 2012) mentioned the sound argument sequence (e.g., argument→counterargument→integration), which is not captured in talk move repertoire presented here. Future studies could enrich this repertoire of single peer talk moves by adding productive move sequences. It is also noteworthy that very few of the sampled studies (Popov et al., 2019; Soller, 2001; Webb, 1982) investigated the temporal patterns of peer interaction, although such patterns may also have important effects on social interdependence and group outcomes (Chen et al., 2017; Csanadi et al., 2018; Kapur, 2011; Reimann, 2009). Therefore, this study also calls on future empirical studies to examine the temporality of productive peer talk moves.

Second, the small number and high level of heterogeneity of the studies used in the meta-analysis might influence the aggregate effect sizes. Although most of the comparisons in the meta-analysis reported positive effects, the high level of heterogeneity may indicate a lack of precision in the measurements. Thus, further meta-analyses are needed to validate the medium-to-large effect sizes reported in this study.

Third, the calculated aggregate effect size for solution quality could be affected by publication bias, which may increase false-positive results. This study only considers peer-reviewed journal articles to ensure the quality of included studies and does not consider grey literature such as conference proceedings, dissertations, book chapters, and government/business/academic reports. Future meta-analyses need to reassess the publication bias issue by including grey literature or more available primary studies in journals. It is also recommended to check the methodological rigor of included studies.

Fourth, the implementation of peer talk interventions may suffer from fidelity issues. Some included studies explicitly discussed the fidelity of their interventions, but most did not have such information. It is therefore challenging to evaluate the degree of infidelity and decide to include certain studies or not. In this meta-review, we did not set the implementation fidelity as one inclusion or exclusion criterion, which may make the reported aggregated effect sizes smaller than true values. Future primary studies are recommended to check and report their implementation fidelity.

The reasons for the failure of productive peer talk moves in some of the sampled studies may also have implications for future studies. First, most of the sampled studies had short intervention periods, which limited the manifestation of talk productivity. Future studies could consider using longer intervention periods (i.e., more than one week) to clarify the function and usage of various talk moves to students and allow students to become familiar with the scaffolds through practice and optimally internalize the productive peer talk moves.

Second, in some of the sampled studies, the offered productive peer talk moves were not used as intended, which reduced the efficacy of the interventions. Students who ignore productive peer talk moves or only superficially adopt them may lack motivation (Stegmann et al., 2007). Furthermore, overly structured talk scaffolds may reduce students' autonomy and motivation (Dillenbourg, 2002; Wise & Schwarz, 2017), although this is not supported by meta-analyses (Radkowitsch et al., 2020; Vogel, Wecker, & Kollar, 2017). Possible reasons for the low motivation of students should be considered in future intervention programs.

If students have insufficient prior knowledge or skills to implement the offered talk moves, they may also fail to adopt them (Ge & Land, 2004). For example, it may be challenging for primary school students to provide sound explanations. Therefore, they may need additional training in how to explain their ideas (King, 1994). However, very few of the reviewed intervention studies—especially those with short intervention periods—provided additional guidance for students. In addition, students may have difficulties in applying the offered talk moves in authentic peer interactions, even after they have acquired relevant knowledge. Students may forget to use the moves or use them superficially, especially in synchronous discussion, due to their unfamiliarity. Additionally, students may fail to select appropriate talk moves in response to the contributions of their peers. Even the completely appropriate usage of single talk moves may not generate the expected benefits if the group lacks interdependence (Hu, 2020). Therefore, further empirical research could consider increasing the teaching as well as the practice of the talk moves.

Finally, inappropriate assessments may fail to reveal the efficacy of productive talk moves. Future studies should ensure that the difficulty of the outcome assessments and learning tasks is comparable. Retention or transfer tests should be considered besides immediate posttests. In addition to targeting domain-specific knowledge, productive talk moves aim to improve domain-general learning (e.g., collaboration skills, argumentation strategies), which might generate long-term and transferable effects (Popov et al., 2019; Stegmann et al., 2007). The beneficial effects of peer interaction may also need time to be apparent (Asterhan & Schwarz, 2007; Howe et al., 2005). However, very few of the sampled studies (King, 1994) examined the retention or transfer effects of productive talk moves. Therefore, future empirical studies could examine these aspects as well.

Acknowledgements This work was supported by Hong Kong Research Grants Council, University Grants Committee (Grant No. 17605221).

Declarations

Conflict of interest The authors have no known conflict of interest to disclose. The submission has not been previously published and is not under consideration for publication elsewhere.

References

*References included in the review are marked with an asterisk.

*+References also included in the meta-analysis are marked with an extra plus sign.

Asterhan, C. S., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99*(3), 626.

*+Avcı, Ü. (2020). Examining the role of sentence openers, role assignment scaffolds and self-determination in collaborative knowledge building. *Educational Technology Research and Development, 68*(1), 109–135. https://doi.org/10.1007/s11423-019-09672-5

*Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13*(3), 175–193. https://doi.org/10.1046/j.1365-2729.1997.00019.x

Bakhtin, M. M. (1981). *The dialogic imagination: Four Essays* (Trans. by C. Emerson and M. Holquist). University of Texas Press.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088–1101. https://doi.org/10.2307/2533446

Blatchford, P., Kutnick, P., Baines, E., & Galton, M. (2003). Toward a social pedagogy of classroom group work. *International Journal of Educational Research, 39*(1–2), 153–172.

*+Bouyias, Y., & Demetriadis, S. (2012). Peer-monitoring vs. micro-script fading for enhancing knowledge acquisition when learning in computer-supported argumentation environments. *Computers and Education, 59*(2), 236–249. https://doi.org/10.1016/j.compedu.2012.01.001

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.

*+Byun, H., Lee, J., & Cerreto, F. A. (2014). Relative effects of three questioning strategies in ill-structured, small group problem solving. *Instructional Science, 42*(2), 229–250. https://doi.org/10.1007/s11251-013-9278-1

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments, 25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Clark, A. M., Anderson, R. C., Kuo, L. J., Kim, I. H., Archodidou, A., & Nguyen-Jahiel, K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review, 15*(2), 181–198. https://doi.org/10.1023/A:1023429215151

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35.

Cook, N. A. (2008). *Online discussion forums: A strategy for developing critical thinking in middle school students*. State University of New York at Buffalo.

Cooper, R. (2018). *Diagnosing the diagnostic and statistical manual of mental disorders*. Routledge.

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 13*(4), 419–438. https://doi.org/10.1007/s11412-018-9292-z

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In Kirschner, P. A. (Ed.), *Three worlds of CSCL. Can we support CSCL* (pp. 61–91). Open Universiteit Nederland.

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*(2), 455–463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Enyedy, N., & Stevens, R. (2014). Analyzing collaboration. In Sawyer, R. (Ed.), *The Cambridge handbook of the learning sciences* (Cambridge Handbooks in Psychology, pp. 191–212). Cambridge University Press. doi:https://doi.org/10.1017/CBO9781139519526.013.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665–694.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66.

Friend, M., & Cook, L. (1992). *Interactions: Collaboration skills for school professionals*. Longman Publishing Group.

Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology, Research and Development, 52*(2), 5.

*+Gelmini-Hornsby, G., Ainsworth, S., & O'Malley, C. (2011). Guided reciprocal questioning to support children's collaborative storytelling. *International Journal of Computer-Supported Collaborative Learning, 6*(4), 577–600. https://doi.org/10.1007/s11412-011-9129-5

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research, 39*(1–2), 35–49.

Gillies, R. M. (2019). Promoting academically productive student dialogue during collaborative learning. *International Journal of Educational Research, 97*(2019), 200–209. https://doi.org/10.1016/j.ijer.2017.07.014

Gillies, R. M., & Haynes, M. (2011). Increasing explanatory behaviour, problem-solving, and reasoning within classes using cooperative group work. *Instructional Science, 39*(3), 349–366. https://doi.org/10.1007/s11251-010-9130-9

*+Gogoulou, A., Gouli, E., & Grigoriadou, M. (2008). Adapting and personalizing the communication in a synchronous communication tool. *Journal of Computer Assisted Learning, 24*(3), 203–216. https://doi.org/10.1111/j.1365-2729.2007.00254.x

Gorard, S., Siddiqui, N., & Huat See, B. (2015). *Philosophy for children: Evaluaiton report and executive summary* (Vol. 1, Issue 2). London: Education Endowment Foundation. https://doi.org/10.13140/RG.2.1.5064.4729

Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Springer. https://doi.org/10.1007/978-94-007-2324-5

*+Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences, 9*(4), 437–469. https://doi.org/10.1207/S15327809JLS0904_3

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O., & Barrera, M. J. (2016). Developing a coding scheme for analysing

classroom dialogue across educational contexts. *Learning, Culture and Social Interaction, 9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001
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*(1), 67–93.
Howe, C., Tolmie, A., Anderson, A., & MacKenzie, M. (1992). Conceptual knowledge in physics: The role of group interaction in computer supported learning. *Learning and Instruction, 2*(3), 161–183.
Hu, L. (2020). When productive talk fails to generate good outcome: Temporality of peer talk in dialogic collaborative problem solving. In Gresalfi, M. & Horn, I. S. (Eds.), *The interdisciplinarity of the learning sciences, 14th International Conference of the Learning Sciences (ICLS)* (pp. 905-906). International Society of the Learning Sciences.
Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning, 6*(1), 39–56. https://doi.org/10.1007/s11412-011-9109-9
*+King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*(4), 664–687. https://doi.org/10.3102/00028312027004664
*+King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology, 83*(3), 307–317. https://doi.org/10.1037/0022-0663.83.3.307
*+King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338–368. https://doi.org/10.3102/00028312031002338
King, A. (1997). ASK to THINK-TEL WHY: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational Psychologist, 32*(4), 221–235.
King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In: Fischer, F., Kollar, I., Mandl, H., & Haake, J. (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp.13–37). Springer. https://doi.org/10.1007/978-0-387-36949-5.
King, A. (2008). Structuring peer interaction to promote higher-order thinking and complex learning in cooperating groups. In R. M. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (pp. 73–92). Springer.
*+Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*(2), 403–420. https://doi.org/10.1016/j.chb.2007.01.028
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(2), 211–224.
*+Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721. https://doi.org/10.1016/j.learninstruc.2007.09.021
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310
Lenhard, W. & Lenhard, A. (2016). *Calculation of effect sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach: Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329.
Lipman, M. (2003). *Thinking in education*. Cambridge University Press.
Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Routledge.
*Mcmanus, M. M., & Aiken, R. M. (1996). Teaching collaborative skills with a group leader computer tutor. *Education and Information Technologies, 1*(1), 75–96. https://doi.org/10.1007/BF00144337
Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*(1), 95–111.
Miyake, N., & Kirschner, P. (2014). The social and interactive dimensions of collaborative learning. In Sawyer, R. (Ed.), *The Cambridge Handbook of the Learning Sciences* (Cambridge Handbooks in Psychology, pp. 418–438). Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9781139519526.026.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Prisma Group. (2009). Reprint-preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Physical Therapy, 89*(9), 873–880.

Molenaar, I., Chiu, M. M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning, 6*(4), 601–624. https://doi.org/10.1007/s11412-011-9130-z
Molenaar, I., Sleegers, P., & van Boxtel, C. (2014). Metacognitive scaffolding during collaborative learning: A promising combination. *Metacognition and Learning, 9*(3), 309–332. https://doi.org/10.1007/s11409-014-9118-y
Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology, 8*(2), 181–192. https://doi.org/10.1002/ejsp.2420080204
*+Noroozi, O., Teasley, S. D., Biemans, H. J., Weinberger, A., & Mulder, M. (2013a). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning, 8*(2), 189–223.
*+Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013b). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers and Education, 61*(1), 59–76. https://doi.org/10.1016/j.compedu.2012.08.013
*Palincsar, A. S., Anderson, C., & David, Y. M. (1993). Pursuing scientific literacy in the middle grades through collaborative problem solving. *The Elementary School Journal, 93*(5), 643–658.
Piaget, J. (1932). *The moral judgment of the child*. Routledge and Kegan Paul.
*+Popov, V., Biemans, H. J. A., Fortuin, K. P. J., van Vliet, A. J. H., Erkens, G., Mulder, M., Jaspers, J., & Li, Y. (2019). Effects of an interculturally enriched collaboration script on student attitudes, behavior, and learning performance in a CSCL environment. *Learning, Culture and Social Interaction, 21*(February), 100–123. https://doi.org/10.1016/j.lcsi.2019.02.004
Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5–47.
Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239–257. https://doi.org/10.1007/s11412-009-9070-z
Reznitskaya, A., Kuo, L. J., Clark, A. M., Miller, B., Jadallah, M., Anderson, R. C., & Nguyen-Jahiel, K. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education, 39*(1), 29–48. https://doi.org/10.1080/03057640802701952
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183.
*+Saab, N., Van Joolingen, W. R., & Van Hout-Wolters, B. H. A. M. (2007). Supporting communication in a collaborative discovery learning environment: The effect of instruction. *Instructional Science, 35*(1), 73–98. https://doi.org/10.1007/s11251-006-9003-4
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal Education in a Knowledge Society, 97*, 67–98.
Sfard, A. (2020). Learning, discursive faultiness and dialogic engagement. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 89–99). Routledge.
Sigafoos, J., O'Reilly, M. F., Ledbetter-Cho, K., Lim, N., Lancioni, G. E., & Marschik, P. B. (2019). Addressing sequelae of developmental regression associated with developmental disabilities: A systematic review of behavioral and educational intervention studies. *Neuroscience & Biobehavioral Reviews, 96*, 56–71.
Sigafoos, J., Green, V. A., Schlosser, R., O'eilly, M. F., Lancioni, G. E., Rispoli, M., & Lang, R. (2009). Communication intervention in Rett syndrome: A systematic review. *Research in Autism Spectrum Disorders, 3*(2), 304–318.
*Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*(1), 40–62.
*+Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science, 40*(2), 297–323. https://doi.org/10.1007/s11251-011-9174-5
*+Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421–447. https://doi.org/10.1007/s11412-007-9028-y
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Teo, Y.-H., & Daniel, C. (2007). Using sentence openers to support students' argumentation in an online learning environment. *Educational Media International, 44*(3), 207–218. https://doi.org/10.1080/09523980701491658

Topping, K. J., & Trickey, S. (2013). The role of dialog in philosophy for children. *INternational Journal of Educational Research, 63*, 69–78. https://doi.org/10.1016/j.ijer.2013.01.002

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

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.

*Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74*(5), 642–655. https://doi.org/10.1037/0022-0663.74.5.642

*Webb, N. M., Ender, P., & Lewis, S. (1986). Problem-solving strategies and group processes in small groups learning computer programming. *American Educational Research Journal, 23*(2), 243–261.

*Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., & Turrou, A. C. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research, 63*, 79–93. https://doi.org/10.1016/j.ijer.2013.02.001

Wecker, C., & Fischer, F. (2010). Fading instructional scripts: preventing relapses into novice strategies by distributed monitoring. In Gomez, K., Lyons, L., & Radinsky, J. (Eds.), *Proceedings of the 9th international conference of the learning sciences: Learning in the disciplines* (pp. 794–801). Chicago, IL: ISLS.

Wilson, D. B. (n.d.). Practical meta-analysis effect size calculator. https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD3.php.

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423–467. https://doi.org/10.1007/s11412-017-9267-5