How collaborative problem solving promotes higher-order thinking skills
A systematic review of design features and processes

Ceballos, Haydeé; Bogaart, Theo van den; van Ginkel, Stan; Spandaw, Jeroen; Drijvers, Paul

DOI
10.1016/j.tsc.2025.102001

Publication date
2026

Document Version
Final published version

Published in
Thinking Skills and Creativity

Thinking Skills and Creativity 59 (2026) 102001

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

How collaborative problem solving promotes higher-order thinking skills: A systematic review of design features and processes

Haydeé Ceballos [a,b,*], Theo van den Bogaart [b], Stan van Ginkel [b], Jeroen Spandaw [c], Paul Drijvers [a]

[a] Freudenthal Institute, Utrecht University, PO Box 85.170, 3508 AD, Utrecht, the Netherlands
[b] Hogeschool Utrecht, University of Applied Sciences, PO Box 14007, 3508 SB, Utrecht, the Netherlands
[c] EWI, Delft University of Technology, PO Box 5031, 2600 GA, Delft, the Netherlands

A R T I C L E I N F O

Keywords:
Collaborative problem solving
Higher-order thinking skills
Mathematics education
Systematic review
Peer communication
Task design
Group processes

A B S T R A C T

Collaborative Problem Solving (CPS) work in mathematics education are widely recognized for engaging students in cognitively demanding activities that foster Higher-Order Thinking Skills (HOTS), like critical thinking and reasoning. However, connections between design features, CPS processes, and learning outcomes remain complex and not fully understood. To address this, we applied a conjecture-based framework to systematically review 45 empirical studies published between 2010 and 2022, focusing on how specific task designs and CPS processes contribute to HOTS. We used a machine learning tool to prioritize relevant studies and streamline the selection process, ending after a threshold number of consecutive irrelevant articles. Guided by the conjecture-based framework, our analysis highlighted how cognitive processes in CPS function as essential mechanisms of learning and measurable outcomes. Specifically, design features, such as technology-supported exploratory tasks and open-ended problems, encourage reflective discourse and deeper cognitive engagement. We also found that structured group procedures, including clear roles and guided interaction protocols, improve collaboration. Nonetheless, challenges like miscommunication and uneven participation can limit CPS from fully realizing its potential to cultivate HOTS. Overall, these findings underscore the importance of aligning task design with CPS processes and using strategies to address collaboration barriers, particularly those related to communication. Without clear protocols and consistent dialogue, even well-designed CPS tasks can fail to cultivate HOTS. In conclusion, this review offers practical insights for educators and researchers implementing CPS effectively in mathematics education, highlighting that fostering open, structured communication is vital for optimizing both collaborative processes and the development of advanced cognitive skills.

1. Introduction

Collaborative Problem Solving (CPS) during mathematics lessons is widely recognized as a powerful approach to foster student

* Corresponding author at: Kapitein Robstraat 20, 1336 HE, Almere, the Netherlands.
E-mail addresses: m.h.ceballos@uu.nl, haydee.ceballos@hu.nl (H. Ceballos), theo.vandenbogaart@hu.nl (T. Bogaart), stan.vanginkel@hu.nl (S. van Ginkel), J.G.Spandaw@tudelft.nl (J. Spandaw), p.drijvers@uu.nl (P. Drijvers).

https://doi.org/10.1016/j.tsc.2025.102001
Received 23 March 2025; Received in revised form 7 September 2025; Accepted 7 September 2025
Available online 9 September 2025

engagement and enhance learning outcomes. Effective CPS tasks rely on well-designed features that not only encourage collaboration but also activate Higher-Order Thinking Skills (HOTS) such as reasoning, analysing, and problem solving (Liljedahl, 2016). Likewise, Schoenfeld (2023) highlights the importance of carefully structured tasks to promote agency, ownership, and positive mathematical identities among students.

Research shows that certain design features in CPS tasks, such as open-ended problems, group composition, and interactive tools, are critical for supporting meaningful peer discussions and cognitive engagement. These features encourage students to articulate, critique, and refine their mathematical arguments, which strengthens their conceptual understanding and problem-solving abilities (Dekker & Elshout-Mohr, 1998; Pijls & Dekker, 2011). Even students who struggle with challenging tasks can contribute constructively by questioning their peers' reasoning or providing counterarguments (Ceballos & Spandaw, 2017; Crouch & Mazur, 2001; Mazur, 1997).

However, simply having students collaborate on mathematical problems does not guarantee the activation of HOTS. Schoenfeld (1985) argues that the reasoning used during problem solving depends not only on task characteristics but also on the interplay between the task, the solver, and the collaborative context. The design of CPS tasks must carefully balance challenge and support to ensure active engagement and productive peer interactions.

What are specific issues for CPS in mathematics education? While many CPS challenges, such as coordinating dialogue and balancing group contributions, are not unique to mathematics, several issues take on a distinctive form within this domain. As Barron (2003) highlights, the quality of interaction among students strongly influences both group performance and individual learning outcomes, yet initiating and sustaining productive dialogue often proves difficult. In mathematics, these challenges are amplified by the discipline's reputation for requiring innate ability (Boaler, 2016), which can deter less confident students from contributing and lead to uneven participation. Moreover, students often lack the meta-knowledge to collaborate effectively in mathematics, where joint reasoning involves abstract concepts, formal representations, and shared problem structures (Goos et al., 2002; Walen et al., 1999). Group work may also reflect motivational disparities: high-achieving students may perceive collaboration as unfair when group assessments do not distinguish individual contributions (Tatsis & Koleza, 2006), especially in a subject where correctness is heavily emphasized.

A further problem in CPS lies in the initial discomfort students have in articulating their mathematical reasoning. As Fitzsimons and Ní Fhloinn (2023) observed in their intervention, students had difficulty explaining their solutions to peers and often experienced problems in reaching a shared understanding. These findings highlight the importance of developing mathematical communication skills as a core component of effective CPS. Here we see a clear role for socio-mathematical norms, which shape how students approach, evaluate, and communicate mathematical arguments (Yackel & Cobb, 1996).

Jarry-Shore and Anantharajan (2015) state that CPS in mathematics presents not only cognitive challenges, but also significant social and interpersonal difficulties. Research shows that students often struggle to communicate their ideas, reach consensus, or feel heard by their peers. These difficulties can impede meaningful engagement with mathematical content. Effective CPS therefore requires attention not only to the complexity of the task, but also to the relational dynamics within student groups.

Regarding the choice of tasks for CPS in mathematics, Canogullari and Radmehr (2025) note in their literature review that various authors assert several pedagogical challenges. One key issue is creating tasks with a 'low floor and high ceiling', that is, tasks that are accessible to all students while still offering depth for advanced learners. Additionally, while real-life contexts can increase engagement, they may also complicate students' understanding of the mathematical objectives. Another frequently reported difficulty is ensuring that the mathematical goals of a task are made explicit to guide student learning effectively. Finally, aligning tasks with instructional goals requires careful attention to multiple, often competing dimensions of classroom practice, such as cognitive demand, interactional structure, and curriculum alignment.

While a substantial body of research has examined collaborative learning, particularly in relation to task features, group composition, and interactional procedures, many of these studies have addressed task features, group composition, or interactional procedures in isolation or focused primarily on social or motivational outcomes. According to Canogullari and Radmehr (2025), for example, various authors highlight challenges in aligning task complexity, accessibility, and instructional goals, yet offer limited integration of these insights with cognitive outcomes. Our review aims to refine this understanding by explicitly linking task design and CPS processes to the development of HOTS. Using a conjecture-based framework, we synthesise how specific design features and group protocols activate key cognitive mechanisms during collaboration. In doing so, we offer an integrative perspective that extends existing literature by foregrounding the interplay between collaborative structure and cognitive depth in mathematics education.

By focusing on the design features that underpin CPS, this review aims to provide a comprehensive framework for understanding how to structure collaborative work that foster both cognitive and collaborative skills. Ultimately, this work seeks to bridge the gap between research and practice, offering actionable insights for educators and researchers alike.

Based on the outlined challenges and gaps in the literature, this review addresses the following research question:

What design features and processes of collaborative problem solving foster higher-order thinking skills in mathematics education according to research literature?

2. Theoretical framework

2.1. Key definitions and concepts

To provide a solid theoretical foundation for this review, this section defines key concepts related to CPS and HOTS. We will analyse these processes and outcomes using Sandoval's (2014) model, which will be introduced in the next section. This model connects design features to cognitive processes, in this case of CPS, and learning results.

According to Robbins (2011), problem solving in general refers to a sequence of actions aimed at achieving a specific outcome defined by an instructor, text, or the learner, within a set of given parameters. Lester and Kehle (2003) describe problem solving as an activity that requires students to actively use their prior knowledge and experiences. This involvement is not passive; it requires coordinating what they already know with their ability to use known representations and inference patterns in addition to their intuition. The ultimate goal of this process is to create new representations that help resolve the tensions or ambiguities inherent in the problem that prompted the problem-solving activity.

While problem solving is a fundamental cognitive activity across various contexts, its role in mathematics education is particularly significant due to the discipline's emphasis on logical reasoning, abstract thinking, and the application of knowledge to novel problems. This perspective underlines the dynamic and constructive nature of problem solving in mathematics, and emphasizes its role in facilitating deeper learning and understanding. Stanic and Kilpatrick (1988) emphasized that problem solving in mathematics education has been embraced as a multifaceted concept, embodying different educational, philosophical, and practical viewpoints. This diversity underlines the complex nature of mathematical problem solving as both a pedagogical goal and a field of study. We use in this study the definition from Lester and Cai (2016): a mathematics problem is a non-routine task presented to students in an instructional setting that poses a question to be answered but for which the students do not have a readily available procedure or strategy for answering it (p. 8). We choose to use this general definition of problem solving in our theoretical framework because one of the parts of the analysis is to identify the types of problems that have been used in the different articles.

To further enhance the effectiveness of problem solving in mathematics, numerous studies highlight a structured approach that integrates interaction, negotiation, and shared reasoning among students, fostering deeper engagement and HOTS. Such approaches emphasize specific types of interactions that facilitate problem solving, including symmetry among participants, ensuring balanced and equitable contributions, and shared goals, which often require ongoing negotiation throughout the collaborative process. Another critical aspect is interactivity, where mutual regulation and real-time adjustments influence each participant's cognitive processes. Moreover, processes such as negotiation, explanation, and conflict mediation jointly stimulate deeper understanding and learning (Dillenbourg, 1999).

In this study, we adopt a conceptually grounded definition of CPS, emphasizing that it is not merely joint task completion but a complex interplay of social and cognitive processes. According to the OECD (2017), CPS is "the capacity of an individual to effectively engage in a process whereby 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" (p. 134). This definition underscores that collaboration is not incidental, but central: it demands mutual dependence, coordinated effort, and joint construction of meaning. Acosta et al. (2025) expand on this view by conceptualizing CPS as comprising two interdependent dimensions: a cognitive dimension (e.g., exploring, planning, monitoring) and a social dimension (e.g., maintaining communication, sharing information, establishing shared understanding, and negotiating). These social interactions are not only organizational but also epistemic, they shape how learners co-construct, adapt, and transform knowledge. In this way, CPS extends beyond individual problem solving by enabling a dynamic form of reasoning that is distributed, dialogic, and responsive.

Building on the collaborative processes inherent in CPS, we will define and contextualizes HOTS by exploring the key terms and cognitive processes relevant to their activation and development in mathematics education. According to He et al. (2017), CPS significantly supports the development of HOTS in mathematics education by engaging students in reasoning, analysis, and creative thinking. Collaborative interactions foster the articulation of ideas, exchange of perspectives, and critical evaluation, deepening understanding of mathematical concepts and problem structures. CPS encourages students to apply HOTS, including logical reasoning, problem formulation, and reflection, while enabling them to regulate efforts, monitor progress, and adapt strategies (Graesser et al., 2020). Research also shows that CPS enhances analytical thinking and mathematical reasoning (Jeannotte & Kieran, 2017). Students navigate strategies, reconcile ideas, and transfer mathematical knowledge to novel situations by working collaboratively, reinforcing higher-order cognitive abilities. The added value of collaboration lies in its power to externalize thought processes, facilitate regulation and feedback, and promote adaptive learning strategies, all essential components of higher-order thinking.

The terminology used in the context of CPS in mathematics related to HOTS is varied, encompassing key concepts such as reasoning, problem solving, analytical thinking, and mathematical thinking, each requiring clear definitions to understand their role in fostering higher-order cognitive processes. HOTS encompass advanced cognitive processes that go beyond memorization and recall, including critical thinking, problem solving, analysis, synthesis, evaluation, creativity, and metacognition. These abilities involve complex activities such as analysing information, generating hypotheses, evaluating evidence, and applying knowledge to new situations (Makmuri et al., 2021). As an umbrella term, HOTS integrates a range of advanced cognitive skills, particularly relevant in CPS in mathematics, where terms such as analytical thinking, problem solving, reasoning, and mathematical thinking are frequently used (Jeannotte & Kieran, 2017; Schoenfeld, 1992; Yackel & Hanna, 2003). Clear definitions of these terms and their interrelations are critical for understanding their role in fostering HOTS.

Analytical thinking involves dissecting problems, innovating ideas, and applying knowledge effectively (Amer, 2005). It requires breaking down situations to better understand them, comparing alternatives, and evaluating attributes (Sternberg, 1985). Mathematical thinking, as Stacey (2006) defines it, encompasses creating strategies for solving problems, recognizing patterns, and forming arguments and proofs. Burton (1984) emphasizes its universal applicability, noting that processes like specializing, conjecturing, generalizing, and convincing are integral to mathematical thinking. These processes also connect deeply with problem solving, which Lester and Cai (2016) define as tasks without readily available procedures, requiring students to integrate prior knowledge and intuition.

Reasoning, crucial for problem solving and analytical thinking, involves drawing conclusions to achieve goals (Leighton &

Sternberg, 2004; Robbins, 2011). Jeannotte and Kieran (2017) distinguish between the structural and process aspects of reasoning in mathematics, the latter including conjecturing, generalizing, and proving. Creative mathematically founded reasoning, as defined by Boesen et al. (2010), reflects innovative strategies rooted in deep understanding, making it a core element of problem solving.

According to Mason and Johnston-Wilder (2004), mathematicians use a variety of processes and actions when working on mathematical problems. They provide an extensive list that includes: 'exemplifying, specializing, completing, deleting, correcting, comparing, sorting, organizing, changing, varying, reversing, altering, generalizing, conjecturing, explaining, justifying, verifying, convincing, and refuting' (p. 109).This comprehensive list includes many of the processes previously described, representing a complete set of actions that we anticipate identifying as enactment processes and outcomes in the corpus yet to be analyzed. Therefore, these processes encapsulate the essence of HOTS in mathematics. Several processes, such as exemplifying, comparing, explaining, and conjecturing, function as both indicators and enablers of HOTS in mathematics (Mason & Johnston-Wilder, 2004). These actions foster deeper learning by enabling students to articulate, refine, and evaluate mathematical ideas. Consequently, the interplay between these processes encapsulates the essence of HOTS, emphasizing their dual role as both outcomes of CPS and drivers of productive collaboration.

The defined terms, reasoning, problem solving, analytical thinking, and mathematical thinking, are interconnected and collectively contribute to fostering HOTS in mathematics education. HOTS serve as an umbrella term encompassing advanced cognitive processes such as critical thinking, evaluation, and creativity, which are crucial for tackling complex tasks. Analytical thinking emphasizes breaking down problems and evaluating alternatives, while mathematical thinking involves creating strategies, recognizing patterns, and forming arguments, making it integral to problem solving. Reasoning acts as the foundational mechanism linking these concepts, enabling students to draw conclusions, justify solutions, and critically evaluate mathematical ideas.

2.2. Structuring the review by Sandoval's model

The Sandoval model (2014) provides a clear framework for analysing how task design, cognitive processes, in this case related to CPS, and learning outcomes are connected in education. This model is particularly relevant for this review, as it helps systematically identify conjectures, design features, processes, and outcomes in the selected studies of CPS in mathematics. The model helps address the dual role of cognitive processes, which act both as processes and as observable outcomes in CPS tasks. Using Sandoval's conjecture map, we analyse the selected articles to highlight key design features and processes of CPS that engage students in HOTS. It also supports the idea that engaging students in problems requiring reasoning and mathematical thinking fosters the development of these skills.

As shown in Fig. 1, Sandoval's model begins with *high-level conjectures*, which outline principles for supporting learning. These guide the formulation of *design conjectures*, connecting features of a learning environment (e.g., tasks, tools) with cognitive processes they aim to activate, such as reasoning or reflection. For instance, 'If students engage with this task using these tools, reflective reasoning will occur.' *Theoretical conjectures* link these processes to broader outcomes, such as improved problem-solving or argumentation skills. While this phrasing is adapted to the context of problem solving, it aligns with Bakker's (2018) general description of conjecture mapping. By distinguishing between design and theoretical conjectures, conjecture mapping clarifies the pathways from task design to learning outcomes. In Fig. 2, we have elaborated an example based on the article by Granberg and Olsson (2015).

Through the lens of Sandoval's model (Sandoval, 2014), we identified a set of terms that align with our theoretical framework. In addition to reasoning and mathematical thinking, these terms include various related processes: illustrating, specializing, supplementing, deleting, correcting, comparing, sorting, organizing, changing, varying, reversing, modifying, generalizing, guessing, explaining, justifying, verifying, convincing, refuting, proving, and arguing. These processes besides reasoning, mathematical and analytical thinking have been found in the literature as underlying processes, but they can also be regarded as outcomes.

Before introducing our sub-questions, it is important to note how we interpreted and applied Sandoval's model in our analysis. It is important to note that this model was chosen because it offers a structured yet flexible lens through which to analyse the layered nature of design-based classroom interventions. It allowed us to systematically trace the logic of each study from theoretical rationale through practical implementation and observable effects. We are aware, however, that in real classroom settings, especially those involving CPS in mathematics, some categories may conceptually overlap. For example, discursive practices such as encouraging students to verbalize their thinking could be both a design feature (as in Tatsis & Koleza, 2006, where students were instructed to "think aloud") and an observed process (e.g., when students spontaneously justified or negotiated ideas during interaction). In such cases, we examined the rhetorical function and location of the passage in the original article: when the practice was described as an intentional part of the lesson design, it was coded as a design feature; when it appeared in the results or analysis as an observed behaviour, it was coded as a process. Another recurring example is reasoning, which could either be framed as a goal of the design (outcome), a practice supported by task structures (process), or even a guiding assumption (high-level conjecture). For instance, in Granberg and Olsson

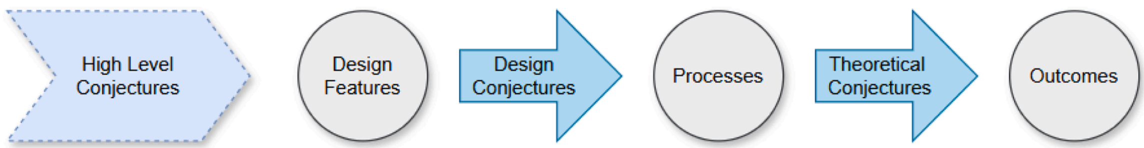

Fig. 1. Generalized conjecture map.

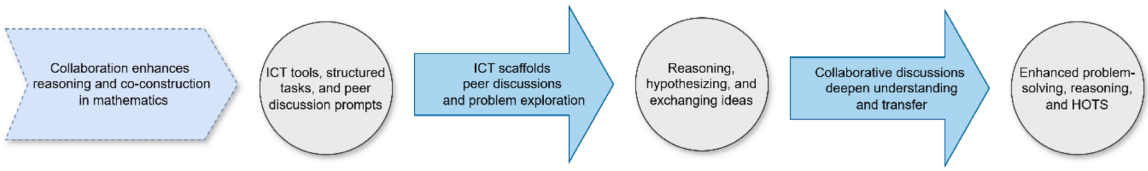

Fig. 2. Conjecture map with examples from CPS and HOTS in mathematics education.

(2015), reasoning was addressed as a learning goal (outcome), described in the results as collaborative activity (process), and justified theoretically as a rationale for using GeoGebra (high-level conjecture). In such multi-layered cases, we consistently assigned the code that best reflected the role that element played within the logic of the article.

Specific design features in educational environments that, according to the authors, promote HOTS include open-ended problems, think-aloud protocols, and so on. Although our research question focuses on the design features and processes of CPS that promote HOTS, we structure our results section around the components of the Sandoval model and have therefore formulated the following sub-questions:

SQ1. Which conjectures are formulated in research literature?

SQ2. Which design features are used in the various experiments reported in the articles?

SQ3. Which processes are observed during collaborative problem solving?

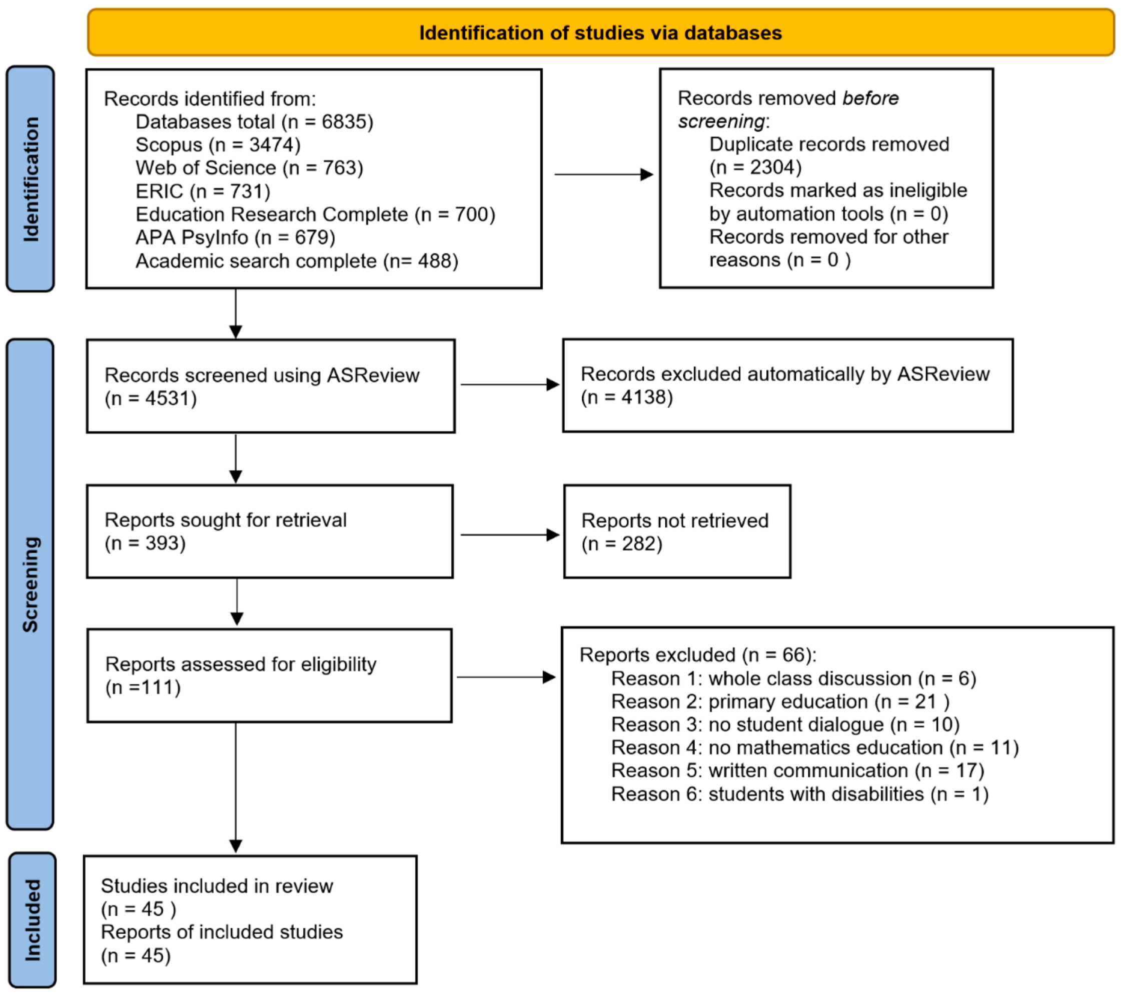

Fig. 3. PRISMA Flow diagram of the search, screening, and inclusion processes.

SQ4. Which outcomes are reported related to higher order thinking?

Through organizing our work around these sub-questions, we ensure a systematic examination of each component in the Sandoval model.

3. Methods

A systematic literature review (concluded in November 2023) was carried out following the methods outlined by Petticrew and Roberts (2006). This involved searching the literature using a predetermined process, selecting studies based on predefined criteria, and then extracting and synthesizing the data. We also followed general procedures consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The study selection process is shown in Fig. 3. Our search strategy comprised three key elements: HOTS, group collaboration, and the domain of mathematics, specifically targeting secondary and tertiary education levels. Keywords were sourced from a database thesaurus and key articles identified in the field. To hone our search, pilot searches were conducted by experimenting with various keyword combinations within titles, abstracts, and keyword sections across databases, ensuring the identification of pertinent articles. We initially tested our search string on a set of 100 articles, which two independent coders evaluated and categorized as "include," "maybe," or "exclude." The intercoder reliability was insufficiently high. Through discussion, a consensus was reached on which articles to include, and the search string was refined. The inclusion and exclusion criteria were also sharpened to focus on peer discussions among students aged 12 and above, excluding whole-class discussions, teacher interventions, and analyses of written dialogues.

In the final search, the following search string was used:

("analy* think*" OR "high* order think skills" OR reasoning OR "analytical competence" OR "analy* ability" OR "problem solving" OR "math* think*")

AND

("peer discussion" OR "peer dialog*" OR collabor* OR coopera* OR "classroom dialog*" OR "classroom discourse" OR "collaborative argumentation" OR "collaborative dialog*" OR "cooperative dialog*" OR "classroom discourse" OR "co-construction" OR "think* aloud" OR "small-group discussion*" OR "small-group conversation*" OR "dialogic pedagogy" OR "math* discourse" OR "peer interaction" OR "peer instruction" OR "dialogic Interaction")

AND math*

NOT ("primary school" OR "elementary school").

Table 1 provides an overview of the databases used, the associated filters and limits, and the numbers of articles found.

The 6835 articles identified were processed through the Rayyan program (Ouzzani et al., 2016) to eliminate duplicates, leaving 4531 articles. To increase the efficiency of literature screening, we used the ASReview software. This software, which is used in many review articles, uses machine learning algorithms (van de Schoot et al., 2021), to screen literature. Using active machine learning, ASReview assesses the relevance of all literature in real time based on the annotation of articles according to the filter and inclusion criteria used by the researchers. Examples of relevant and irrelevant articles used to train the software are shown in Table 2. In this way, ASReview can immediately put forward the most relevant literature for priority annotation. We chose that If 100 (Boetje & van de Schoot, 2024) consecutive articles are marked as not relevant, it can be assumed that the remaining literature is also not relevant. At that point, the screening of the literature can be stopped (Bloodgood & Vijay-Shanker, 2014).

A member of the research team screened a total of 393 articles using ASReview, making initial selections based on the abstracts to determine their potential relevance, categorizing them as either to be included for further review or excluded. Based on the title and abstract, 282 articles were excluded. The remaining 111 warranted a more thorough examination due to uncertainties regarding their relevance; upon detailed review, it was determined that 66 out of these 111 articles did not meet the inclusion criteria and were deemed irrelevant to our study's objectives. Since only one reviewer conducted the screening, no separate consistency measure was calculated. The iterative learning process of the tool helped ensure that relevant articles were prioritized in a consistent manner. The 45 selected articles form the corpus of the further analysis, were stored in Atlas Ti and are summarized in Table 3.

3.1. Extraction process

Initially, data including study title, authors, publication year, journal, country, and education level were collected to describe the corpus. For the research question, information on conjecture maps, based on the theoretical framework was extracted from relevant sections (van der Linden et al., 2022). High-level conjectures were mainly derived from study goals or, if absent, by analysing the

Table 1
Databases, filters and resulting frequencies.

Database	Filters and limits	N
Scopus	Title, Abstract, Keywords	3474
Web of Science	Title, Abstract, Keywords Plus	763
ERIC	Title, Abstract, Keywords, Peer reviewed	731
Education Research Complete	Title, Abstract, Peer reviewed	700
APA PsycInfo	Title, Abstract, Keywords, Peer reviewed	679
Academic search complete	Title, Abstract, Keywords, Peer reviewed	488

Table 2
Examples of the articles used to train the software.

Examples of relevant papers	Examples of irrelevant papers
Bjuland, 2007	Abrahamson, Blikstein, & Wilensky, 2007
Goos et al., 2002	Abulencia et al., 2012
Hansen, 2022	Adamson et al., 2003
Hoek & Seegers, 2005	Aerts, n.d.
Mueller, 2009	Aggarwal, 2004

theoretical framework. Method sections provided design features, while results sections detailed implementation processes. Occasionally, artifacts' information came from the method section. Outcomes were determined through analysis of article conclusions, including findings not directly related to our objectives, which were also considered.

3.2. Data analysis

The data analysis consisted of two phases. Initially, three researchers independently analysed a subset of ten articles to identify key codes, which led to the development of a preliminary codebook. Inter-rater reliability was not calculated, but discrepancies were discussed extensively during consensus meetings. Through these discussions, definitions were clarified and the codebook was refined to improve consistency and clarity among the coders. This collaborative and iterative process established the foundation for coding the full dataset. The coding categories were based on Sandoval's conjecture mapping framework, comprising high-level conjectures, design features, mediating processes, and outcomes. Text relevant to each category was segmented into meaningful units, ranging from paragraphs to sentences or fragments, each representing a single idea. Descriptive codes were assigned to these units and organized into concept code groups. During early rounds, some conceptual ambiguity was encountered, particularly in distinguishing between processes and outcomes. For instance, elements such as "reasoning" or "mathematical thinking" could either represent in-the-moment learning processes or post hoc learning outcomes. To resolve this, coders considered not only the content but also the location and rhetorical function of each segment within the original article. For example, "reasoning" discussed in the methods section was coded as a process, whereas its appearance in the results section as an observed effect was coded as an outcome. This approach helped ensure fidelity to the internal logic of each article. Coding continued through iterative rounds using Atlas.ti software (2023), with the codebook updated as needed. Coding saturation, indicated by the emergence of no new codes or category shifts, was observed after analysing 12 articles and confirmed after an additional 20 articles. Ultimately, the framework was applied consistently across all 45 empirical classroom studies, enabling the exploration of both the frequency and relationships of conjecture map elements across the dataset.

3.3. Description of the corpus

Articles included in the review are indicated with an asterisk (*) in the bibliography. Fig. 4 shows the articles' publication years and countries of origin.

4. RESULTS

The results are organized according to the categories described in Sandoval's (2014) model and are shown in Fig. 1: high-level conjectures, design features, design conjectures, processes, theoretical conjectures, and outcomes.

Tables 4 to 6 show the codes for each subcategory and their frequencies. In addition, we describe the different components of the

Table 3
Studies included in this review.

(No) Authors		
(1)Lai and White (2012)	(16) Aksu and Zengin (2022))	(31) Zhang et al. (2022)
(2) Granberg and Olsson (2015)	(17) Demir and Zengin (2023))	(32) Abdu and Schwarz (2020)
(3) Kerrigan et al. (2021)	(18) Carlsen (2010)	(33) Hansen (2022)
(4) Bleiler-Baxter et al. (2023)	(19) Walter and Barros (2011)	(34) Kumpulainen and Kaartinen (2003)
(5) Schwarz and Linchevski (2007)	(20) Prusak et al. (2012)	(35) Hansen and Naalsund (2022)
(6) Kieran (2002)	(21) Francisco (2013)	(36) DeJarnette (2022)
(7) Bjuland (2007)	(22) Nieminen et al. (2022)	(37) Haataja et al. (2022)
(8) Walen et al. (1999)	(23) Yemen-Karpuzcu et al. (2017)	(38) Goos (1994)
(9) Mevarech and Kramarski (2003)	(24) Hoek and Seegers (2005)	(39) Abdullah et al. (2017)
(10) Szabo et al. (2024)	(25) Tatsis and Koleza (2006)	(40) Kim and Moore (2019)
(11) Keh et al. (2019)	(26) White (2006)	(41) Goos and Galbraith (1996)
(12) White et al. (2012)	(27) Lantz-Andersson (2009)	(42) Goos et al. (2002)
(13) Moate et al. (2021)	(28) Oner (2013)	(43) Goos (2002)
(14) DeJarnette and González (2015)	(29) Huang et al. (2023)	(44) Mueller (2009)
(15) DeJarnette et al. (2014)	(30) Chan and Clarke (2017)	(45) Sivaraj et al. (2021)

model and refer to one article for each category as an illustrative example.

Table 5

4.1. High level conjectures, design conjectures and theoretical conjectures related to HOTS

In addressing SQ1, we discuss High-Level Conjectures, which according to Sandoval (2014) High-Level Conjectures are broad, theoretical ideas about how to support learning in a particular context, typically derived from prior research or problem analysis. Design Conjectures are hypotheses about how specific design elements (tasks, tools, roles) will generate mediating learning processes. Theoretical Conjectures are hypotheses about how these mediating processes result in desired learning outcomes.

In most of the reviewed articles, the three types of conjectures, high-level, design, and theoretical, are distributed across different sections. High-level conjectures typically appear in the introduction, as they provide overarching principles that frame the study and set its broader goals. These conjectures often outline the intended educational impact or the theoretical rationale for the intervention. Design conjectures are usually found in the methods section, where the specific features of the designed learning environment are described. This includes details about tasks, tools, or activities that are expected to activate specific cognitive processes related to CPS. Theoretical conjectures are most often addressed in the results or discussion sections. These conjectures link the observed processes to the intended outcomes, reflecting on how the design and enactment of the intervention influenced learning. We have listed the types of conjectures related to HOTS in Table 4, these categories may overlap.

We identified high-level conjectures, design conjectures, and theoretical conjectures related to HOTS in various articles. Notable examples of these conjectures, particularly those associated with **reasoning and mathematical thinking**, are found in Granberg and Olsson (2015). The authors present the following overarching assumption (high-level conjecture): GeoGebra supports students' collaborative creative reasoning by providing a shared working space and feedback that becomes the subject of students' reasoning processes (p. 49). This assumption is then translated into a specific design conjecture: The design of tasks using GeoGebra should ensure that students work on non-routine, intellectually challenging problems to encourage collaboration and creative reasoning. These tasks require adidactical situations where the teacher minimizes direct guidance, allowing GeoGebra to offer feedback (p. 53). Finally, the authors link the processes and outcomes through a theoretical conjecture: Collaborative creative reasoning during problem solving with GeoGebra leads to enhanced conceptual understanding and joint problem space maintenance, fostering skills in mathematical reasoning and effective collaboration (p. 60).

Another example comes from Mevarech and Kramarski (2003), where the authors explicitly address **metacognition** in their conjectures. They propose the following high-level conjecture: *Mathematical reasoning and achievement can be enhanced in cooperative learning settings through structured approaches that focus on reflection, monitoring, and metacognitive questioning, as these encourage deeper engagement with the material and long-term knowledge retention* (p. 451). Design Conjecture: *Students in cooperative learning groups benefit more when they are guided to actively use metacognitive questioning* (*e.g., comprehension, connection, strategy, and reflection questions*) *rather than passively studying worked-out examples, as the questioning promotes understanding of relationships between prior knowledge and new problems* (p. 453). Theoretical Conjecture: *Metacognitive questioning helps students develop better mathematical reasoning by encouraging them to reflect on and monitor their problem-solving processes. This leads to higher-quality verbal explanations, algebraic representations, and solutions compared to those who rely solely on worked-out examples* (p. 451).

We observed that several articles mention the terms **'mathematical understanding' or 'mathematical learning'** within their conjectures.' Consequently, we have categorized these articles together. For example, according to Keh et al. (2019) High-Level Conjecture: Open-ended mathematical problems encourage students to use multiple approaches to solve problems, fostering their creativity and **mathematical understanding** (p. 2). Design Conjecture: Tasks designed with open-ended problems should allow for diverse strategies and solutions to cultivate mathematical understanding and creativity (p. 2–3). Theoretical Conjecture: Engaging students in open-ended mathematical problem-solving processes leads to enhanced HOTS, including flexibility, fluency, and novelty in

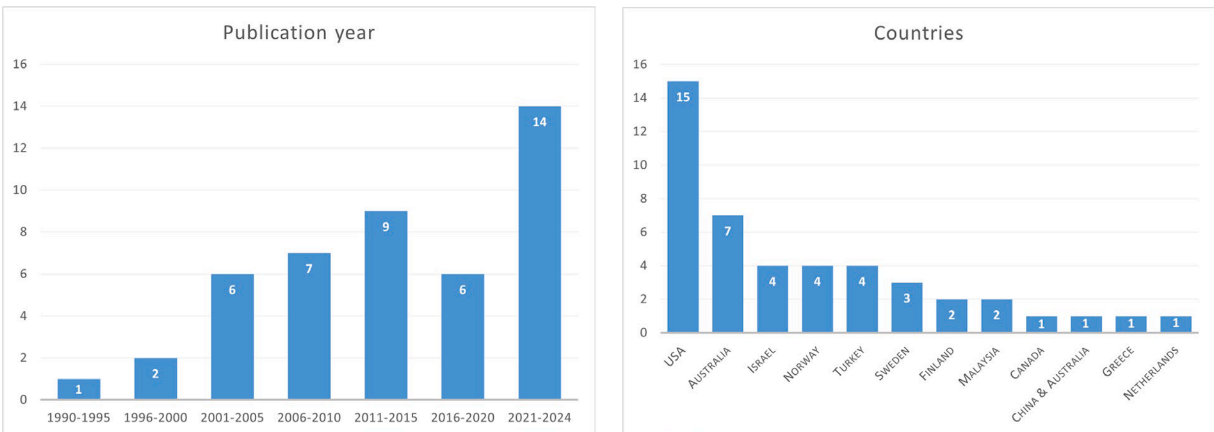

Fig. 4. Temporal and geographic distribution of the corpus.

Table 4
Types and frequencies of higher-level conjectures in the corpus classified by categories.

Category	Freq.
Reasoning	28
Mathematical Thinking	13
Metacognition	8
Mathematical Understanding or Learning	12

Table 5
Types and frequencies of Tools and Materials in the corpus classified by subcategories.

Category	Description	Freq.
Tools and materials	Dynamic Computer Environment	15
	Calculators	5
Task Structures Related To Assignment	Open-Ended Problems	18
	Genuine, Unfamiliar, Challenging	13
	Exploratory And Inquiry-Based Using Technology	15
	Creating Cognitive Conflict	5
Task Structures Related To The Subject	Other General Subjects (e.g. Arithmetical Problems)	12
	Multiple Mathematical Subjects	5
	Geometry	16
	Algebra (Including Functions)	12
Task Structures Related To Collaboration	Working In Small Groups	29
	Working In Pairs (Dyads)	19
Participant Structures	Vocational Education	1
	Tertiary Education	11
	Secondary Education	33
	Inexperienced on Similar Tasks	2
	Higher and/or lower achievers	4
	Experienced On Similar Tasks	2
Discursive Practices	Think Aloud (verbalizing ideas)	10

problem solving (p. 16–18).

4.2. *Design features*

In addressing SQ2, we begin this section with a reflection on the role of the teacher within Sandoval’s (2014) model. We begin this section with reflection on the role of the teacher within Sandoval’s model (2014). It is well known that teachers play a crucial role in shaping the design of a learning environment, including how tasks are structured and how classroom discourse is facilitated. Their actions and decisions directly influence the outcomes that the design aims to deliver. In this review, we focus on articles in which students collaborate and communicate with each other, rather than on teacher-student dialogues. Therefore, we chose not to consider teacher actions as part of embodiment.

According to Sandoval (2014), design features are the practical implementation of high-level conjectures within a learning environment. These include tools and materials such as designed resources like software or manipulatives; task structures such as the goals and expected procedures for learners; participant structures including the roles and interaction patterns of students; and discursive practices such as the expected ways of speaking and reasoning. Below, the various design features in CPS are elaborated.

4.2.1. *Tools and materials*

In all the experiments outlined in the articles, writing utensils were employed; thus, we opted not to list them as materials. Besides pen and paper, two categories of tools were utilized. The first encompasses various dynamic computer environments that facilitated collaborative work among students, primarily in exploratory tasks. In the study of Granberg and Olsson (2015) GeoGebra was used for constructing and adjusting algebraic formulas and visualizing them graphically. The second tool is the calculator, which was sometimes used just to calculate something as in Kieran (2002) ‘He expressed the calculation he entered into the calculator: “One hundred and fifty divided by ten.” ‘ (p.204). Calculators were used in other cases to check answers as in Goos (1994): ‘Despite their previous experience with trigonometry, the students continued until their calculator’s "Error" message alerted them to their mistake’ (p. 161).

4.2.2. *Assignments*

Open-Ended Problems were commonly used, these problems vary between ’riddles’ such as those used in the article written by (1999), where the specific task involves a well-known mathematical problem: three men sharing a hotel room, leading to a paradoxical missing dollar. In other cases there were real-world problems where the students had to find a solution to the real world problem as described in the article of Sivaraj et al. (2021), where students collaboratively design a laser security system for a travelling museum exhibit, applying knowledge of light properties and geometry. Some articles use the terms genuine, unfamiliar and challenging to characterize the problems. These terms have been used together or separately, for example by Goos (1994): ‘For the purposes of this study, a task is considered to be a genuine problem for the student…’(p. 146). An good example of a task of the type ‘exploratory and inquiry-based using technology’, can be found in the study by Lai and White (2012). It features a small-group computing setting in which each student manipulates a distinct point within a collective geometric space, collaboratively creating a quadrilateral. Another category identified was tasks in which students are confronted with situations that cause a cognitive conflict, namely that they are placed in uncertainty by encountering new information that contradicts existing beliefs or knowledge. For example in Kumpulainen and Kaartinen (2003): ‘Students are faced with cognitive and social conflicts that are left unresolved‘ (p. 342).

4.2.3. Mathematical subject

The articles used tasks within many different mathematics domains, most of them within geometry (Oner, 2013), and algebra (DeJarnette & González, 2015). But we have created two other categories for cases where, for example, experiments were conducted for a long time and therefore many topics were covered (multiple mathematical subjects) (Francisco, 2013), and a group in which we put the tasks that are not directly linked to a particular mathematical topic (other general subjects) (Mueller, 2009).

4.2.4. Group sizes

While all studies involved small student groups, it was observed that some articles featured authors working with pairs (Kumpulainen & Kaartinen, 2003). This distinction led us to consider it important to separate these two categories.

4.2.5. School level and student characteristics

The participants structures include the school level, secondary (Nieminen et al., 2022), tertiary (Yemen-Karpuzcu et al., 2017) and vocational (Hoek & Seegers, 2005) education. We also deemed the students’ prior experience with group work and the nature of the assignments significant. Consequently, we categorized them as either ’Experienced on Similar Tasks’ or ’Inexperienced on Similar Tasks’. One example is the article of Szabo et al. (2024), where the participants are described as students that were already experienced in CPS using whiteboards from previous years. In the study of DeJarnette and González (2015), where students were unaccustomed to problem-based instruction and collaborative work. Furthermore, in some articles a distinction is made between higher and/or lower achievers, so this group includes a study where performance was measured for these two types of students. For example, Carlsen (2010) focused on upper secondary school students in Norway, specifically in a voluntary course called 3MX, designed for high achievers preparing for university-level mathematics.

4.2.6. Discursive practices

In terms of discursive practices, we have a category that we use to identify the articles where students where explicitly asked to think aloud and share ideas (Think Aloud). It is clear that in all cases students are speaking to each other, but it seems important to make this distinction. For example: ‘The only instructions given to the students were that they should verbalize every thought they make and…’ (Tatsis & Koleza, 2006, p. 447).

4.3. Observed processes in CPS

In addressing SQ3, we examine the mediating processes in CPS. According to Sandoval (2014), mediating processes are the interactions and activities that arise from the design and are hypothesized to lead to learning. We focus on the processes related to HOTS. In Table 6 we have distinguished different categories: reasoning underlying processes (exemplifying, specializing, completing, deleting, correcting, comparing, sorting, organizing, changing, varying, reversing, altering, generalizing, conjecturing, explaining, justifying, verifying, convincing, and refusing ’) monitoring, regulation, planning and evaluation. Since our goal is to highlight both the design features and the processes that help promote HOTS, we choose here to also pay attention to the processes related to student collaboration. In this context, the authors of the revised articles refer to the following processes: verbal/non-verbal, interactions, exploratory talk, discussion and negotiation.

Table 6
Occurrences of enactment processes.

Category	Description	Freq.
Processes Related to HOTS	Reasoning	17
	Underlying Processes	43
	Monitoring	4
	Regulation	1
	Planning	2
	Evaluation	8
Processes Related to Collaboration Skills	Verbal/ Non-Verbal Interactions	5
	Exploratory Talk	1
	Discussion	8
	Negotiation	16

4.3.1. Processes related to HOTS

Leighton and Sternberg (2004) broadly define reasoning as the process of drawing conclusions in order to achieve goals, numerous authors within the corpus indeed describe reasoning as a process. For example this quote in Granberg and Olsson (2015): 'However, based on the main part of the students' dialog, that is, their arguments and the extent to which their propositions were anchored in mathematical concepts, the study shows that creative reasoning was traceable within all student groups' (p. 60). In developing our theoretical framework, we aimed to deeply understand the processes central HOTS in mathematical problem-solving contexts. While certain researchers in the corpus have not identified reasoning as a noticeable process, they have highlighted other processes like explaining or justifying; Kieran (2002) report that 'But even though the computer was facilitating the public discourse related to the testing of conjectures.' (p. 210). Therefore, we intentionally looked for these practices within the results and conclusions sections of the articles. In the review, two articles did not reference either reasoning or the foundational processes (Lantz-Andersson, 2009; Zhang et al., 2022)

Within the corpus, certain articles focus explicitly on metacognition, such as Mevarech and Kramarski (2003). However, metacognitive processes also surface in other articles. Here are two quotes linked to two of the categories of metacognitive processes found in articles that do not primarily focus on metacognition: Monitoring: 'The dialogue throughout this segment has shown that the three strategies monitoring, questioning, visualising (translating the problem into a visual representation) stimulate the students to become aware of two alternative ways of interpreting distance from a point to a line' (Bjuland, 2007)(p.14). Evaluation: 'Similarly, during small group evaluation of design solutions, students often pushed beyond the scope of criteria and constraints set up by the client and occasionally contemplated additional criteria that led to a greater diversity of outcomes' (Sivaraj et al., 2021) (p.5).

4.3.2. Processes related to collaboration

We aimed to catalogue the most prevalent processes associated with collaboration, though this compilation is not all-

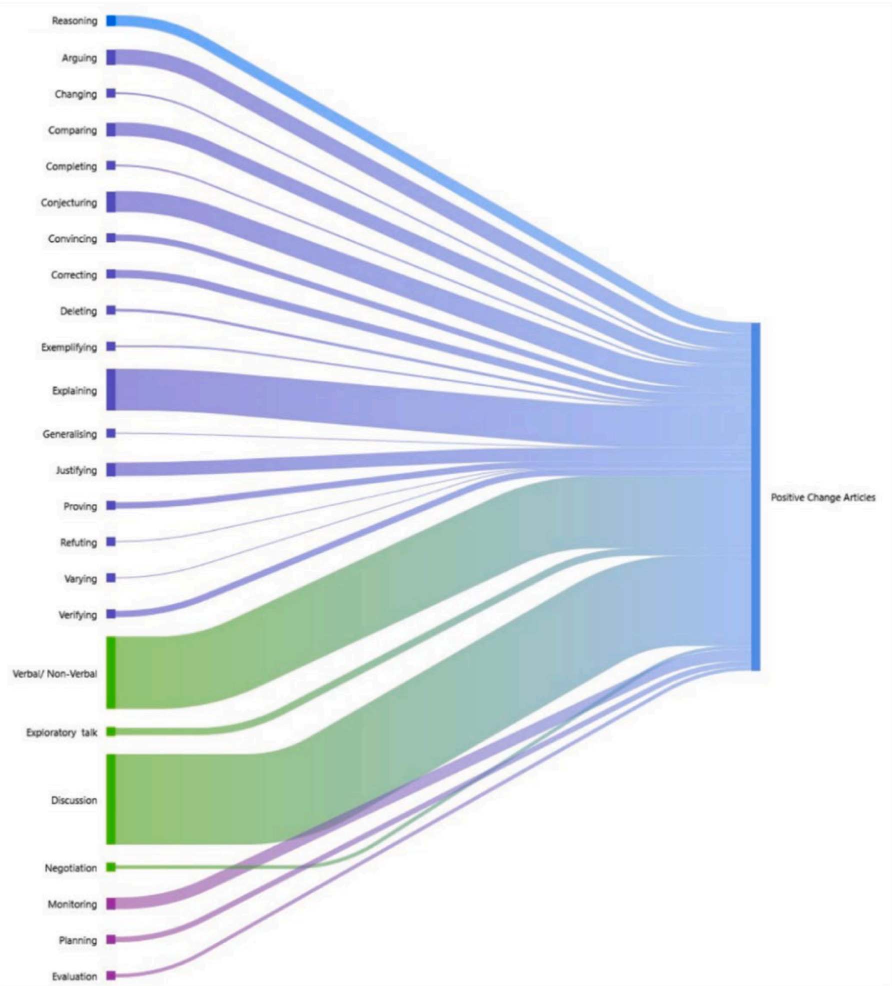

Fig. 5. Extent of processes in articles with positive outcomes for HOTS (group 1) Colours indicate process type: blue = HOTS-related, purple = metacognitive, green = communication.

encompassing, and further terms related to collaboration might emerge from the corpus. We've steered clear of delving into finer details at this juncture, such as the nature of questions posed during peer discussions or the extent of group member participation. These more intricate analyses are reserved for the discussion segment, where we'll explore the processes detailed in articles that reported positive outcomes more thoroughly. Below are several quotes associated with the four categories of processes pertinent to collaboration: verbal/ non-verbal interactions: 'Using both verbal and visual data as the dataset in this study recognises that oral language is contextualised by complex actions and omitting nonverbal actions can distort interpretation'(Moate et al., 2021) (p. 35). Exploratory talk: 'These challenges are justified and sometimes alternative hypotheses are offered. These elements are characteristic of exploratory talk' (Hoek & Seegers, 2005)(p.31). Discussion: 'At this point in the segment the group was still concerned simply with the process of building the triangle and had not yet reached a discussion of the relationship between side lengths and angle measures of the triangle' (DeJarnette, 2022) (p. 259) Negotiation: 'In coding interactions between groups of students we allowed for multiple students to perform K1 moves within a single exchange, as long as they all referred to the same object of negotiation' (DeJarnette & González, 2015)(p.393)

To relate the results concerning the design features to the processes and outcomes, we chose to look at the articles that had conjectures concerning HOTS. We categorized these articles into two groups: group 1) the studies that found or measured a positive effect on HOTS, and group 2) the studies that reported no or minimal change in this area. Figs. 5 and 6 detail the CPS processes observed in experiments in group 1 and group 2, respectively. The thickness of each branch reflects how frequently these processes were observed in the articles' data. Each branch corresponds to a distinct process, and its width is proportional to the number of

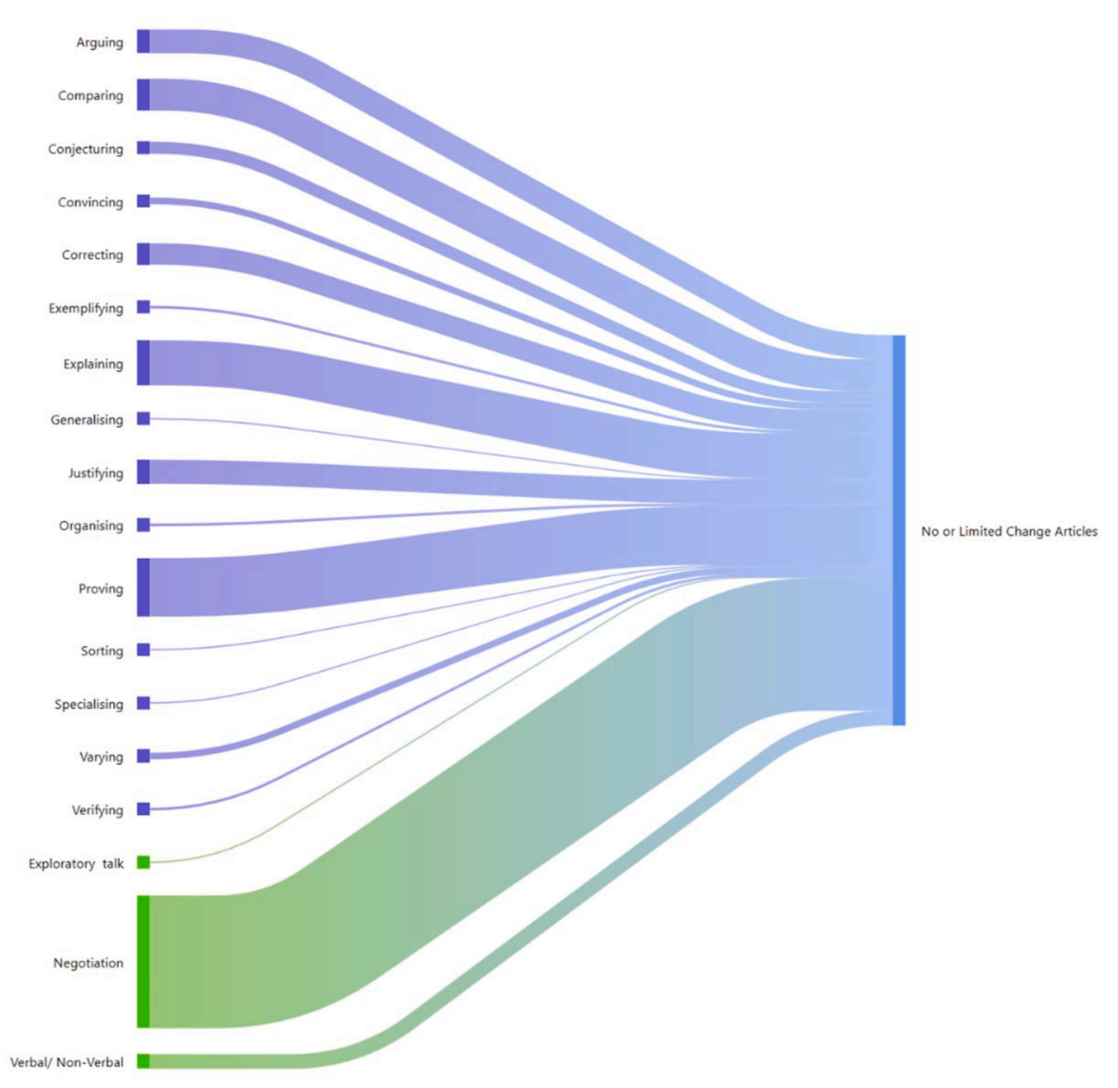

Fig. 6. Extent of processes in articles with 'no or limited' outcomes for HOTS (group 2). Colours indicate process type: blue = HOTS-related, purple = metacognitive, green = communication.

occurrences noted by the researchers. This design offers an immediate visual comparison: wider branches represent processes that appeared more often, while narrower branches indicate those that occurred less frequently. To enhance interpretability, colours were used to distinguish between process types: blue indicates processes related to HOTS, purple represents metacognitive processes, and green refers to communication-related processes. By visually scaling each branch according to its relative occurrence, the diagram highlights both dominant and underrepresented processes at a glance.

What stands out the most is the absence of reasoning and any processes related to metacognition in these articles. Negotiation emerges as the most prevalent, with discussion not being identified at all.

Observing the processes identified in both groups, a few key observations emerge. Firstly, reasoning was not mentioned in any article within group 2, similar to the absence of any codes linked to metacognition, such as planning or evaluation. When examining the underlying processes, the contrast between the two groups is minimal. Secondly, negotiation appeared frequently in group 2′s articles, whereas discussion was notably absent. This is curious because negotiation typically presupposes the occurrence of a discussion. It is worth mentioning that the negotiation predominantly featured in three articles (Chan & Clarke, 2017; DeJarnette, 2022; DeJarnette & González, 2015), where the authors describe negotiation as encompassing discussions. This suggests that in these articles, the emphasis is placed on the negotiation facet within the conducted discussions. In the realm of communication, it's noteworthy that verbal/non-verbal interactions were notably more frequent in group 1 than in group 2.

4.4. Occurrences of outcomes

In addressing SQ4, we consider outcomes, which according to Sandoval (2014) outcomes are the learning goals or changes expected as a result of the design and its mediating processes. In this section, the results of the studies that have conjectures regarding HOTS are presented. We again look at the articles in the two groups mentioned earlier separately: group 1) those that found or measured a positive effect on HOTS, and group 2) those that reported no or minimal change in this area. See Table 7 for more details. Among the articles reviewed, 62 % contain a conjectures associated with reasoning processes. Of these, 64 % document favourable results linked to reasoning or the underlying processes. In the subset where 29 % conjecture on mathematical thinking, 31 % highlight beneficial outcomes. Additionally, among the articles lacking 'high-level conjectures' about analytical thinking, two demonstrated positive results in relation to analytical thinking (Kerrigan et al., 2021; Oner, 2013).

Fig. 7 illustrates the percentage-based relationship between the articles in group 1 and those in group 2, in relation to the design features. Frequently observed features in the positive outcome group include: Working in small groups (67 %), Secondary education (62 %), Tertiary education (52 %), Geometry (48 %), and Exploratory tasks using technology (38 %). In the limited outcome group, the most frequent were: Secondary education (93 %), Open-ended problems (43 %), and Genuine, unfamiliar, challenging tasks (43 %). A closer examination of Fig. 7 reveals substantial differences between the studies in group 1 and those in group 2. Notably, group 2′s experiments featured fewer exploratory tasks involving technology and a substantially reduced focus on geometry. In contrast, these studies had a higher proportion of open-ended problems and challenging assignments. Another noteworthy observation is that in group 2, no study assigned tasks designed to provoke cognitive conflict. The size of the student groups didn't appear to influence the outcomes in group 2, unlike in group 1, where there was a marked preference for group work over pair activities. Regarding educational levels, group 1 exhibited a balanced mix across three tiers, whereas almost all the experiments in group 2 were conducted at the secondary school level, with just one exception. In both groups, only a small fraction of experiments requested participants to verbalize their thought processes.

Based on Fig. 7, one of the most striking results was the role of technology in improving higher- order thinking skills like mathematical thinking and problem-solving skills, among students. Digital tools have significantly supported student collaboration and reasoning. Granberg and Olsson (2015) highlight its role in enhancing students' understanding through interactive feedback, while promotes collective reasoning and the linkage of reasoning with proving (Aksu & Zengin, 2022). The studies also observe the positive influence of visual and analytical strategy adoption in learning mathematical structures like slopes (White et al., 2012) and underscore the evolution from intuitive to deductive reasoning as critical for mathematical proof skills (Prusak et al., 2012). The shift towards reflective discourse underlines the value of collaborative tasks in reasoning (Hoek & Seegers, 2005), and Oner (2013) emphasizes that collaboration via dynamic software deepens understanding of geometry. In summary, these findings collectively underscore the important role of digital tools in promoting mathematical reasoning. The evolution of mathematical reasoning in these environments is attributed to the dynamic interplay between individual and collaborative efforts, the meditative role of technology, and the structured design of learning tasks.

But not all results are positive. Lantz-Andersson (2009) research shows that integrating educational software into education can fundamentally change the way students approach and solve problems, sometimes leading them to wrongly blame the software for their mathematical mistakes. This misattribution can hinder their understanding and reasoning in mathematics.

Open-ended problems are highlighted in several articles as effective tasks for fostering reasoning, a notion well-established in

Table 7
Types and frequencies of Outcomes in the corpus classified by categories.

Aspects of HOTS	Freq. Group 1	Freq. Group 2
Reasoning (including underlying processes)	20	10
Mathematical Thinking	4	8
Metacognition	2	5

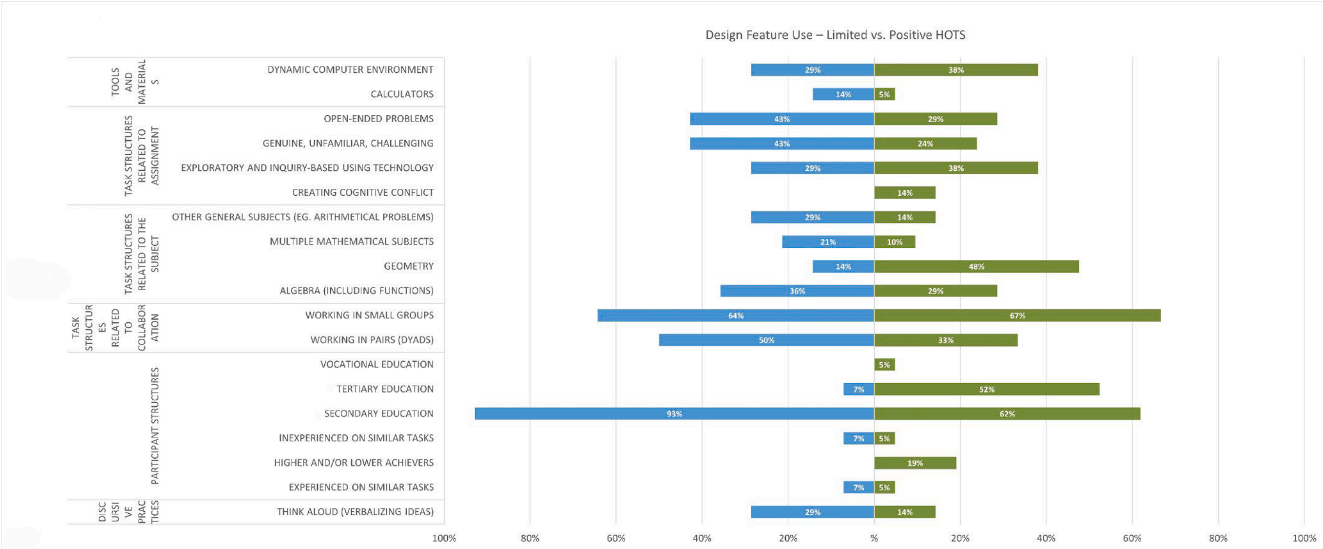

Fig. 7. Comparison of design feature use in articles with positive vs. limited HOTS outcomes.
Note: Green bars represent articles reporting positive outcomes for Higher-Order Thinking Skills (HOTS); blue bars represent articles reporting limited or no outcomes.

academic literature. Three studies in particular associate positive higher -order thinking outcomes with these kinds of tasks. Keh et al. (2019) illustrate that open-ended mathematical problems encourage higher-order thinking and creativity in problem solving. Furthermore, Hoek and Seegers (2005) and Kumpulainen and Kaartinen (2003) underline the transformative impact of open-ended tasks within a social learning context. They observe a notable transition in student discourse from non-reflective to reflective, alongside enhancements in teamwork and cognitive processes. Two authors have presented negative outcomes associated with these kinds of tasks. Chan and Clarke (2017) point out that the negotiation over socio-mathematical and other social aspects of tasks can sometimes detract from a deeper focus on developing sophisticated mathematical understanding. When students spend significant time discussing the social dynamics or the acceptability of various approaches, it might limit their engagement with rigorous mathematical analysis.

Examining the communication theme in group 1′s articles reveals varied outcomes. Communication between students is characterized as dynamic, interactive and multimodal. Students engage in verbal discussions, use mathematical language, and use nonverbal cues such as gestures and body orientation, especially in technology-enhanced environments such as VR (Huang et al., 2023; Kumpulainen & Kaartinen, 2003). Other authors speak of various functions of communication, including sharing and clarifying ideas, negotiating meaning, and constructing shared understanding of mathematical concepts. It is not just about exchanging information, but also about building on each other's contributions to jointly develop a deeper understanding (Bjuland, 2007; Lai & White, 2012; Prusak et al., 2012). Discourse among students is emphasized as a crucial part of the learning process. Adopting the language of peers and engaging in argumentative and constructive dialogue are key to fostering rich mathematical understanding (Granberg & Olsson, 2015; Prusak et al., 2012). Software, such as GeoGebra, plays an important role in improving communication. These tools provide a shared workspace that supports visualization, manipulation of mathematical objects, and CPS, which in turn promotes more coherent and productive peer discussions (Huang et al., 2023; Oner, 2013). Some authors view communication between students through a social constructivist lens, emphasizing the importance of social interaction in the construction of knowledge. (Keh et al., 2019; White et al., 2012). Notably, of the nine authors who documented successful outcomes in communication and collaboration, five experimented with adult learners (Bjuland, 2007; Huang et al., 2023; Keh et al., 2019; Oner, 2013; Prusak et al., 2012). Within this subset, the students in the experiment of Prusak et al. (2012) had extensive experience in collaboration. In the experiment of Huang et al. (2023) the groups benefited from the presence of a 'facilitator' who led the peer discussions, and in the work of Bjuland (2007) the students were informed about collaboration and communication strategies before starting their joint work. In the work from Granberg and Olsson (2015) upper secondary students were involved in solving problems in groups for a longer period of time.

The authors in both groups report a range of communication and collaboration issues. Some authors in group 1 argue that conflicts in collaborative tasks may reduce reasoning and lead to disengagement or misunderstanding without proper management, highlighting the importance of productive communication. Issues like incoherent transactions and strategies, resulting from conflicts, create one-sided interactions and a focus on specific tasks without overall comprehension. Additionally, such conflicts can cause peer domination and hinder effective problem-solving due to a lack of shared understanding (Kumpulainen & Kaartinen, 2003; Mueller, 2009). In VR environments, maintaining effective group work and ensuring equal contribution pose significant challenges (Huang et al., 2023). Kumpulainen and Kaartinen (2003) warn that technology-enhanced environments like VR can result in uneven student participation, affecting collaborative reasoning and the learning experience's effectiveness. In group 2 the following challenges were reported. Some student couples struggle with fragmented communication, where peer discussions are not coherent or complete, leading to misunderstandings. This hinders the depth of shared understanding and the ability to reach consensus on mathematical reasoning (DeJarnette & González, 2015; Kieran, 2002). Problems of authority in the group in which one student often dominates the problem-solving process, relegating the other to a passive role. (DeJarnette et al., 2014; Hansen, 2022). Interactions with peers are generally positive, but can have negative effects on problem solving if there is a lack of mutual respect, an uneven distribution of knowledge, or a balance of power. These factors can cause tensions in the collaboration process, affecting the quality and effectiveness of communication (Goos & Galbraith, 1996). When students do not use or do not use effective metacognitive strategies in a group setting, this can lead to poor problem-solving results (Goos, 2002; Goos & Galbraith, 1996). Open-ended tasks require students to engage in various forms of negotiation, including mathematical and social-mathematical issues. If these negotiations are not managed effectively, it can lead to communication breakdowns, limiting the group's ability to reach a shared understanding or resolve issues effectively (Chan & Clarke, 2017). When students focus more on procedural aspects of mathematical tasks, possibly influenced by educational software or other tools, there may be a neglect of deeper conceptual understanding. This can lead to communication that is more about completing steps than understanding underlying mathematical principles (Lantz-Andersson, 2009). The findings suggest that teachers should design tasks and learning environments that promote effective communication, taking into account the importance of both verbal and non-verbal interactions in the process of collaborative learning (Hoek & Seegers, 2005;Keh et al., 2019).

5. Conclusion and discussion

5.1. Addressing the sub-questions

This study aims to address the research question: What design features and processes of CPS foster HOTS in mathematics education according to research literature? As mentioned before, although our research question focuses on the design features and processes of CPS that promote HOTS, we structure our results section around the components of the Sandoval model. In this section, we will first answer the sub-questions and then the main question.

SQ1: Which conjectures are formulated in research literature?

The study shows that authors formulate conjectures about various aspects of HOTS in mathematics education, most of which relate to reasoning and mathematical thinking, as well as the corresponding CPS processes. Metacognition emerges prominently in these conjectures as well. Less specific terms such as mathematical understanding and learning are also mentioned, although not all articles define these terms consistently.

SQ2: Which design features are used in the various experiments reported in the articles?

Our study offers insights into creating educational settings that foster higher-order thinking through mathematical problem-solving activities. It shows that using technology in an exploratory and investigative manner within geometry proves to be more beneficial than simply presenting open-ended problems. Confronting students with tasks that cause cognitive conflicts are also beneficial. Furthermore, studies involving older students more often report stronger HOTS-related outcomes, and group work is more commonly featured in studies that document positive results than those that involve pair work; however, these patterns do not reflect direct comparative evidence and should be interpreted with caution. Commonly mentioned problems include problems within group dynamics, which often hinder analytical thinking.

SQ3: Which processes are observed during CPS?

If we look at the processes of CPS involved in developing HOTS we can conclude that a clear distinction emerges between the articles reporting positive impacts on HOTS and those showing no or minimal change. In the first group, researchers consistently highlight explicit reasoning and metacognitive processes, such as monitoring, evaluation, and planning, alongside collaborative practices including explanations, discussions, and verbal/non-verbal interactions. These findings suggest that an intentional focus on fostering deeper cognitive and reflective processes contributes significantly to students' development of HOTS. By contrast, the second group of studies, which reports limited or no improvements in HOTS, rarely references reasoning or metacognition. Although negotiation appears prominently in these articles, it is not accompanied by the same depth of discussion or explicit cognitive engagement seen in the first group. This discrepancy implies that negotiation alone, without robust opportunities for reflection and structured discourse, does not suffice to enhance higher-order thinking. Overall, the presence of deeper cognitive and metacognitive processes and structured collaborative interactions appears to be a key driver of successful HOTS development in mathematics education.

SQ4: Which outcomes are reported related to higher order thinking?

In reviewing the articles that formulated conjectures about HOTS, two major groups emerged: studies reporting a positive effect on HOTS and those observing no or minimal change. Among the studies focused on reasoning, 64 % showed beneficial results linked to reasoning processes, while 31 % of those emphasizing mathematical thinking documented gains. Notably, a few articles achieved positive outcomes despite not explicitly formulating conjectures about higher-level analytical thinking. Examining design features, technology frequently proved instrumental in fostering collaboration and reasoning, as digital tools allowed students to visualize mathematical relationships and engage in dynamic problem solving. Many of these technology-enhanced environments supported reflective discourse that strengthened understanding of complex ideas. Meanwhile, open-ended tasks often encouraged creativity and advanced thinking skills. Another design aspect associated with improved HOTS is the purposeful introduction of cognitive conflict tasks, which appear to stimulate critical and reflective thinking. Across both groups, collaboration and communication emerged as important factors for achieving higher-order thinking, with well-structured group activities and guided discussions tending to yield deeper engagement.

Research on CPS in mathematics education consistently highlights several design features and processes that foster HOTS. First, many studies emphasize the importance of tasks that provoke deep reasoning, often through challenges that encourage students to generate and test conjectures, explain their thinking, and engage in metacognitive activities such as monitoring and evaluation. Open-ended problems in particular allow for a broader range of solution strategies, promote creativity, and support the transition from intuitive to more formal, deductive reasoning. Additionally, technology-enhanced tasks are frequently identified as valuable for fostering HOTS, as digital tools (e.g., dynamic geometry software) provide a shared workspace for visualizing concepts, engaging in trial-and-error exploration, and prompting students to articulate and refine their reasoning. Several studies suggest that cognitive conflict may play an important role in leveraging the benefits of such tools, particularly in prompting students to revisit and deepen their mathematical reasoning. Effective collaborative structures also play a critical role. The reviewed studies show that small-group facilitate the exchange of diverse ideas, while well-managed negotiation and exploratory talk help learners build on one another's contributions. Such dialogue supports higher-order processes like analysing, justifying, and synthesizing, especially when guided by intentional facilitation or structured roles that keep learners focused on conceptual understanding. Finally, an emphasis on metacognition, including planning, monitoring, and evaluating problem-solving steps, foster higher-order thinking. When instructional designs explicitly integrate metacognitive prompts or reflection phases, students develop a stronger awareness of their strategies, recognize errors more readily, and refine their approaches as they collaborate.

5.2. Study limitations

This review's findings may be influenced by specific selection criteria and analytical approaches. We limited our review to peer-reviewed articles published in high-quality scientific journals to ensure data integrity. Consequently, book chapters, unpublished dissertations, conference papers, research reports, and other forms of grey literature were excluded. Additionally, the exclusion of non-English studies and studies not available in PDF format may have restricted the diversity of perspectives and findings, potentially introducing language bias. This is evidenced by the majority of included studies being conducted in the United States. A significant limitation of this literature review is that most of the examined articles originate from the United States. This introduces a potential bias due to the specific characteristics of the American education system, which differ in curriculum content, teaching methods, and the degree of teacher and school autonomy compared to other countries. Consequently, the findings may not be directly applicable to other educational contexts. Moreover, while the U.S. is known for its diverse student population, the demographic and cultural contexts can vary significantly from those in other countries. Cultural norms and values that influence peer interactions during discussions may not be present elsewhere. Therefore, caution is needed when generalizing these findings to other settings. Future research should aim for a broader geographical spread of studies to provide a more representative understanding of how peer discussions can enhance higher-order thinking in various educational contexts worldwide.

Furthermore, we did not differentiate between studies based on whether they performed statistical analysis of their results, nor did we extensively compare the outcomes of short-term experiments to those conducted over longer periods. As a result, our literature review may not fully capture the nuances in the strength of empirical support across different studies. This limitation could affect the robustness and generalizability of our conclusions, as the varying methodological rigor and duration of studies might influence the observed impact of peer discussions on students' higher-order thinking.

Although all articles in our review included transcripts of dialogues between students, we relied on the original authors' interpretations of these peer discussions without conducting our own analysis. Our focus was primarily on higher-order thinking in the context of collaborative math problem solving, which may not fully capture other dimensions emphasized by different studies. This focus, while deliberate, highlights the methodological limitations and potential biases that could affect the generalizability and applicability of our conclusions.

5.3. Discussion and implications about the role of communication

Building on the analysis of processes that foster HOTS, it is crucial to consider how communication dynamics shape the success or failure of CPS. The studies in the first group, which reported positive effects on HOTS, often emphasized interactive discourse characterized by processes such as explaining, justifying, and discussing. By contrast, the second group's limited improvement in HOTS was frequently associated with minimal or surface-level communication. We argue that the quality of communication critically determines the success or failure of CPS efforts and, by extension, influences the development of students' HOTS. Based on Liljedahl's (2016) work, it is recommended that to promote a 'thinking classroom', teachers should incorporate engaging, collaborative problem-solving activities that promote communication among students. Although task selection generally aligns with this goal, the failure to adequately teach students communication and interaction skills can dilute outcomes. According to Mazur (1997), fellow students often communicate more effectively at student level than a teacher. However, our analysis suggests that such communication is not inherently effective, nor is it guaranteed, particularly within the context of secondary education. This finding challenges the assumption that peer interactions always facilitate better understanding and highlights the complexities of communication in educational settings.

While several studies demonstrate the potential of collaborative communication to support HOTS, the literature also highlights persistent challenges in enacting effective CPS. These include difficulties in maintaining balanced participation, students reverting to superficial or procedural talk, and the tendency of dominant individuals to steer group interactions. Such issues were evident in studies that reported limited or no gains in HOTS, where interactions often lacked depth and clarity of reasoning. Moreover, even when tasks were well-designed, the absence of explicit support for communication skills or collaborative norms frequently led to underwhelming outcomes, particularly in secondary education settings. These observations echo long-standing concerns in the field about the fragile nature of productive collaboration and the risk of inequality in student contributions.

These challenges reflect findings discussed earlier in the introduction. Simply engaging students in collaborative tasks does not inherently lead to higher-order thinking. As Schoenfeld (1985) points out, the reasoning during problem solving depends not only on task characteristics but on how students interact with the task and each other. Several studies in our review reinforce this: ineffective CPS is often linked to students' difficulty articulating reasoning (Fitzsimons & Ní Fhloinn, 2023), lack of meta-collaborative skills (Barron, 2003), and discomfort in resolving disagreements or reaching shared understanding (Jarry-Shore & Anantharajan, 2015). These issues were especially prominent in studies that reported minimal or no gains in HOTS. Moreover, as Canogullari and Radmehr (2025) note, even task selection can become a barrier when students struggle to grasp the mathematical goals or when tasks fail to offer both accessibility and challenge. Taken together, these findings highlight that successful CPS requires not only well-designed tasks but also explicit instruction in communication norms, collaborative competencies, and strategic reasoning. Without addressing these factors, the potential of CPS to foster HOTS remains fragile.

In examining the peer discussions among students from the studies that reported positive outcomes in tertiary education, it becomes clear that these interactions were characterized by a dynamic exchange of ideas. Students not only shared and challenged each other's thoughts but also engaged in questioning with the intent to clarify concepts and monitor the progression of their work. Furthermore, these peer discussions often led to explicit reasoning following a critical inquiry, illustrating the depth and constructiveness of their communication.

In the introduction, we explore the theoretical ideal of collaboration as highlighted by Boaler (2016) and Schoenfeld (2023), although it becomes apparent that these ideals are often not fully realized in practice. Research highlights the critical role of productive peer discussions in education, emphasizing the understanding of others' perspectives and the joint construction of knowledge as fundamental for effective learning (Crouch & Mazur, 2001; Mazur, 1997). Yackel and Cobb (1996) contribute to this discussion by defining sociomathematical norms, the classroom's accepted practices for reasoning and solving mathematical problems, which are instrumental in shaping these productive peer discussions. By establishing classroom rules that encourage meaningful group and class conversations, as suggested by Littleton and Mercer (2013), teachers can promote these norms, thereby fostering an environment where dialogue enhances professional growth and helps students engage in peer discussions that benefit all participants' learning processes (Ruthven et al., 2017). This integrated approach not only stresses the importance of dialogue in learning but also enhances students' critical thinking and deepens their mathematical understanding.

Looking ahead, future research should concentrate on identifying the most effective strategies to assist teachers in creating the optimal conditions for teaching students to collaboratively solve mathematical problems. Understanding these strategies will equip educators to better facilitate environments where collaboration and higher-order thinking are nurtured, ultimately contributing to more profound educational outcomes.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) solely to assist in improving English language clarity and readability. The authors reviewed and edited the content as needed and take full responsibility for the final version of the manuscript.

Funding

This research received no external funding.

Ethics approval

Not applicable. This study did not involve human participants.

Informed consent

Not applicable.

CRediT authorship contribution statement

Haydeé Ceballos: Writing – original draft, Project administration, Formal analysis. **Theo van den Bogaart:** Writing – review & editing, Supervision. **Stan van Ginkel:** Writing – review & editing, Supervision. **Jeroen Spandaw:** Writing – review & editing, Supervision. **Paul Drijvers:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

No data was used for the research described in the article.

References

**Abdu, R., & Schwarz, B. (2020). Split up, but stay together: Collaboration and cooperation in mathematical problem solving. *Instructional Science, 48*(3), 313–336. https://doi.org/10.1007/s11251-020-09512-7

**Abdullah, A. H., Ahmed, S., Abd Rahman, S. N. S., Mun, S. H., & Mokhtar, M. (2017). Metacognitive behaviour of Malaysian students while solving mathematical problems of Form Three Assessment (PT3). *Bolema - Mathematics Education Bulletin, 31*(59), 907–927. https://doi.org/10.1590/1980-4415v31n59a03

Abrahamson, D., Blikstein, P., & Wilensky, U. (2007). *Classroom model, model classroom: Computer-supported methodology for investigating collaborative-learning pedagogy, (Vol. 8, Part 1,* pp. 46–55). International Society of the Learning Sciences. http://ccl.northwestern.edu/research/conferences/CSCL2007/CSCL2007.html.

Abulencia, J. P., Vigeant, M. A., & Silverstein, D. L. (2012). Using video media to enhance conceptual learning in an undergraduate thermodynamics course. In *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2–22731

Acosta, H., Hong, D., Lee, S., Min, W., Mott, B., Hmelo-Silver, C., & Lester, J. (2025). Collaborative game-based learning analytics: Predicting learning outcomes from game-based collaborative problem solving behaviors. In *15th International Conference on Learning Analytics and Knowledge, LAK 2025* (pp. 429–438). https://doi.org/10.1145/3706468.3706522

Adamson, S. L., Banks, D., Burtch, M., Cox, F., Judson, E., Turley, J. B., Benford, R., & Lawson, A. E. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching, 40*(10), 939–957. https://doi.org/10.1002/tea.10117

Aerts, C. (n.d.). *ENEAS: The European Network of Excellence in AsteroSeismology.* https://www.researchgate.net/publication/234233906.

Aggarwal, C. C. (2004). A human-computer interactive method for projected clustering. *IEEE Transactions on Knowledge and Data Engineering, 16*(4), 448–460. https://doi.org/10.1109/TKDE.2004.1269669

**Aksu, N., & Zengin, Y. (2022). Disclosure of students' mathematical reasoning through collaborative technology-enhanced learning environment. *Education and Information Technologies, 27*(2), 1609–1634. https://doi.org/10.1007/s10639-021-10686-x

Amer, A. (2005). *Analytical thinking.* Cairo, Egypt: CAPSCU, Faculty of Engineering, Cairo University, ISBN 977-403-011-7.

ATLAS.ti Scientific Software Development GmbH. (2023). *ATLAS.ti mac (Version 23.2.1) [Qualitative data analysis software]*. https://atlasti.com.

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers.* Routledge.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

**Bjuland, R. (2007). Adult students' reasoning in geometry: Teaching mathematics through collaborative problem solving in teacher education. *The Mathematics Enthusiast, 4*(1), 1–30. https://doi.org/10.54870/1551-3440.1056

**Bleiler-Baxter, S. K., Kirby, J. E., & Reed, S. D. (2023). Understanding authority in small-group co-constructions of mathematical proof. *Journal of Mathematical Behavior, 70*, Article 101015. https://doi.org/10.1016/j.jmathb.2022.101015

Bloodgood, M., & Vijay-Shanker, K. (2014). A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping. *arXiv.* http://arxiv.org/abs/1409.5165.

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching.* Jossey-Bass/Wiley.

Boesen, J., Lithner, J., & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics, 75*, 89–105.

Boetje, J., & van de Schoot, R. (2024). The SAFE procedure: A practical stopping heuristic for active learning-based screening in systematic reviews and meta-analyses. *Systematic Reviews, 13*(1). https://doi.org/10.1186/s13643-024-02502-7

Burton, L. (1984). Mathematical thinking: The struggle for meaning. *Journal for Research in Mathematics Education, 15*(1), 35–49.

Canogullari, A., & Radmehr, F. (2025). Task design principles in mathematics education: A literature review. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2025.2457365

**Carlsen, M. (2010). Appropriating geometric series as a cultural tool: A study of student collaborative learning. *Educational Studies in Mathematics, 74*(2), 95–116. https://doi.org/10.1007/s10649-010-9230-0

Ceballos, H., & Spandaw, J. (2017). Peer instruction: Leren van je klasgenoten. *Euclides, 92*(7), 4–5.

**Chan, M. C. E., & Clarke, D. (2017). Structured affordances in the use of open-ended tasks to facilitate collaborative problem solving. *ZDM - Mathematics Education, 49*(6), 951–963. https://doi.org/10.1007/s11858-017-0876-2

Crouch, C., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970–977.

**DeJarnette, A. F. (2022). What do correct answers reveal? The interpersonal and mathematical aspects of students' interactions during groupwork in seventh grade mathematics. *Journal of the Learning Sciences, 31*(4–5), 509–544. https://doi.org/10.1080/10508406.2022.2073232

**DeJarnette, A. F., Dao, J. N., & González, G. (2014). Learning what works: Promoting small-group discussions. *Mathematics Teaching in the Middle School, 19*(7), 414–419. https://doi.org/10.5951/mathteacmiddscho.19.7.0414

**DeJarnette, A. F., & González, G. (2015). Positioning during group work on a novel task in Algebra II. *Journal for Research in Mathematics Education, 46*(4), 378–422. https://doi.org/10.5951/jresematheduc.46.4.0378

Dekker, R., & Elshout-Mohr, M. (1998). A process model for interaction and mathematical level raising. *Educational Studies in Mathematics, 35*, 303–314. https://doi.org/10.1023/A:1003187204737

**Demir, M., & Zengin, Y. (2023). The effect of a technology-enhanced collaborative learning environment on secondary school students' mathematical reasoning: A mixed method design. *Education and Information Technologies, 28*(8), 9855–9883. https://doi.org/10.1007/s10639-023-11587-x

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Elsevier. https://telearn.hal.science/hal-00190240v1.

Fitzsimons, A., & Ní Fhloinn, E. (2023). The COPS model for collaborative problem-solving in mathematics. *Irish Educational Studies*. https://doi.org/10.1080/03323315.2023.2189137

**Francisco, J. M. (2013). Learning in collaborative settings: Students building on each other's ideas to promote their mathematical understanding. *Educational Studies in Mathematics, 82*(3), 417–438. https://doi.org/10.1007/s10649-012-9437-3

**Goos, M. (1994). Metacognitive decision making and social interactions during paired problem solving. *Mathematics Education Research Journal, 6*(2), 85–104. https://doi.org/10.1007/BF03217269

**Goos, M. (2002). Understanding metacognitive failure. *Journal of Mathematical Behavior, 21*, 183–202.

**Goos, M., & Galbraith, P. (1996). Do it this way! Metacognitive strategies in collaborative mathematical problem solving. *Educational Studies in Mathematics, 30*(3), 229–260. https://doi.org/10.1007/BF00304567

**Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics, 49*(2), 193–223. http://www.jstor.org/stable/3483075.

Graesser, A. C., Greiff, S., Stadler, M., & Shubeck, K. T. (2020). Collaboration in the 21st century: The theory, assessment, and teaching of collaborative problem solving. *Computers in Human Behavior, 104*, Article 106168. https://doi.org/10.1016/j.chb.2019.09.010

**Granberg, C., & Olsson, J. (2015). ICT-supported problem solving and collaborative creative reasoning: Exploring linear functions using dynamic mathematics software. *Journal of Mathematical Behavior, 37*, 48–62. https://doi.org/10.1016/j.jmathb.2014.11.001

**Haataja, E. S. H., Chan, M. C. E., Salonen, V., & Clarke, D. J. (2022). Can noncomplementarity of agency lead to successful problem solving? A case study on students' interpersonal behaviors in mathematical problem-solving collaboration. *Learning and Instruction, 82*, Article 101657. https://doi.org/10.1016/j.learninstruc.2022.101657

**Hansen, E. K. S. (2022). Students' agency, creative reasoning, and collaboration in mathematical problem solving. *Mathematics Education Research Journal, 34*(4), 813–834. https://doi.org/10.1007/s13394-021-00365-y

**Hansen, E. K. S., & Naalsund, M. (2022). The role of teacher actions for students' productive interaction solving a linear function problem. *International Electronic Journal of Mathematics Education, 17*(3), em0685. https://doi.org/10.29333/iejme/11921

He, Q., von Davier, M., Greiff, S., Steinhauer, E. W., & Borysewicz, P. B. (2017). Collaborative problem solving measures in the programme for international student assessment (PISA). *Methodology of educational measurement and assessment* (pp. 95–111). Springer Nature. https://doi.org/10.1007/978-3-319-33261-1_7

**Hoek, D. J., & Seegers, G. (2005). Effects of instruction on verbal interactions during collaborative problem solving. *Learning Environments Research, 8*, 75–96. https://doi.org/10.1007/s10984-005-7949-9

**Huang, W., Walkington, C., & Nathan, M. J. (2023). Coordinating modalities of mathematical collaboration in shared VR environments. *International Journal of Computer-Supported Collaborative Learning, 18*(2), 163–201. https://doi.org/10.1007/s11412-023-09397-x

Jarry-Shore, M., & Anantharajan, M. (2015). In *Proceedings of the 43rd annual meeting of PME-NA: Student struggle during collaborative problem-solving in one mathematics classroom*.

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics, 96*(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8

**Keh, L. K., Ismail, Z., & Yusof, Y. M. (2019). A review of open-ended mathematical problem. *Anatolian Journal of Education, 1*(1). https://doi.org/10.29333/aje.2016.111a

**Kerrigan, J., Weber, K., & Chinn, C. (2021). Effective collaboration in the productive failure process. *Journal of Mathematical Behavior, 63*, Article 100892. https://doi.org/10.1016/j.jmathb.2021.100892

**Kieran, C. (2002). The mathematical discourse of 13-year-old partnered problem solving and its relation to the mathematics that emerges. In C. Kieran, & E. S. A. Forman (Eds.), *Learning discourse* (pp. 191–228). Springer. https://doi.org/10.1007/0-306-48085-9_6.

**Kim, Y. R., & Moore, T. J. (2019). Multiple levels of metacognition: Circumstances interfering with students' spontaneous metacognitive activities. *Journal of Educational Research and Practice, 9*(1). https://doi.org/10.5590/jerap.2019.09.1.12

**Kumpulainen, K., & Kaartinen, S. (2003). The interpersonal dynamics of collaborative reasoning in peer interactive dyads. *Journal of Experimental Education, 71*(4), 333–370. https://doi.org/10.1080/00220970309602069

**Lai, K., & White, T. (2012). Exploring quadrilaterals in a small group computing environment. *Computers and Education, 59*(3), 963–973. https://doi.org/10.1016/j.compedu.2012.04.006

**Lantz-Andersson, A. (2009). The power of natural frameworks: Technology and the question of agency in CSCL settings. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 93–107. https://doi.org/10.1007/s11412-008-9058-0

Leighton, J., & Sternberg, R. (2004). In R. Sternberg (Ed.), *The nature of reasoning*. Cambridge University Press.

Lester, F. K., & Cai, J. (2016). Can mathematical problem solving be taught? Preliminary answers from 30 years of research. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems* (pp. 117–135). Springer International Publishing. https://doi.org/10.1007/978-3-319-28023-3_8.

Lester, F. K., & Kehle, P. E. (2003). From problem solving to modeling: The evolution of thinking about research on complex mathematical activity. In R. Lesh, & H. Doer (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 501–517). Lawrence Erlbaum Associates.

Liljedahl, P. (2016). Building thinking classrooms: Conditions for problem-solving. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems* (pp. 361–386). Springer International Publishing. https://doi.org/10.1007/978-3-319-28023-3_21.

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Routledge. https://doi.org/10.4324/9780203809433
Makmuri, Aziz, T. A., & Kharis, S. A. A (2021). Characteristics of problems for developing higher-order thinking skills in mathematics. *Journal of Physics: Conference Series, 1882*(1), Article 012074. https://doi.org/10.1088/1742-6596/1882/1/012074
Mason, J., & Johnston-Wilder, S. (2004). *Designing and using mathematical tasks*. Tarquin.
Mazur, E. (1997). Peer instruction: A user's manual. In A. Reeves (Ed.), *Series in educational innovation)* (1st ed.). Prentice Hall.
**Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73*(4), 449–471. https://doi.org/10.1348/000709903322591181
**Moate, J., Kuntze, S., & Chan, M. C. E. (2021). Student participation in peer interaction: Use of material resources as a key consideration in an open-ended problem-solving mathematics task. *LUMAT, 9*(1), 29–55. https://doi.org/10.31129/LUMAT.9.1.1470
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097
**Mueller, M. F. (2009). The co-construction of arguments by middle-school students. *Journal of Mathematical Behavior, 28*(2–3), 138–149. https://doi.org/10.1016/j.jmathb.2009.06.003
**Nieminen, J. H., Chan, M. C. E., & Clarke, D. (2022). What affordances do open-ended real-life tasks offer for sharing student agency in collaborative problem-solving? *Educational Studies in Mathematics, 109*(1), 115–136. https://doi.org/10.1007/s10649-021-10074-9
OECD. (2017). *PISA 2015 assessment and analytical framework*. OECD. https://doi.org/10.1787/9789264281820-en
**Oner, D. (2013). Analyzing group coordination when solving geometry problems with dynamic geometry software. *International Journal of Computer-Supported Collaborative Learning, 8*(1), 13–39. https://doi.org/10.1007/s11412-012-9161-0
Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell.
Pijls, M., & Dekker, R. (2011). Students discussing their mathematical ideas: The role of the teacher. *Mathematics Education Research Journal, 23*(4), 379–396. https://doi.org/10.1007/s13394-011-0022-3
**Prusak, N., Hershkowitz, R., & Schwarz, B. B. (2012). From visual reasoning to logical necessity through argumentative design. *Educational Studies in Mathematics, 79*(1), 19–40. https://doi.org/10.1007/s10649-011-9335-0
Robbins, J. (2011). Problem solving, reasoning, and analytical thinking in a classroom environment. *The Behavior Analyst Today, 12*(1), 41–48. https://doi.org/10.1037/h0100710
Ruthven, K., Mercer, N., Taber, K. S., Guardia, P., Hofmann, R., Ilie, S., Luthman, S., & Riga, F. (2017). A research-informed dialogic-teaching approach to early secondary school mathematics and science: The pedagogical design and field trial of the epiSTEMe intervention. *Research Papers in Education, 32*(1), 18–40. https://doi.org/10.1080/02671522.2015.1129642
Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204
Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Macmillan.
Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
Schoenfeld, A. H. (2023). Ways to help students become powerful mathematical thinkers. *Mathematics Teacher: Learning and Teaching PK-12, 116*(11), 888–889. https://doi.org/10.5951/MTLT.2023.0221
**Schwarz, B. B., & Linchevski, L. (2007). The role of task design and argumentation in cognitive development during peer interaction: The case of proportional reasoning. *Learning and Instruction, 17*(5), 510–531. https://doi.org/10.1016/j.learninstruc.2007.09.009
**Sivaraj, R., Wieselmann, J. R., & Roehrig, G. (2021). Designing solutions in middle school engineering: An exploration of epistemic practices of engineering in small group contexts (Work in Progress). In *2021 ASEE Virtual Annual Conference Content Access*. https://doi.org/10.18260/1-2–36930
Stacey, K. (2006). What is mathematical thinking and why is it important?. In *Proceedings of the APEC-Tsukuba International Conference 2007: Innovative teaching mathematics through lesson study* (pp. 4–13). https://www.researchgate.net/publication/254408829.
Stanic, G., & Kilpatrick, J. (1988). Historical perspectives on problem solving in the mathematics curriculum. In R. I. Charles, & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 1–22). National Council of Teachers of Mathematics.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.
**Szabo, A., Tillnert, A.-S., & Mattsson, J. (2024). Displaying gifted students' mathematical reasoning during problem solving: Challenges and possibilities. *The Mathematics Enthusiast, 21*. https://doi.org/10.54870/1551-3440.1623.
**Tatsis, K., & Koleza, E. (2006). The effect of students' roles on the establishment of shared knowledge during collaborative problem solving: A case study from the field of mathematics. *Social Psychology of Education, 9*(4), 443–460. https://doi.org/10.1007/s11218-006-9005-8
van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence, 3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7
van der Linden, S., van der Meij, J., & McKenney, S. (2022). Teacher video coaching, from design features to student impacts: A systematic literature review. *Review of Educational Research, 92*(1), 114–165. https://doi.org/10.3102/00346543211046984
**Walen, S., Williams, S., & Barton, H. (1999). Dollars and sense: A case of distributed cognition in a problem-solving session. *Mathematics Education Research Journal, 11*(1), 37–52. https://doi.org/10.1007/BF03217350
**Walter, J. G., & Barros, T. (2011). Students build mathematical theory: Semantic warrants in argumentation. *Educational Studies in Mathematics, 78*(3), 323–342. https://doi.org/10.1007/s10649-011-9326-1
**White, T. (2006). Code talk: Student discourse and participation with networked handhelds. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 359–382. https://doi.org/10.1007/s11412-006-9658-5
**White, T., Wallace, M., & Lai, K. (2012). Graphing in groups: Learning about lines in a collaborative classroom network environment. *Mathematical Thinking and Learning, 14*(2), 149–172. https://doi.org/10.1080/10986065.2012.656363
Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477. https://doi.org/10.2307/749877
Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 227–236). National Council of Teachers of Mathematics.
**Yemen-Karpuzcu, S., Ulusoy, F., & Işıksal-Bostan, M. (2017). Prospective middle school mathematics teachers' covariational reasoning for interpreting dynamic events during peer interactions. *International Journal of Science and Mathematics Education, 15*(1), 89–108. https://doi.org/10.1007/s10763-015-9668-8
**Zhang, S., Cao, Y., Chan, M. C. E., & Wan, M. E. V. (2022). A comparison of meaning negotiation during collaborative problem solving in mathematics between students in China and Australia. *ZDM - Mathematics Education, 54*(2), 287–302. https://doi.org/10.1007/s11858-022-01335-9