Thinking Skills and Creativity 42 (2021) 100964

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

Conceptualization and operationalization of group thinking sustainability in dialogic collaborative problem solving

Liru Hu

Faculty of Education, The University of Hong Kong, Pokfulam Road, Hong Kong, China

A R T I C L E I N F O

Keywords:
Collaborative problem solving
Peer talk
Group thinking
Sustainability
Talk move

A B S T R A C T

Peer talk shapes the trajectory of group thinking. Studies have explored productive peer talk moves that can facilitate high-order group thinking, yet few have focused on the extent to which students consecutively take up these talk moves to sustain group thinking. There is no consensus on how to understand or measure the sustainability of productive peer talk. This study establishes a construct to help characterize a group's capacity to consecutively engage in high-order collective thinking and to investigate the impact of such sustainability on group outcomes. The proposed construct, group thinking sustainability (GTS), was conceptualized as a three-level nested hierarchy (comprised of reciprocity, productivity, and constructiveness) and further operationalized as the average length of a corresponding overt turn-taking sequence in group discussions. This study applied this construct to a sample of 168 primary school students who were divided into groups of four and asked to collaboratively solve three mathematical problems within 30 minutes. The results revealed that GTS can help characterize and differentiate a group's capacity to sustain productive peer talk. GTS can also help predict group outcomes and explain why some groups were more successful than others. This study provides novel insights into understanding and measuring GTS across groups. It also suggests a three-level scaffolding (i.e., turn-taking, productive talk, and knowledge construction) that teachers can use to support sustainable group thinking in collaborative peer talk.

1. Introduction

Involving students in collaborative problem solving has the potential to produce social, emotional, and cognitive gains (e.g., Blatchford et al., 2003; Johnson & Johnson, 2016; Slavin et al., 2014). However, only certain kinds of well-structured peer talk that involve high-level cognitive processing can fulfill the potential benefits of collaborative problem solving (King, 2008). It has been well established that high-quality peer talk is often characterized by a set of fine-grained productive talk moves (e.g., Gillies, 2019; King, 2008; Webb & Mastergeorge, 2003; Webb et al., 2014).

There is growing awareness that sustained, effective peer interaction does not happen naturally but requires dedicated teaching efforts or various scaffolding techniques (Gillies & Haynes, 2011; Scheuer et al., 2010; van der Veen & van Oers, 2017), and much effort has been put into identifying productive peer talk moves to guide peer interaction and promote positive group outcomes (Asterhan & Schwarz, 2009; Gillies, 2019; King, 1997; Noroozi et al., 2013; Webb et al., 2014). A talk move is a stretch of talk that forms a unit and is functionally related to the conversation to which it belongs (Goffman, 1981). Most talk moves are verb–object and content-free phrases with local social or cognitive goals, such as *express new idea, explain oneself*, and *disagree with others*. Peer talk

E-mail address: liruhu@connect.hku.hk.

https://doi.org/10.1016/j.tsc.2021.100964
Received 25 November 2020; Received in revised form 14 September 2021; Accepted 10 October 2021
Available online 16 October 2021

moves refer to those that occur in conversation between equals and not by a facilitator. Productive peer talk moves are also described as questioning frames (e.g., Ge & Land, 2004; King, 2008), sentence openers (e.g., Gogoulou et al., 2008; Teo & Daniel, 2007), or micro collaboration scripts (e.g., Noroozi et al., 2013; Stegmann et al., 2012), such as *what do you think about…, why…, I think…, I agree/-disagree with…, can you say more about…*, and *do you agree or disagree with…*. These peer talk moves aim to perform or elicit collective reasoning and knowledge construction. Studies have established the benefits of these various talk moves in eliciting high-level cognitive processing (Bouyias & Demetriadis, 2012; Popov et al., 2019) and facilitating high-level social interdependence (Gelmi-ni-Hornsby et al., 2011; Kirschner et al., 2008), which further lead to better collaborative discourse, domain learning, and solution quality (King, 1994; Noroozi et al., 2013; Popov et al., 2019; Stegmann et al., 2012). However, it is still not clear to what extent students take up these moves in peer talk or what, if any, impact the sustaining of such talk moves has on group outcomes.

Some studies have made efforts to understand the sequentiality of high-quality peer talk, which underpins the persistent usage of productive talk moves. For example, Chi and Menekse (2015) claimed that students achieved the largest amount of learning when they engaged in co-constructive talk in which they elaborated, explained, and built on each other's ideas. High-performing groups have been shown to engage more often in co-constructive processes, rather than only in discursive information sharing or isolated contributions (Heo et al., 2010; Rojas-Drummond et al., 2006). It has also been found that students' reasoning ability improves when they engage in exploratory talk in which they *interthink* with each other (Mercer & Littleton, 2007), and that the progressiveness in whole-class collective inquiry typically emerges from a sequential social mechanism where students critically build on each other's ideas (Clarà, 2019). Stahl (2014) suggested that group work requires longer sequences of response where students construct a series of adjacency pairs to accomplish the group cognitive goal. Studies on knowledge building have emphasized supporting students' sustained engagement in idea improvement (e.g., Scardamalia & Bereiter, 2014). All of these studies have shown the importance of the sustained use of productive talk moves in achieving a better group outcome; however, there has been no consensus on how to describe such sustainability in peer talk or compare it across various groups.

To address this gap, this study proposes a construct called *group thinking sustainability* (GTS) to characterize the persistency of dialogic thinking-together (Webb, 2019) and to enable cross-group comparisons. The following sections introduce the conceptualization and operationalization of GTS and then apply GTS to authentic data to validate whether it could characterize the sustainability of productive peer talk and explain the variety of group outcomes.

2. Theoretical framework

2.1. Definition of dialogic collaborative problem solving

Mikhail Bakhtin (1895–1975), a representative and influential scholar of dialogism, claimed that there is no end to dialog; a genuine dialog always stimulates new voices: "If an answer does not give rise to a new question from itself, it falls out of the dialog" (Bakhtin, 1986, p. 168). Bakhtin (1999) expressed the view that there is no fixed and final knowledge or truth, but that truth emerges from unlimited dialog involving "*a plurality of [opaque, non-transparent] consciousnesses*, with equal rights and each with its own world, [that] combine but are not merged in the unity of the event" (p. 6, italics in original). He compared genuine dialog to a polyphony, which suggests some essential features of dialogic interaction: equality, multivocality, and interanimation (Kolikant & Pollack, 2020; Matusov et al., 2019; Wegerif, 2013, 2020).

The Organisation for Economic Co-operation & Development (2013) described collaborative problem solving as occurring when "two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (p. 6). According to Friend and Cook (1992, p. 5), "interpersonal collaboration is a style of direct interaction between at least two co-equal parties voluntarily engaged in shared decision making as they work toward a common goal." Therefore, we can say that collaborative interaction involves the interanimation of two or more independent and co-equal consciousnesses, which echoes what Bakhtin's dialogic interaction implied.

Referring to Bakhtin's dialogic framework (1999), the present study defines *dialogic collaborative problem solving* as a complex dynamic process whereby two or more consciousness, with equal rights and each with its own world, combine but are not merged in the unity of solving a shared problem. This definition acknowledges the value that being *dialogic* adds to collaborative problem solving on at least two levels. First, it emphasizes the role of dialog in collaborative problem solving. This is consistent with other theoretical perspectives that have emphasized the role of language in thinking (Piaget, 1932; Vygotsky, 1978). Second, dialogic collaborative problem solving emphasizes collaborators treating each other as equals, treating each other with respect, and engaging in internally persuasive discourse (Bakhtin, 1981; Wegerif, 2020).

2.2. Conceptualization of group thinking sustainability

Positive interdependence between group members has long been emphasized for successful collaborative problem solving (Brush, 1998; Swiecki, 2020; Wang, 2009). Such interdependence refers to the connections between group members in terms of contributions, objectives, rewards, and accountability (Brush, 1998; Wang, 2009). GTS, as constructed in the present study, aims to describe the persistence of interdependence in individual thinking within dialogic collaborative problem solving. It is a group-level feature characterizing the extent to which group members continuously think with each other around a common target.

This study considers that the interdependence of individual thinking in group talk could be shaped by the use of productive talk moves by group members. Greeno (2015) suggested that talk sequence implied humans' underlying cognitive structure, and some scholars have further claimed that discourse *is* cognition/thinking (e.g., Resnick et al., 1997; Sfard, 2015). Interpersonal dialog reveals a social mode of thinking and does not only serve as a stimulus for individual thought (Mercer & Littleton, 2007). At the same time, individual thinking is an individualized form of interpersonal dialog (Sfard, 2015). A large body of literature has attempted to identify the productive peer talk moves characterizing high-performing groups (e.g., Gillies, 2019; King, 2008; Webb et al., 2014). These productive talk moves involve or elicit high-level cognitive processing (King, 2008). In the present study, *thinking* in GTS also refers to high-level cognitive processing. GTS describes the extent to which a group persistently engages in high-level group cognition, which could be shaped by the extent to which students consecutively adopt these productive talk moves.

However, it is not necessary for students to adopt productive talk moves in consecutive turns to sustain collective thinking. Many studies have abstracted principles or features of peer talk that could generate productive learning outcomes, such as *exploratory talk* (Littleton & Mercer, 2013; Mercer & Littleton, 2007), *knowledge-building talk* (van Aalst, 2009), and *co-constructive dialog* (Chi & Menekse, 2015; Rojas-Drummond et al., 2006). In these macro discourse patterns, students adopt productive talk moves but not necessarily in consecutive turns.

To tackle this tension between GTS and the consecutive use of productive talk moves, this study also considers the reciprocity of peer interaction and adopts a hierarchical perspective to evaluate GTS. Specifically, the group thinking trajectory is first segmented based on the reciprocity of peer talk. GTS considers a thinking trajectory to be sustained if a talk sequence involves discursive reciprocal feedback, such as expressing listening, and to be broken if consecutive turns in the talk sequence are not reciprocal, even though these turns all involve productive talk moves. The number of productive talk moves in one thinking trajectory segment is then evaluated.

In addition to the emphasis on reciprocal and productive talk, another essential dimension of collective thinking processes is whether such talk constructs something new to the group's knowledge (Chi & Wylie, 2014; Chi et al., 2018; van Aalst, 2009). The present study thus also evaluates the constructiveness of a sustained group thinking trajectory to better understand the efficacy of group thinking.

Based on the above discussions, the present study has structured GTS as a three-level nested hierarchy consisting of reciprocity, productivity, and constructiveness (see Fig. 1). The model first requires the reciprocity of a talk sequence, and reciprocal group thinking sustainability (RGTS) denotes the extent to which a talk sequence is reciprocal in consecutive turns. The two further levels of productivity and constructiveness characterize the efficacy of sustained collective thinking; productive group thinking sustainability (PGTS) refers to the quality of a reciprocal talk sequence (that is, the degree to which a group engages in high-order thinking), and constructive group thinking sustainability (CGTS) denotes the cognitive outcomes of a reciprocal talk sequence (that is, the degree to which new contributions emerge). The following sections interpret these three levels in detail.

2.2.1. Reciprocity

Reciprocity emphasizes the interdependence of collective thinking. Reciprocity is an essential feature of human interaction (Blau, 1964) and forms the foundational structure of GTS, similar to the construct of *mutuality* proposed by Barron (2000). Mutuality describes the extent to which peer interactions are reciprocal and balanced. Barron (2000) differentiated various types of treatment for solution proposals to characterize the *transactiveness* of peer talk (Azmitia & Montgomery, 1993). He found that a high-performing group engaged in productive conflicts, transactional responses, and turn-taking norms. Reciprocal interaction has also been emphasized in many studies as an essential dimension of high-quality collaboration processes (King, 2008; Meier et al., 2007; Webb & Mastergeorge, 2003). It denotes a symmetrical interpersonal relationship that involves respectful turn-taking, productive conflict resolution, and equitable participation (Meier et al., 2007).

In the present study, sustained group thinking flow requires the constant responsiveness of group members to their peers' thinking; reciprocity describes such responsiveness between adjacent utterances. A turn is viewed as being reciprocal if it involves a response to the previous turn, such as providing an answer, asking a follow-up question, or offering evaluation or feedback.

2.2.2. Productivity

Productivity describes the quality of thinking, echoing a large number of studies that have abstracted features of productive peer interactions. Many scholars have concretized coarse-grained talk rules and principles into detailed questioning frames or sentence openers to better scaffold student co-constructive talk with peers (e.g., Gillies, 2019; King, 1997; Webb & Mastergeorge, 2003; Webb

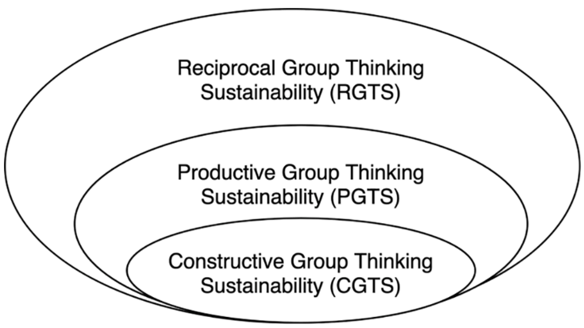


Fig. 1. The three-level nested hierarchy of group thinking sustainability (GTS).

et al., 2014). These scaffolding questions are similar to productive talk moves identified for teachers to use in dialogic teaching (Michaels & O'Connor, 2015). Hennessy et al. (2016) proposed a universally applicable coding scheme that defines productive talk moves for both teacher–student talk and peer talk. Despite differences across frames, some productive talk moves are common across various studies. The present study extracts several core sets of productive talk moves (see Table 1). These sets contain different specific talk moves that take various forms, such as being invitational, extending one's own thinking, or building on someone else's contributions. For example, the *explain* set includes pressing someone else for reasoning, explaining oneself, and explaining others' viewpoints.

2.2.3. Constructiveness

Constructiveness emphasizes the outcome of thinking; that is, whether the utterance contributes something new. Constructivism assumes that students should construct understandings of ideas in relation to prior knowledge and to the situation (von Glasersfeld, 1995). The construct should involve "qualitative changes in the complexity of students' thinking about and conceptualization of context-specific subject matter" (Moore, 2002, p. 27). Students in a constructive activity generate some new knowledge and inferences beyond what is presented in the materials, such as through note-taking, diagram drawing, self-explanation, etc. (Chi & Wylie, 2014; Chi et al., 2018). Chi and Menekse (2015) defined an utterance in peer interaction as constructive if it contained elaborations on one's own or one's partner's contributions. They further identified *co-constructive* dialog as the most productive pattern, where dyads were built on each other's ideas. Van Aalst (2009) suggested that knowledge-construction discourse involved high-level cognitive operations, such as interpretation, explanation, evaluation, synthesis, and creation.

In the present study, a constructive utterance is one that contains knowledge content that moves beyond current group knowledge, such as new ideas, new elaborations, and new inferences. It denotes the outcomes of a sustainable thinking flow and aids in understanding the connection between GTS and the group's final solution quality.

2.3. Operationalization of group thinking sustainability

2.3.1. Two methodological traditions

GTS is indicated by the persistence of interdependent peer talk; its operationalization depends on how the connection between individual utterances is quantified. Methodology on collaborative discourse follows either a socio-cognitive tradition or an interpretive

Table 1
Some core sets of productive peer talk moves.

Set	Interpretation	Examples
Express	Contributing to group knowledge, such as by proposing a new solution, viewpoint, or suggestion; or initiating task-related individual or collective actions.	1. I think the sign should be positive. 2. How should we use this condition? 3. What do you think? 4. Let's skip this question and go to the next one due to limited time.
Elaborate	Clarifying/extending one's own or another's contributions.	1. Beside, I also found this condition redundant. 2. One example of this is… 3. Can you say more on your plan?
Explain	Explaining/justifying one's own or another's contributions.	1. The reason I disagree is that… 2. Why do you think this plan cannot work? 3. Can you help me explain why she said the first choice was not correct?
Speculate	Speculating/hypothesizing/imagining different possibilities/theories based on one's own or another's contributions.	1. According to my knowledge, it could be A but also could be B. 2. So, if what she says is correct, then I wonder whether there are multiple solutions to this problem? 3. If this proposition was correct, was the second one true or false?
Evaluate	Judging/evaluating or stating (dis)agreement with another's or collective contributions.	1. You are wrong. 2. I think you are right with this question. 3. What do you think of my idea?
Revoice	Paraphrasing what another said to ensure engagement and shared understanding.	1. Do you mean we should add another triangle? 2. We should reverse this formula. Is this what you mean?
Reflect	Reflecting on one's own or collective learning processes/outcomes.	1. I did not perform well because I am not familiar with this topic. 2. We could find more solutions if allowed more time. 3. What do you think of our group's solution? Is it perfect?
Challenge	Challenging another's view/assumption/argument.	1. Do you really think these angles are the same? They may just look similar. 2. Did you forget this condition? Could your solution still be right with this?
Refer back	Building connections to previous contributions.	1. I remember we just talked about a similar idea… 2. She also mentioned this just now.

one (Dyke et al., 2012; Fu et al., 2016). The former usually adopts a quantitative content analysis approach, also called the frequency-based coding-and-counting approach (Chi, 1997). For example, Chi and Menekse (2015) coded paired utterances into five dialog patterns—passive-active, active-active, active-constructive, constructive-constructive, and co-constructive—and counted the number of different patterns and investigated how they were related to group outcomes. This coding-and-counting approach has been criticized for ignoring temporality and failing to detect temporal patterns of collaborative discourse (Csanadi et al., 2018; Kapur, 2011; Reimann, 2009).

In contrast, the interpretative tradition adopts a conversational analysis approach and emphasizes the contextual nature of collaborative discourse. For example, Fu et al. (2016) adopted narrative analysis to complement their coding-and-counting approach to extend the three online discourse modes proposed by van Aalst (2009) into nine detailed discourse patterns. However, this approach is time-consuming and limited in both its generalizability and applicability in cross-group comparisons (Dyke et al., 2012).

Reimann (2009) concluded that these two methodological traditions fit into two types of analysis technique: variable-based and event-based. Variable-based models assume that independent variables continuously act on dependent variables and thus cannot effectively capture the fragmented nature of most process variables. However, process variables may not be measurable at the same time due to the dynamism of interactions, which may violate the assumptions of variable-based methods. This is why the coding-and-counting approach usually fails to describe the temporality of peer talk. Although Chi and Menekse (2015) coded at the paired-utterance level, they still lost information on the connection among paired utterances.

The event-centered approach, in contrast, analyzes the world through "conceptualizing development and change processes as sequences of events which have unity and coherence over time" (Poole et al., 2000, p. 36), rather than as fixed entities with different attributes. It can generalize through pattern extraction and process modeling. This approach has drawn much attention due to increased interest in temporal analyses of peer talk (e.g., Chen et al., 2017; Csanadi et al., 2018). Furthermore, the event-centered approach can be integrated with the variable-based approach by extracting some event variables—i.e., quantitative aspects of events or event sequences, such as duration, intensity, periodicity, rate of change, or persistency (Reimann, 2009).

In the present study, GTS describes the extent to which a group can sustain collective thinking. It is an event variable that characterizes the continuity of high-level cognitive operations shaped by a coherent sequence of productive peer talk. A sustained flow of collective thinking involves more than simply providing feedback to the previous speaker. It emphasizes the involvement of productive peer talk moves and the construction of new understandings through reciprocal peer interactions. The present study analyzes the interdependence of peer talk at the turn level. GTS for a group is quantified as the average length of turn-taking sequences that should be first reciprocally connected and is then characterized by the degree of productivity and constructiveness.

2.3.2. Mathematical expressions of group-thinking sustainability

In this study, the impact of one turn on sustaining the flow of collective thinking is mathematically simplified as a sequence of three-dimensional vectors (see Table 2). The *reciprocal* dimension refers to whether the speaker holding the conversation floor responds to the last speaker. A reciprocal turn contains feedback relating to the previous speaker, such as giving an answer, asking a follow-up question, or building on previous contributions. The *productive* dimension refers to whether the turn contains productive peer talk moves, as shown in Table 1. Finally, the *constructive* dimension refers to whether the turn contributes new knowledge to the current discussion. It is noteworthy that a constructive turn is also productive (according to the coding protocol), but a productive or constructive turn could be non-reciprocal.

As such, a turn-taking sequence could be expressed as a sequence of three-dimensional vectors concerning reciprocity (R), productivity (P), and constructiveness (C): $\{(R_1, P_1, C_1), (R_2, P_2, C_2), (R_3, P_3, C_3), \ldots, (R_n, P_n, C_n)\}$. To operationalize reciprocal group thinking sustainability (RGTS), the corresponding turn-taking sequence could be expressed as a sequence of one-dimensional vectors on reciprocity: $\{R_1, R_2, R_3, \ldots, R_n\}$. RGTS could be quantified as the average length of a reciprocal turn-taking sequence.

For productive group thinking sustainability (PGTS), the turn-taking sequence could be represented as a sequence of two-dimensional vectors: $\{(R_1, P_1), (R_2, P_2), (R_3, P_3), \ldots, (R_n, P_n)\}$. PGTS could accordingly be quantified as the average number of turns containing productive talk moves in a reciprocal talk sequence.

To calculate constructive group thinking sustainability (CGTS), the turn-taking sequence could be characterized as a sequence of

Table 2
Coding protocol.

Dimension	1	0	NULL
Reciprocal	The turn contains feedback to the last speaker including simple feedback, such as brief acknowledgement and repetition, and productive feedback, such as elaborations, justifications, evaluations, and challenges.	The turn ignores the last speaker and initiates a new thinking trajectory.	The turn is off-task. That is, it contains no information that is relevant to the current problem-solving activity.
Productive	The turn contains productive talk moves such as *express, revoice, elaborate, explain, speculate, evaluate, refer back*, and *reflect*.	The turn does not contain productive talk moves and usually engaged in low-level cognition, such as simple repetition and information sharing.	
Constructive	The turn contributes at least one type of new content knowledge to the current task, including new ideas, new inferences, new speculations, or new proposals.	The turn does not contain relevant content knowledge. Or it refers back or repeats previously contributed content knowledge.	

three-dimensional vectors: {(R_1, P_1, C_1), (R_2, P_2, C_2), (R_3, P_3, C_3), …, (R_n, P_n, C_n)}. CGTS could accordingly be quantified as the average number of turns producing new contributions in a reciprocal and productive talk sequence.

The operationalization of GTS is consistent with its theoretical three-level hierarchy. Some additional assumptions are made for the calculation of the three types of GTS:

(1) Off-task turns will be excluded and assumed as parentheses and as not breaking the current thinking trajectory. In other words, the sustainability of a thinking trajectory only depends on whether the current thinking thread shifts to another thinking thread, not whether it pauses in the dimension of time.
(2) A thinking trajectory is broken only when the reciprocity dimension equals 0. Regarding reciprocity, the present article evaluates a turn's connection to the previous turn rather than the next because of the assumption that the previous speaker has issued a default invitation to all other speakers, even if only addressing a specific person. When the current speaker responds to the previous speaker, the connection is viewed as having been formed; if, in turn, the current speaker addresses the next speaker but the invitation to reciprocate is overlooked by the group, the connection is broken.

Therefore, in a coded turn-taking sequence {(R_1, P_1, C_1), (R_2, P_2, C_2), (R_3, P_3, C_3), …, (R_i, P_i, C_i), …, (R_n, P_n, C_n)},
The number of break points $n_{break} = n\ (R_i = 0)$

$$\text{RGTS} = n(\text{R}_\text{i} = 1\ \text{OR}\ 0)/n_{break}$$

$$\text{PGTS} = n(\text{P}_\text{i} = 1)/n_{break}$$

$$\text{CGTS} = n(\text{C}_\text{i} = 1)/n_{break}$$

In the following sections, this study further investigates how GTS develops in the process of dialogic collaborative problem solving, how it diversifies across different groups, and whether it is related to group outcomes, using both quantitative and interpretative methodologies.

3. Method

3.1. Participants

This study was conducted in a high-ranking primary school in a third-tier city of mainland China in 2019. The participants were 168 fourth graders from five classes (41% females, 59% males) (see Table 3). They were informed of the overall project background (i.e., to study their collaborative dialog) and their task (i.e., to finish three challenging mathematical problems) in class. All of the participants volunteered for this study. Those students not willing to participate were assigned other regular tasks and told to be quiet and not disturb the other students.

3.2. Settings and procedure

Participants were organized into groups of four in their own classrooms, without computers and during regular schooltime (see Fig. 2). Gender and prior mathematics grades were balanced across groups with the help of teachers. Specifically, the author categorized students into three levels based on their previous academic performance (high, middle, and low) and then assigned to the extent possible into groups with one high-level student, one low-level student, and two middle-level students, trying to assign two boys and two girls in one group at the same time. The mathematics teachers then helped adjust the preliminary groupings based on their knowledge of the students. After all groupings were settled, the participants were required to write down the names of their group

Table 3
Background information on the study participants.

	N	Min.	Max.	*M*	*SD*
Age (in years)	168	8	12	10.50	0.57
Recent mathematics grade[a]	144	11	120	100.97	13.78
Recent Chinese grade[a]	144	58	117.5	101.49	10.09
Mother's education level[b]	112	1	6	2.97	1.31
Father's education level[b]	112	1	6	3.39	1.28
Mathematics self-concept[c]	146	1.44	4.00	3.16	0.62
Mathematics enjoyment[c]	146	1.33	4.00	3.58	0.53
Social anxiety[d]	146	1.00	3.00	1.51	0.42

[a] The maximum score is 120.
[b] Education level. 1: Primary school or below; 2: middle school; 3: high school or technical high school; 4: junior college; 5: undergraduate; 6: graduate or above.
[c] Four-point Likert scale. The maximum score is 4.
[d] Three-point Likert scale. The maximum score is 3.

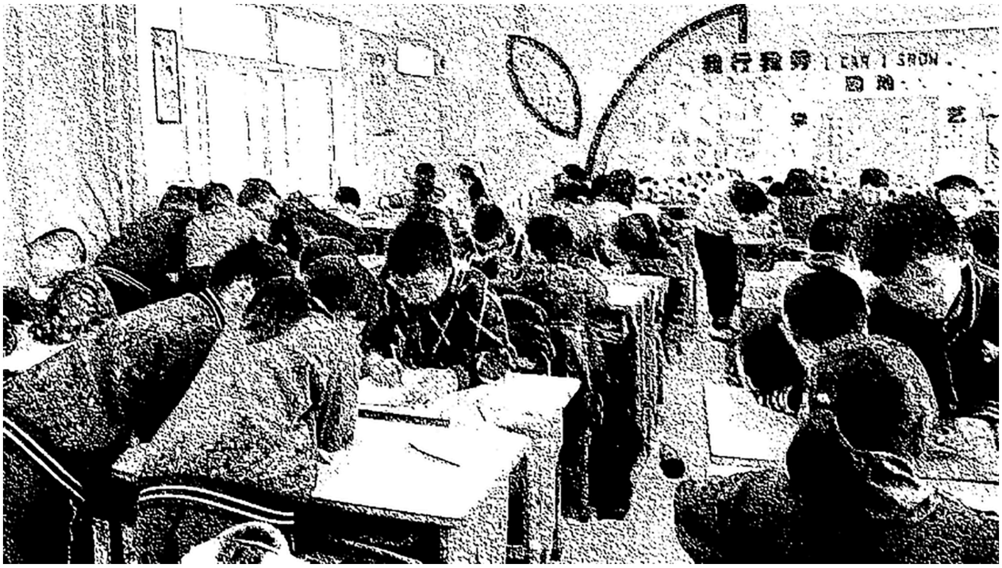


Fig. 2. A typical setting in the study.

members and report their willingness to collaborate with each of their assigned group members on a scale of 1 to10, with 10 representing the highest degree of willingness. Reported willingness was adopted to reveal the level of friendship between students.

The groups were then instructed to collaboratively solve three structured, open-process mathematical problems within half an hour and told not to discuss with other groups or to touch the recorder in the middle of their table. To facilitate the identification of speakers, each group member was also required to introduce themselves following a structured format before solving the problem. During the test, the teachers and the researcher did not moderate the group discussions except to clarify task instructions or maintain classroom discipline.

After the test, students independently completed a questionnaire concerning their demographic information, mathematics learning enjoyment, mathematics self-concept, and social anxiety. Mathematics self-concept and mathematics learning enjoyment were measured using items adapted from the Trends in International Mathematics and Science Study (TIMSS) 2015 questionnaire for fourth graders in Taiwan (Mullis & Martin, 2013), with students being asked to indicate their agreement with each statement on a four-point Likert scale (1 = *strongly agree*, 2 = *somewhat agree*, 3 = *somewhat disagree*, and 4 = *strongly disagree*). Social anxiety was measured using the 10-item Chinese version of the Social Anxiety Scale for Children–Revised (La Greca & Stone, 1993). The students were asked to indicate the frequency of specific behaviors on a three-point Likert scale (1 = *always do this*, 2 = *sometimes do this*, 3 = *never do this*). The measures had a relatively high internal reliability, as indicated by Cronbach's alpha values for social anxiety ($\alpha = 0.835$), mathematics enjoyment ($\alpha = 0.734$), and mathematics self-concept ($\alpha = 0.882$) (Tavakol & Dennick, 2011).

3.3. Materials

The level of difficulty increased throughout the three problems (featuring *ice cream*, a *snake*, and a *bridge*). The *ice cream* (item ID: M041132) and *snake* (item ID: M051006) problems were adapted from the TIMSS survey conducted in 2015 (TIMSS & PIRLS International Study Center, 2015). The *ice cream* problem required students to calculate the unit prices of an ice cream and a popsicle based on the spending of two children on different combinations of these items. The *snake* problem required students to estimate the number of stones that a snake would occupy when it straightened its body. The *bridge* problem was the most difficult and was adapted from the Junior Mathematical Olympiad (Database of Mathematical Olympiad, 2021); it required students to design a bridge-crossing plan for four people that would take the least amount of time.

Students did not have any prior instruction on similar mathematical problems or how to collaboratively solve a problem. To promote collaborative peer talk, none of the problems had any explicit routine for students to follow, although each problem had a unique correct answer. In addition, the solutions to these problems mainly required students' reasoning ability (rather than their specific content knowledge), which helped ensure that the students with high levels of prior knowledge were not at an advantage. The three problems, which were presented to the students in Chinese, have been translated into English and are appended (see Table 1 in the Appendix).

3.4. Data analysis

Written solutions submitted by the groups were graded by the author according to a standard scoring criterion that considered the correctness of the final solution first and then awarded partial credit for solution steps heard in the group discussion audio files if the final answer was wrong (see Table 2 in the Appendix).

The group discussions were transcribed by turns. Each turn was coded for three dimensions: reciprocity, productivity, and constructiveness. The break points and the three types of GTS were automatically calculated using Excel. The author and one trained coder completed the coding of five groups sequentially; they fully discussed their results after coding each group, settled all disagreements, and only then moved on to the coding for the next group. The initial average inter-coder agreement levels for reciprocity, productivity, and constructiveness were 0.63 ($SD = 0.04$), 0.60 ($SD = 0.08$), and 0.71 ($SD = 0.03$), respectively, indicating a substantial initial level of agreement (McHugh, 2012). After discussion, the average inter-coder agreement was 0.87 ($SD = 0.08$), 0.91 ($SD = 0.10$), and 0.94 ($SD = 0.05$), respectively, which is high. The author finished the coding of the remaining 37 groups alone.

Statistical analyses were conducted to examine the features of GTS across the groups and their impact on group outcomes. The unit of analysis is groups as a whole across three tasks. Group outcome refers to one group's average solution scores for the three assigned tasks. Group friendship refers to all group members' average willingness to collaborate with each other. All of the other individual characteristics were averaged for the group-level analysis. Qualitative analysis was conducted on two example groups to investigate how GTS might impact group outcomes.

4. Results

4.1. Dynamics of group thinking sustainability

As shown in Table 4, the 42 groups produced an average of 253 on-task turns ($SD = 111$). Each group talked reciprocally for an average of 3.08 consecutive turns, of which 2.33 turns were productive and 0.72 were constructive. There were strong correlations between RGTS, PGTS, and CGTS, with all correlation coefficients being greater than 0.8 (RGTS and PGTS, $r(42) = 0.926$, $p < .001$; RGTS and CGTS, $r(42) = 0.832$, $p < .001$; PGTS and CGTS, $r(42) = 0.851$, $p < .001$).

Hierarchical regression analysis was conducted to determine the impact of the three GTS metrics on the quality of group final solution. Given the relatively high correlations between RGTS, PGTS, and CGTS, the variance inflation factor (VIF) was checked to see if the data met the assumption of collinearity. A common rule of thumb is that a VIF of 10 or lower is acceptable (Hair et al., 2010), while more conservative estimates indicate the VIF should be 5 or lower (Kock & Lynn, 2012). The results indicated that multicollinearity was a concern based on the conservative estimate but not the common rule of thumb (RGTS, Tolerance = 0.11, VIF = 9.16; PGTS, Tolerance = 0.11, VIF = 9.40; CGTS, Tolerance = 0.21, VIF = 4.88). Given that PGTS is a core variable in the present study, this study did not delete PGTS from the regression models.

A significant regression equation on group solution quality was found ($F(13, 24) = 2.37$, $p < .05$), with an R^2 of 0.56 (see Table 5). Model 3 was a significant improvement over Model 2, with the addition of RGTS, PGTS, and CGTS. RGTS and CGTS significantly predicted group solution quality after accounting for group demographics, participation inequity (i.e., the standard deviation of individual participation rates), and number of on-task turns (RGTS, standardized $\beta = -1.11$, $t(37) = -2.72$, $p < .05$; CGTS, standardized $\beta = 0.72$, $t(37) = 2.42$, $p < .05$). Groups with lower RGTS and higher CGTS achieved better solutions. PGTS was not a significant predictor for final group solution quality (standardized $\beta = 0.658$, $t(37) = 1.59$, $p = .13$). However, the number of on-task turns was found to be a significant predictor of solution quality, standardized $\beta = 0.43$, $t(37) = 2.75$, $p < .05$. Groups that had more intense on-task discussions tended to generate better solutions.

4.2. Two example groups

This study further selected two representative groups to illustrate how the dynamics of GTS differed between them and how the three-level GTS metric helped explain and compare the quality of the various group discussions. The two selected groups were similar to each other regarding the intensity of group interaction, member demographics, and average recent mathematics grades but differed in terms of the final solution quality for the three tasks. Fig. 3 illustrates the dynamics of cumulative RGTS, PGTS, and CGTS for the two groups. It shows that the high-performing group initially had a higher RGTS than the low-performing group but ended up with an overall lower RGTS. The PGTS dynamics followed a pattern similar to that of the RGTS in both groups: it fluctuated greatly in the beginning but soon achieved near stability. The high-performing group persistently had a lower PGTS than the low-performing group after around turn 80. Consistent with the quantitative analysis, the high-performing group was only advantaged in terms of CGTS, which was almost consistently higher than that of the low-performing group across the three tasks.

I now zoom in on the same "stuck" phase that the two groups experienced in solving the *ice cream* problem (see the red dotted boxes

Table 4
Descriptive statistics of group-thinking sustainability.

Variable	*N*	*M*	*SD*	Min.	Max.
RGTS	42	3.08	0.81	1.81	6.43
PGTS	42	2.33	0.63	1.24	4.04
CGTS	42	0.72	0.36	0.13	2.00
Number of on-task turns	42	252.60	110.61	101	500
Number of productive turns	42	189.43	84.79	77	385
Number of constructive turns	42	53.69	22.74	17	99

Table 5
Standardized beta coefficients of predictors for group solution quality in the hierarchical regression analysis.

	Model 1	Model 2	Model 3
Demographics			
Prior Chinese grade	-0.03	0.03	0.02
Prior mathematics grade	0.11	0.17	0.21
Mother's education level	-0.19	-0.17	-0.33
Father's education level	0.03	-0.04	0.10
Mathematics learning enjoyment	0.34	0.35	0.34
Mathematics self-concept	0.16	0.20	0.16
Social anxiety	-0.12	-0.12	-0.20
Friendship	-0.15	-0.11	-0.10
Interaction features			
Participation inequity		0.15	-0.13
Number of on-task turns		0.33	0.43*
Group-thinking sustainability			
RGTS			-1.11*
PGTS			0.66
CGTS			0.72*
R^2	0.27	0.36	0.56
ΔR^2	0.27	0.09	0.20
ΔF	1.35	1.89	3.67*

*$p < .05$.

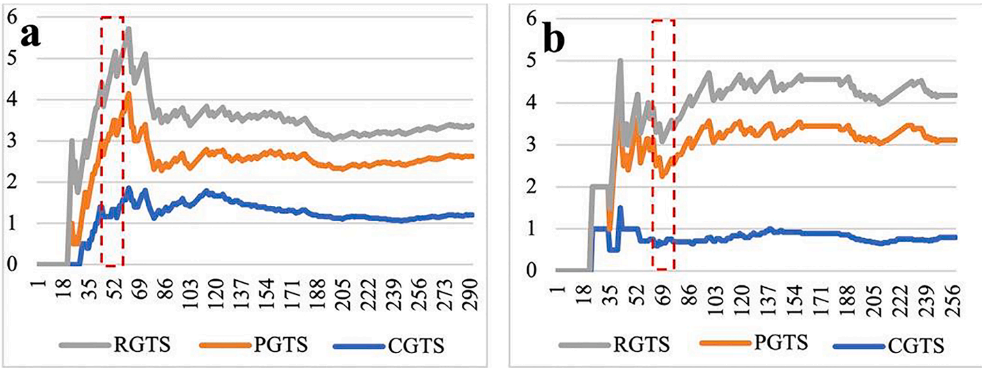


Fig. 3. Cumulative RGTS, PGTS, and CGTS of (a.) the high-performing group and (b.) the low-performing group. The red dotted box denotes a similar "stuck" phase during the solving of the first task for both groups.

Table 6
Translated extract of the stuck phase of the high-performing group.

Turn	Speaker	Content	R	P	C
37	Wang	What do you think, Gu? *(4 s pause)*.	1	1	0
38	Gu	I agree with Gan because they spend 22 and 14 in total. Ming has one more ice cream than Lin. Then… *(interrupted)*.	1	1	0
39	Wang	Oh, wait a moment.	1	0	0
40	Gu	Then for popsicle, one is one more than the other, use… use 22 minus 14 first… *(3 s pause)*. Equals 8 yuan… then…	1	1	1
41	Wang	Yao, you do not do it anymore! *(muffled sound)*.	NULL		
42	Gan	8 divided by 2 equals 4.	1	1	1
43	Gu	Why? Their prices may not be the same.	1	1	1
44	Gan	8 divided by 2 equals 4. Listen to me *(5 s pause)*, 8 yuan…	1	1	0
45	Yao	Gan, I wanna ask a question, … *(muffled sound)*.	0	1	0
46	Gan	It means 8 is equal to one popsicle and one ice cream.	1	1	0
47	Gu	It has begun. The teacher has pressed it for us. Then, how can we calculate the prices for one popsicle and one ice cream?	1	1	0
48	Gan	One popsicle…	1	0	0
49	Wang	I think we can calculate like this.	1	1	0
50	Gan	Say it.	1	1	0
51	Wang	2 times 8 is 16. That is two popsicles and two ice creams. Then 22 minus 16 equals to two popsicles. Then divide by 2. It is one popsicle.	1	1	1
52	Gan	Then you calculate! *(4 s pause)*. Math problem again. I do not wanna do it anymore. I should quit.	1	1	0

in Fig. 3). This task required the students to calculate the unit prices for one ice cream and one popsicle, with known conditions of 22 yuan for two ice creams and four popsicles and 14 yuan for one ice cream and three popsicles. It measured students' reasoning ability with whole numbers. Most of the participating groups became stuck after figuring out the total price for one ice cream and one popsicle using 22 minus 14; the two selected groups reached the same impasse. The following qualitative analysis compares the stuck phase for the two selected groups from a GTS perspective.

The thinking trajectory for the stuck phase of the high-performing group (see Table 6) can be mathematically expressed as {… (1, 1, 0), (1, 1, 0), (1, 0, 0), (1, 1, 1), NULL, (1, 1, 1), (1, 1, 1), (1, 1, 0), (0, 1, 0), (1, 1, 0), (1, 1, 0), (1, 0, 0), (1, 1, 0), (1, 1, 0), (1, 1, 1), (1, 1, 0) …}

$$\text{The RGTS was}: (7+8)/2 = 7.50;$$

$$\text{The PGTS was}: (6+7)/2 = 6.50;$$

$$\text{The CGTS was}: (3+1)/2 = 2.00.$$

For this 16-turn discussion, the RGTS was relatively high, with only one break point (at turn 45; see Fig. 4). The first subsequence comprised turns 37 to 44 except for the off task turn 41, which was skipped in the analysis. Six turns in the first subsequence adopted productive talk moves and three contained new contributions. At turn 37, Wang invited Gu to express her viewpoint, which elicited Gu's agreement with Gan's idea in turn 38. Gu further built on Gan's idea at turn 40 and worked out the total price for one ice cream and one popsicle. Gan built on Gu at turn 42 and proposed getting the price by dividing 8 by 2. Gan's proposal was wrong because of his assumption that an ice cream and a popsicle had the same price. This mistake was immediately found by Gu, who pressed for Gan's reasoning at turn 43 and challenged his proposal by pointing out that "their prices may not be the same." Gan tried to explain but failed at the following turn.

Yao jumped in at turn 45 and asked Gan another clarification question that initiated a new 8-turn thinking trajectory. Gan helped Yao understand the meaning of 8 at the following turn; however, he could not figure out the next step when Gu pressed him to say more. Wang jumped in at turn 49 and at turn 51 added that they could get the price of a popsicle by subtracting 2 times 8 from 22.

For the low-performing group, the thinking trajectory for the stuck phase (see Table 7) can be expressed as {… (0, 1, 1), (1, 1, 0), (1, 1, 0), (0, 0, 0), (0, 0, 0), (1, 1, 0), (1, 1, 0), (1, 0, 0), (0, 1, 1), (1, 1, 0), (1, 0, 0), (1, 1, 1), (1, 1, 0), (1, 1, 0), (1, 1, 0), (0, 1, 0)…}

$$\text{The RGTS was}: (3+1+4+7+1)/5 = 3.20;$$

$$\text{The PGTS was}: (3+0+2+6+1)/5 = 2.40;$$

$$\text{The CGTS was}: (1+0+0+2+0)/5 = 0.60.$$

This 16-turn discussion was cut into five subsequences (see Fig. 5). The RGTS was thus much lower than for the high-performing group; PGTS and CGTS were accordingly lower as well. In the first subsequence, Bao proposed that 8 yuan was the total price of one ice cream and one popsicle; however, he made a mistake when referring to the given conditions, leading to subsequent corrections by Li

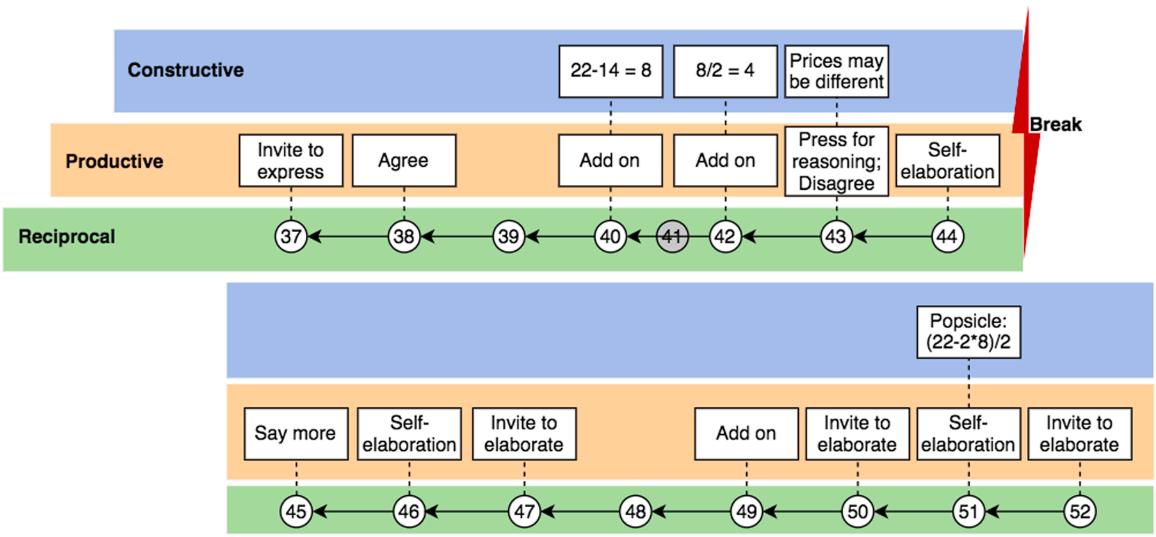


Fig. 4. Dynamics of group thinking sustainability in the high-performing group.

Table 7
Translated extract of the stuck phase of the low-performing group.

Turn	Speaker	Content	R	P	C
61	Bao	So, one popsicle plus one ice cream equals 8. So, back to the figure above, we could know Ming has two popsicles and two ice creams…	0	1	1
62	Li	Ah?! Four popsicles!	1	1	0
63	Wang	Four popsicles *(muffled sound)*	1	1	0
64	Xiao	I got it. I got it.	0	0	0
65	Li	Four popsicles and two ice creams. Haha.	0	0	0
66	Xiao	No, the 8 more yuan should be this…	1	1	0
67	Li	The 8 more yuan is this and this. After subtracting these two…	1	1	0
68	Xiao	Two popsicles! *(interjected)*	1	0	0
69	Li	After subtracting these two, there is one and three popsicles.	0	1	1
70	Bao	You two talked so much which amounts to nothing!	1	1	0
71	Li	One ice cream and three popsicles!	1	0	0
72	Xiao	No, this should be one group. One group costs 8. There are two groups. Two more left. The remaining money equals the price of two popsicles.	1	1	1
73	Li	That is exactly what I meant.	1	1	0
74	Xiao	Then why did you say that again…	1	1	0
75	Bao	You two talked so much which amounts to nothing! Haha, one ice cream and three popsicles…	1	1	0
76	Li	Next welcome Yao to express his idea. What you just said is quite good. You should talk to the recorder.	0	1	0

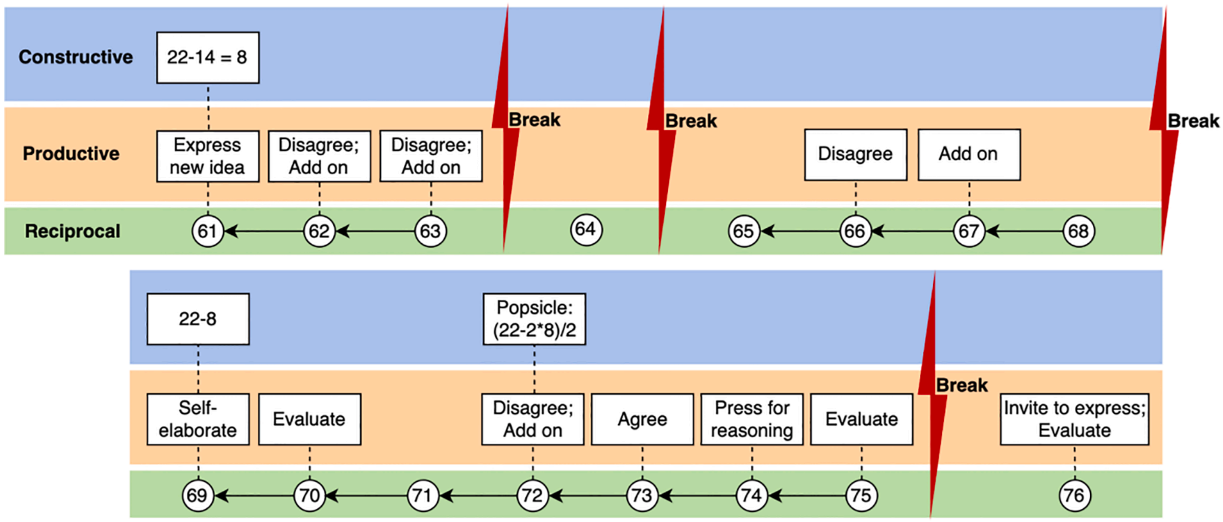


Fig. 5. Dynamics of group-thinking sustainability in the low-performing group.

and Wang at turns 62 and 63, respectively. The second subsequence was initiated by Xiao, who claimed he had figured out how to solve the problem at turn 64. However, his voice was overlooked by the group, which caused the second 1-turn subsequence.

In the following two subsequences (from turn 65 to turn 75), Li and Xiao competed for turns to express their ideas, while Bao jumped in to express his evaluations at turns 70 and 75. At turn 65, Li repeated her correction (in turn 62) of Bao's mistake. Xiao immediately disagreed and attempted to express his idea but was interrupted by Li, who built on his idea. However, it seemed that Li did not understand what Xiao wanted to propose; therefore, Xiao jumped in again and tried to express his idea at turn 68. Li overlooked Xiao's voice again and further elaborated her claim at turn 69. Bao seemed not to follow their discussion and thus commented on the ineffectiveness of Xiao and Li's discussion at turn 70. This did not receive constructive feedback from Li, who continued insisting on her idea through repetition. At turn 75, Bao again commented on the ineffectiveness of the discussion but was ignored, causing turn 76 to begin another (1-turn) subsequence.

By comparing the stuck stages of the two groups, this study shows that both produced the same intensity of interaction but differed dramatically in terms of GTS, with the high-performing group having relatively stronger GTS than the low-performing group. Specifically, their discussions were more reciprocal and involved more productive talk moves as well as more new contributions. The high-performing group also involved all four group members in the stuck phase. They asked each other many reciprocal questions to induce their peers' elaborated reasoning, such as "What do you think?" (turn 37), "Why?" (turn 43), "I wanna ask a question…" (turn 45), and "How can we calculate the prices?" (turn 47). These questions enabled co- and socially shared regulative processes. In contrast, the low-performing group seldom adopted such invitation-format questions, and one of the four members did not participate in the extracted discussion at all. Li and Xiao competed for turns to express their ideas and dominated the whole discussion. In particular, it seemed Li focused on expressing her own ideas without fully understanding Xiao's, even ignoring Xiao's words. This caused the fragmentation of their discussion and indicated a lack of co- and socially shared regulation.

5. Discussion

The present study has shed light on the sustainability of collective thinking in dialogic collaborative problem solving, with its most important contribution being the conceptualization and operationalization of group thinking sustainability (GTS), defined as a group's capacity to persistently engage in high-order collective thinking in peer talk. GTS was conceptualized as a three-level nested hierarchy (comprised of reciprocity, productivity, and constructiveness) and operationalized as the average length of a corresponding turn-taking sequence. Reciprocal group thinking sustainability (RGTS) provides the foundation of GTS by determining the segmentation of group talk sequence. Productive group thinking sustainability (PGTS) illustrates the quality of thinking underlying group talk, while constructive group thinking sustainability (CGTS) captures the cognitive outcomes in the process of group talk.

5.1. Discussion of the study findings

The analysis of the interactions of 42 four-person groups in primary schools revealed that RGTS, PGTS, and CGTS were highly correlated with each other due to the design of the three-level GTS hierarchy. The productivity and constructiveness of peer talk both depended on students talking reciprocally and not overlooking peer voices, making the interdependence of individual utterances a prerequisite. Furthermore, this study revealed that PGTS had a stronger correlation with RGTS than did CGTS. This is understandable, as most of the productive peer talk moves identified in previous studies were reciprocal and aimed to perform or elicit high-order thinking (King, 2008); therefore, utterances involving productive peer talk moves are very likely to also be reciprocal. CGTS and PGTS were also highly correlated. This is explainable, as turns involving new contributions always contained such productive talk moves as *express, explain*, or *elaborate*. There is a close correspondence between students' overt productive peer talk and their underlying high-level collaborative knowledge construction (e.g., knowledge assimilation and integration) (King, 1994).

The present study found that groups with higher CGTS achieved better solutions. This is in line with a vast body of literature supporting the predictive power of constructive talk for group final outcomes (e.g., Chi & Menekse, 2015; Heo et al., 2010; Rojas--Drummond et al., 2006). CGTS denotes the amount of emergent new knowledge in a reciprocal talk sequence. It captures group performance in process and thus belongs to the same type of measure as group final performance. It is thus understandable that performance in process could significantly predict final performance. The present study extends previous findings on the predictive power of aggregated co-constructive talk on group final outcomes. It demonstrates that the persistence of such co-constructive talk also matters.

RGTS was also a significant predictor of final solution quality. However, groups with lower RGTS achieved a higher solution quality, seemingly contradicting previous studies that emphasized reciprocity as a feature of effective group talk (Barron, 2000; Heo et al., 2010; Meier et al., 2007). This study proposes two possible explanations for this inconsistency.

First, it might be due to the general claim that good peer talk does not necessarily lead to a good group solution. This might also explain the non-significant effect of PGTS. Each student group's ability to solve a given problem was a measure of the group's short-term success, but solution quality might have been affected by multiple factors in addition to peer talk. For example, although the tasks assigned in this study were challenging for individual students, the study showed that some groups efficiently solved the problems by trusting those group members with a relatively high academic status. The nature of mathematical problems might also matter. In one instance in the study, a single student in a group suddenly thought of the solution without intense peer discussion, suggesting reciprocal and productive interaction might facilitate individual group members learning from group work, but without the group necessarily achieving short-term success. Previous studies on productive failure have demonstrated the long-term benefits of failing in tasks (Kapur, 2008; Kapur & Bielaczyc, 2012). From a learning-oriented perspective, the present study still encourages persistence in reciprocal and productive interaction, which might not lead to success in the task at hand, but which may benefit individual learning and long-term group success.

Second, there may be differences between smooth phases and stuck phases, and high-performing groups might have higher RGTS in stuck phases but not when things are going well. The qualitative analysis of the dynamics of GTS over the three problems showed fairly consistent overall effects of the three types of GTS with the quantitative analysis of the same; however, the result for the selected stuck phase revealed a different pattern, with the high-performing group being advantaged over the low-performing one on all three GTS metrics. This might suggest an alternative explanation for the present study's inconsistency with previous findings regarding the benefits of reciprocal interaction and productive peer talk; that is, a high-performing group might differ from a low-performing group in its degree of persistence in high-order collective thinking during some critical problem-solving phases, rather than through the whole problem-solving process. This possibility remains open for further studies to investigate.

5.2. Implications for understanding group thinking

This study provides novel insights into understanding and comparing the sustainability of collective thinking in group talk. It echoes the emerging research interest in the temporality of peer talk (Chen et al., 2017; Csanadi et al., 2018; Kapur, 2011) and extends previous research on what qualifies as high-quality peer talk (King, 2008; Littleton & Mercer, 2013; van Aalst, 2009; Webb & Mastergeorge, 2003). GTS concerns not only the quality of students' collective thinking but also the sustained quality thereof.

GTS provides a process-oriented approach to characterizing the collective thinking of various groups, which may also extend previous efforts to measure group thinking (Wegerif et al., 2017; Woolley et al., 2010). Woolley et al. (2010) extracted one single factor that explained a fairly large percentage of group performance variance over a wide variety of tasks and therefore demonstrated the existence of a collective intelligence factor, or *c* factor. They also found that this *c* factor was correlated not only with a group's composition features (i.e., percentage of female students, individual social sensitivity) but also its interaction patterns (i.e., participation equality). Wegerif et al. (2017) proposed a Ravens reasoning test to measure the *c* factor of groups, which could be more easily implemented in practice. They also suggested additional qualitative analysis on group interaction to further explain how social sensitivity might lead to effective group thinking. GTS provides an alternative way to help analyze group interaction processes in specific tasks that may help explain how the *c* factor manifests itself in various patterns of interaction and leads to various group outcomes. It allows the analysis of one group across a wide range of tasks and the comparison of various groups on the same task.

As a three-level metric, GTS enables a fine-grained analysis at the turn level. It provides an alternative approach to characterizing peer talk. For example, it may help differentiate the various types of talk identified by Mercer (1995) and Wegerif and Mercer (1997) (i.e., cumulative talk, disputational talk, exploratory talk) and describe the emergent group cognition proposed by Stahl (2014). It also avoids the disadvantages of coding-and-counting approaches (Csanadi et al., 2018) by capturing the interdependence of utterances. Instead of directly analyzing the sustainability of productive peer talk moves or constructive contributions, the three-level hierarchical metric adopts the reciprocity of peer talk as the first-level structure, which is more consistent with the discursiveness of peer talk. In this study, students sometimes consecutively adopted productive talk moves and/or made new contributions. However, their discussions were more often filled with non-productive talk that maintained rather than broke the thinking flow.

The three-level metric also provides a new segmentation technique based on the reciprocity of peer talk, allowing fine-grained analysis within segmented reciprocal talk sequences. Segmentation is crucial to interaction analysis. The granularity of segmentation could be at the level of the speech act, the move, the exchange, the sequence, or the interaction, based on the "hierarchical-functional" model (Roulet, 1991). However, how to make these segmentations is still unclear (Baker et al., 2007). The three-level GTS metric might provide an effective way to segment interactions into sequences that engage in sustainable collective thinking. The present study suggests that this three-level GTS metric could guide future analysis on the sequential structure of collective thinking.

5.3. Implications for scaffolding

GTS not only helps researchers understand and compare the sustainability of group thinking but also bears important implications for practice. Reciprocal, productive, and constructive thinking represent three aspects of high-quality peer talk: turn-taking, productive talk, and knowledge construction. Therefore, the three-level hierarchical GTS metric also implies a three-level scaffolding for peer talk.

First, RGTS illustrates whether each voice in group talk gets feedback from its peers. Highly reciprocal talk requires students to listen to each other and voice their acknowledgement of each contribution. This is consistent with previous studies on listenership that have suggested that students should respond to the previous speaker when taking turns and that all interruptions and overlaps should be supportive to ensure a smooth interaction flow (Liu et al., 2016; McCarthy, 2003). The reciprocity also reflects the virtue of respect in dialogic interaction, where everyone is treated as equal and each voice is heard and responded to (Bakhtin, 1981; Matusov et al., 2019). Turn-taking affects whether and to what extent different individuals in a group can access the conversational floor and influence a group discussion through their verbal and non-verbal contributions (Engle et al., 2014). Access to a conversational floor can be manipulated by changing the context (Shah & Lewis, 2019). Teachers are therefore urged to ensure equitable access for each group member and to help students build awareness of and develop respect for each member's contributions (Boaler, 2008).

Second, PGTS quantifies the extent to which students adopt productive talk moves in a reciprocal talk sequence. There have been many efforts made toward helping students take up these productive talk moves (Gillies, 2019; King, 1997; van de Pol et al., 2019). PGTS is consistent with these studies, also suggesting the importance of encouraging students to adopt productive peer talk moves. Furthermore, it highlights the important of urging students to adopt productive peer talk moves in a reciprocal way rather than only focusing on how often they adopt these moves. To maintain reciprocity, students could be guided to engage with each other's ideas, to connect their own ideas to those of others, and to ask more reciprocal questions with the aid of productive peer talk moves.

Third, CGTS quantifies the extent to which a reciprocal talk sequence generates new contributions. This metric illustrates whether productive peer talk moves in a reciprocal talk sequence lead to new contributions. It is highly related to group final performance. CGTS emphasizes the outcomes of peer talk and the importance of encouraging students to build on each other's contributions to promote joint knowledge. This is in line with the co-constructive discourse pattern reported in Chi and Menekse's (2015) work as well as the idea-centered knowledge building discourse advocated by the knowledge building community (e.g., Scardamalia & Bereiter, 2014; Tong, 2020; van Aalst, 2009). Similar to PGTS, CGTS also emphasizes the reciprocity of peer talk. That is, it highlights the importance of students building on existing knowledge or ensuring a smooth change of line of inquiry, rather than only focusing on how many new ideas they have contributed.

Despite the great potential of peer talk for promoting individuals' domain-related learning and cognitive development, productive peer talk seldom happens spontaneously in classrooms (e.g., Cohen & Lotan, (2014); Gillies, 2019; Miller & Hadwin, 2015). Thus, students need explicit guidance on how to use language effectively and how to regulate group interactions (e.g., Belland et al., 2013; King, 2008; Näykki et al., 2017). The three-level GTS metric helps integrate the three critical scaffolding dimensions of peer talk suggested above (i.e., turn-taking, productive talk, and knowledge construction) to sustain group thinking. The implications for scaffolding are roughly consistent with the three dimensions of accountable talk in the classroom: accountability to community, reasoning, and knowledge (Michaels et al., 2008). This three-level schema provides an alternative framework for future studies to design peer talk intervention programs. For example, possible design-based research studies could use GTS as a pre-test on group thinking at the beginning, then progressively intervene in students' talk at the three levels of reciprocity, productivity, and construction, and then retest students' group thinking at the end of the study to validate the efficacy of such a three-level schema and improve the design of the intervention program.

5.4. Limitations

There are some limitations to this study that could be addressed by future studies. First, this study used group average performance on three various tasks as the unit of analysis, which may have failed to fully capture the predictive power of GTS on group performance. GTS is a content-free schema to characterize the sustainability of high-quality collective thinking. Although reciprocity, productivity, and constructiveness are theoretically essential to high-quality peer talk, the quantified GTS may vary greatly over various tasks. The three tasks in the present study differed in the level of structuredness though they were all open-processed mathematical reasoning problems. The snake and bridge tasks were relatively more ill-structured and more unfamiliar to students than the ice cream task. Different types of tasks are likely to produce different styles of peer talk. For example, sequencing of group talk is often a common practice in typical mathematical problem solving (Stahl, 2014), while creative tasks in other subjects or without unique correct answers may lead to different kinds of talk even though high-quality talk around such tasks does contain some shared communicative acts or patterns with that around the reasoning tasks (Rojas-Drummond et al., 2006). Therefore, the author suggests future studies to validate the predictive power of GTS on group performance based on a single problem or single types of task. It would also be useful to compare a group's GTS over various types of task to understand how GTS is affected by task features.

Second, this study was somewhat limited by its context. It only applied GTS to foursomes in primary school, and it remains unknown whether the findings could be generalized to other samples and group sizes. Further, the quantification of GTS is mainly applicable to the face-to-face, whole-group context, but RGTS only considers feedback on the last turn and does not currently support the inclusion of time lag. Therefore, it fails to deal with such complex interactions as parallel discussions, which might happen when subgroups emerge in group work, or online discussions. Future studies could adapt GTS to fit more complex contexts such as these.

6. Conclusion

This study established a construct called group-thinking sustainability (GTS), which characterized the sustainability of high-quality peer talk at three levels: whether students responded to each voice and talked in a reciprocal way; if students talked reciprocally, the degree to which they adopted productive peer talk moves; and if students talked reciprocally and productively, the degree to which they made new contributions. It further applied this construct to describe the collective thinking of primary school foursomes in solving mathematical problems and partially validated the predictive power of GTS on group outcomes.

This study provides an alternative tool to aid researchers in understanding and measuring the sustainability of high-level collective thinking. The three-level schema also implies a three-level scaffolding (i.e., turn-taking, productive talk, and knowledge construction) for teachers to use to support sustainable collective thinking in collaborative peer talk.

This study had also generated many open questions for further investigation, such as how the GTS of one group differs across various types of tasks, whether and how GTS can help predict group performance in a single type of task, and whether GTS mediates the impact of collective intelligence on group performance. This study calls for further efforts within this line of research on the sustainability of productive peer talk.

CRediT authorship contribution statement

Liru Hu: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft.

Declaration of Competing Interest

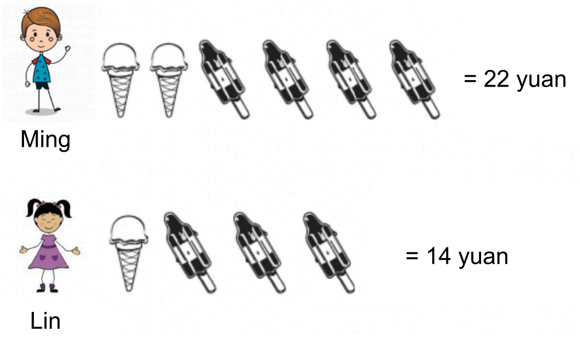
The author declares that there is no conflict of interest.

Appendix

Tables A1 and A2.

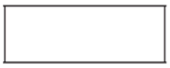
Table A1
Mathematical problems set in the study.

Ice Cream. Ming buys two ice creams and four popsicles. He spends 22 yuan in total. Lin buys one ice cream and three popsicles. She spends 14 yuan in total. How much do one ice cream and one popsicle cost? Please write out your problem-solving process in detail.

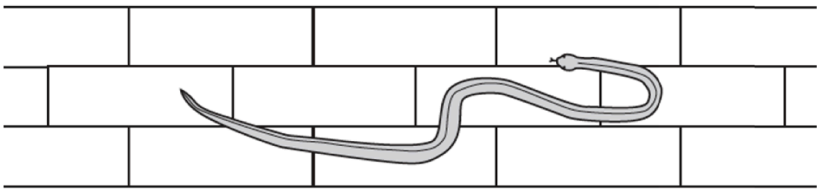


Answer: One ice cream costs _______ yuan.
One popsicle costs ______ yuan.
Your problem-solving process:

Snake. There is a snake on a pathway in a park. The pathway is made of stones, as shown below.



If we straightened the snake out to its full length, how many stones would it occupy? Please try to solve this problem by using as many approaches as you can and write out all of the solutions that you can think of.



(*continued on next page*)

Table A1 (*continued*)

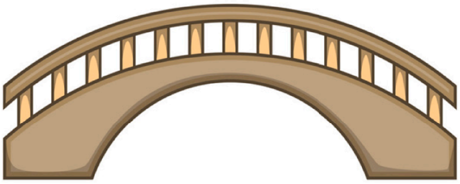
Answer: The straightened snake would occupy _____ stones.

Solution 1:

Solution 2:

Solution 3:

Bridge. Jia, Yi, Bing and Ding want to cross a bridge. It takes them 1 minute, 2 minutes, 5 minutes and 10 minutes, respectively. They have to use a flashlight because it is dark. However, they only have one flashlight, and the bridge can only support two people at most, due to its limited loading capacity. They want to cross the bridge as quickly as possible. How quickly can they cross? Please help them plan their crossing and calculate the smallest amount of time it will take.



Your crossing plan is:

__________ and__________ cross the bridge

↓

__________ returns

↓

__________ and__________ cross the bridge

↓

__________ returns

↓

__________ and__________ cross the bridge

Table A2
Scoring criteria of three tasks.

Task	Features of the solution	Score
Ice cream	Both unit prices are correct.	10
	One unit price is correct.	5
	Two prices are reversed.	5
	No one is correct.	0
Snake	Final answer is four.	5
	Final answer is five.	3
	Final answer is bigger than five or smaller than four.	0
	All three solutions are correct.	5
	Only two solutions are correct.	4
	Only one solution is correct.	2
	No solution is correct.	0
Bridge	The plan is optimal (i.e., 17 min) and the reported total time is correct.	10
	The plan is optimal, but the reported total time is wrong.	8
	The plan is suboptimal (i.e., 19 min) and the reported total time is correct.	5
	The plan is suboptimal, but the reported total time is wrong.	3
	The plan is correct but time-consuming (i.e., longer than 19 min) and the reported total time is correct.	1
	The plan is wrong.	0

Note. The total score for each item is 10.

References

Asterhan, C. S., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive science, 33*(3), 374–400.

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development, 2*(3), 202–221. https://doi.org/10.1111/j.1467-9507.1993.tb00014.x

Baker, M., Andriessen, J., Lund, K., Van Amelsvoort, M., & Quignard, M. (2007). Rainbow: A framework for analysing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning,, 2*(2–3), 315–357. https://doi.org/10.1007/s11412-007-9022-4

Bakhtin, M. M., Emerson, C., Holquist, M., & McGee, V. W. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press. Trans.

Bakhtin, M. M. (1999). *Problems of dostoevsky's poetics*. Minneapolis: University of Minnesota Press (Vol. 8).

Bakhtin, M. (1981). *The dialogic imagination*. Austin: University of Texas Press.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences, 9*(4), 403–436. https://doi.org/10.1207/S15327809JLS0904_2

Belland, B. R., Kim, C. M., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist, 48*(4), 243–270. https://doi.org/10.1080/00461520.2013.83892

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. https://doi.org/10.1016/S0883-0355(03)00078-8

Blau, P. M. (1964). *Exchange and power in social life*. New York, NY: John Wiley.

Boaler, J. (2008). Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach. *British Educational Research Journal, 34*(2), 167–194. https://doi.org/10.1080/01411920701532145

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

Brush, T. A. (1998). Embedding cooperative learning into the design of integrated learning systems: Rationale and guidelines. *Educational Technology Research and Development, 46*(3), 5–18. https://doi.org/10.1007/BF02299758

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

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271–315.

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., et al. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science, 42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns in peer collaboration that promote learning. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 263–274). Washington DC: American Educational Research Association.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Clarà, M. (2019). Building on each other's ideas: A social mechanism of progressiveness in whole-class collective inquiry. *Journal of the Learning Sciences, 28*(3), 302–336.

Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogeneous classroom* (third edition). New York: Teachers College Press.

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

Database of Mathematical Olympiad (2021). (n.d.). Mathematical Olympiad problems for fifth grade students. Retrieved from https://www.aoshuku.com/timu/23064.html.

Dyke, G., Kumar, R., Ai, H., & Rosé, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), *Proceedings of the 10th international conference of the learning sciences (ICLS 2012), The future of learning* (pp. 363–370). Sydney: ISLS.

Engle, R. A., Langer-Osuna, J. M., & McKinney de Royston, M. (2014). Toward a model of influence in persuasive discussions: Negotiating quality, authority, privilege, and access within a student-led argument. *Journal of the Learning Sciences, 23*(2), 245–268. https://doi.org/10.1080/10508406.2014.883979

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

Fu, E. L. F., Van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 441–478. https://doi.org/10.1007/s11412-016-9245-3

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. (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

Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.

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

Greeno, J. G. (2015). Classroom talk sequences and learning. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 255–262). Washington DC: American Educational Research Association.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* ((7th Edition)). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., et al. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction,, 9*, 16–44.

Heo, H., Lim, K. Y., & Kim, Y. (2010). Exploratory study on the patterns of online interaction and knowledge co-construction in project-based learning. *Computers and Education, 55*(3), 1383–1392. https://doi.org/10.1016/j.compedu.2010.06.012

Johnson, D. W., & Johnson, R. (2016). Cooperative learning and teaching citizenship in democracies. *International Journal of Educational Research, 76*, 162–177. https://doi.org/10.1016/j.ijer.2015.11.009

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. https://doi.org/10.1080/07370000802212669

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*, 39–56. https://doi.org/10.1007/s11412-011-9109-9

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

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. (2008). Structuring peer interaction to promote higher-order thinking and complex learning in cooperating groups. In R. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (pp. 73–91). New York: 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

Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems, 13*(7), 546–580.

Kolikant, Y. B., & Pollack, S. (2020). The power of a dialogical framework to articulate collaborative learning in the 21st century. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 634–646). London, United Kingdom: Routledge.

La Greca, A. M., & Stone, W. L. (1993). Social anxiety scale for children-revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology, 22*(1), 17–27.

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Abingdon: Routledge.

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on computational tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI-Global.

Matusov, E., Marjanovic-Shane, A., & Gradovski, M. (2019). *Dialogic pedagogy and polyphonic research art: Bakhtin by and for educators*. New York: Springer. https://doi.org/10.1057/978-1-137-58057-3

McCarthy, M. J. (2003). Talking back: 'Small' interactional response tokens in everyday conversation. *Research on Language in Social Interaction, 36*(1), 33–63.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning,, 2*(1), 63–86. https://doi.org/10.1007/s11412-006-9005-x

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon: Multilingual Matters.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London and New York: Routledge.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 346–361). Washington, DC: American Educational Research Association.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*(4), 283–297. https://doi.org/10.1007/s11217-007-9071-1

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior, 52*, 573–588. https://doi.org/10.1016/j.chb.2015.01.050

Moore, W. S. (2002). Understanding learning in a postmodern world: Reconsidering the Perry scheme of intellectual and ethical development. In B. K. Hofer, & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 17–36). Mahwah, NJ: Lawrence Erlbaum Associates.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Näykki, P., Isohätälä, J., & Järvelä, S. (2017). Facilitating socio-cognitive and socio-emotional monitoring in collaborative learning with a regulation macro script – an exploratory study. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 251–279.

Noroozi, O., Teasley, S. D., Biemans, H. J., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning,, 8*(2), 189–223.

Piaget, J. (1932). *The language and thought of the child* (2nd Edition). London: Routledge & Kegan Paul.

Poole, M. S., van de Ven, A., Dooley, K., & Holmes, M. E. (2000). *Organizational change and innovation processes: Theories and methods for research*. New Oxford: Oxford University Press.

Popov, V., Biemans, H. J. A., Fortuin, K. P. J., van Vliet, A. J. H., Erkens, G., Mulder, M., et al. (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

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

Resnick, L. B., Pontecorvo, C., & Säljö, R. (1997). Discourse, tools, and reasoning: Essays on situated cognition. In L. B. Resnick, C. Pontecorvo, R. Säljö, & B. Burge (Eds.), *Discourse, tools, and reasoning: Essays on situated cognition* (pp. 1–20). Berlin: Springer.

Rojas-Drummond, S., Mazón, N., Fernández, M., & Wegerif, R. (2006). Explicit reasoning, creativity and co-construction in primary school children's collaborative activities. *Thinking Skills and Creativity, 1*(2), 84–94.

Roulet, E., et al. (1991). On the structure of conversation as negotiation. In J. R. Searle, et al. (Eds.), *(On) searle on conversation* (pp. 91–99). Amsterdam: John Benjamins.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The cambridge handbook of the learning sciences*. New York: Cambridge University Press (2nd ed, pp. 397- 417).

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 43–102. https://doi.org/10.1007/s11412-009-9080-x

Sfard, A. (2015). Why all this talk about talking classrooms? Theorizing the relation between talking and learning. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 245–253). Washington DC: American Educational Research Association.

Shah, N., & Lewis, C. M. (2019). Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and Instruction, 37*(4), 423–452. https://doi.org/10.1080/07370008.2019.1631825

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 51*(7), 870–901.

Stahl, G. (2014). The constitution of group cognition. In L. Shapiro (Ed.), *Handbook of embodied cognition*. New York: Routledge (ch.32, pp. 335–346).

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

Swiecki, Z. (2020). *Modeling interdependence in collaborative problem-solving*. University of Wisconsin–Madison [Doctoral dissertation].

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55.

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

TIMSS & PIRLS International Study Center (2015). TIMSS 2015 item information tables–fourth grade. Retrieved from https://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_G4_ItemInformation.zip.

Tong, Y. (2020). *Developing epistemic understanding and productive inquiry in knowledge building through meta-discourse*. The University of Hong Kong [Doctoral dissertation]. HKU Theses Online (HKUTO).

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

van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences, 28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258

van der Veen, C., & van Oers, B. (2017). Advances in research on classroom dialogue: Learning outcomes and assessments. *Learning and Instruction, 48*, 1–70. https://doi.org/10.1016/j.learninstruc.2017.04.002

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer Press.

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

Wang, Q. (2009). Design and evaluation of a collaborative learning environment. *Computers & Education, 53*(4), 1138–1146. https://doi.org/10.1016/j.compedu.2009.05.023

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., et al. (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

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39*(1–2), 73–97. https://doi.org/10.1016/S0883-0355(03)00074-0

Webb, P. (2019). Dialogic thinking together towards abstract reasoning. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 570–580). London, United Kingdom: Routledge.

Wegerif, R. (2020). Towards a dialogic theory of education for the internet age. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 14–26). London, United Kingdom: Routledge.

Wegerif, R. (2013). *Dialogic: Education for the internet age*. London, United Kingdom: Routledge.

Wegerif, R., Fujita, T., Doney, J., Linares, J. P., Richards, A., & Van Rhyn, C. (2017). Developing and trialing a measure of group thinking. *Learning and Instruction, 48*, 40–50.

Wegerif, R., & Mercer, N. (1997). A dialogical framework for researching peer talk. In R. Wegerif, & P. Scrimshaw (Eds.), *Computers and talk in the primary classroom* (pp. 49–61). Clevedon: Multilingual Matters.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688 (New York, N.Y.).