WILEY

Higher Education Quarterly

REVIEW ARTICLE

Argument Mapping in Higher Education: A Systematic Review

John C. Nesbit | Qing Liu

Faculty of Education, Simon Fraser University, Burnaby, Canada

Correspondence: John C. Nesbit (jcnesbit@sfu.ca)

Received: 1 March 2025 | **Revised:** 4 September 2025 | **Accepted:** 8 September 2025

Funding: This research was supported by funding from the Social Sciences and Humanities Research Council of Canada (420-2024-00378) awarded to Qing Liu (principal investigator) and John C. Nesbit (co-investigator).

Keywords: argument | diagram | map | postsecondary | review

ABSTRACT

Given the emphasis on critical thinking across the undergraduate curricula, research on argument visualisation has significant implications for designing learning activities in higher education. This systematic review examines research on the use of argument maps or diagrams by postsecondary students. The goals were to identify the themes, research questions, and results of systematically identified studies, and to assess the current prospects for meta-analyses. Relevant databases were searched for qualitative, observational and experimental studies. We coded 124 studies on research design, mapping software, student attitudes, collaborative mapping and thinking skills. There were 102 empirical studies, of which 44% assessed student attitudes toward argument mapping, 40% investigated collaborative argument mapping and 51% examined the quality or structure of student-constructed argument maps. The causal relationship most frequently investigated was the effect of argument mapping on critical thinking skills. We present the results from selected studies and consider their significance for learning design.

Developing students' ability to argue, the core component of critical thinking, has long been held as a goal of higher education (Wolfe 2011). Whether taught in dedicated courses, 'Critical Thinking 101', or integrated into a myriad of courses across disciplines, informal reasoning and argumentation have continued to receive considerable attention across the undergraduate curricula (Kleemola et al. 2022). Argument mapping has been increasingly adopted as an approach to developing argumentation and critical thinking ability in undergraduate education, and argument mapping researchers have reported gains in critical thinking ability relative to other common methods of teaching argumentation (van Gelder 2015). An additional and complementary purpose for adopting argument mapping in the postsecondary classroom is to deepen students' knowledge of the concepts that are the subject of the argument (Asterhan and Schwarz 2016).

Despite substantial research that has investigated argument mapping, and the obvious importance of understanding its potential benefits as an instructional approach in higher education, we could find no peer-reviewed systematic reviews that focus on the topic. Scheuer et al. (2010) conducted a systematic review of computer-supported argumentation which emphasised system features and learning effects across educational levels. A thesis by Alvarez (2007) reported a meta-analysis which investigated the effects of philosophy instruction on critical thinking ability. Although both reviews considered the effects of argument mapping as a facet of their work, neither examined the full set of studies on argument mapping for postsecondary learners available at the time they were published, and, of course, many studies have appeared since that time. This systematic review examines the range of research questions investigated in the primary research, not only effects on critical thinking ability.

We cover additional dependent variables such as domain knowledge, writing skills, and student attitudes toward argument mapping, and we consider independent variables such as collaborative versus individual argument mapping.

1.1 | What Is Argument Mapping?

Whether drawn on paper or composed with specialised software, an argument map or diagram is a visual representation of an argument's structure. An argument map often consists of nodes representing propositions (i.e., statements) connected by links that indicate support or opposition. Figure 1 shows a simple argument map representing an argument implicit in a commonplace conversation.

More complex argument map formats give greater detail about the role of each statement in an argument, allowing arguers to specify that a statement (i.e., a qualifier) limits or restricts a claim to which it is connected, or that a supporting or opposing node is justified by another statement (i.e., a warrant) to which it is conjoined.

Although they represent the same fundamental support/oppose relations between statements, some argument maps are designed more like charts than node-link maps. They tend to have a bilateral structure with statements supporting a main claim listed on one side and statements opposing the claim on the other (Nesbit et al. 2024). In contrast, argument maps with a node-link structure tend to allow more choice in the placement of statements (nodes).

Although it is the visuospatial features of argument maps that distinguish them from conventionally written arguments, the meaning that argument maps convey is still highly dependent on the textual statements that appear within their boxes or panels. Davies (2011) compared three popular mapping techniques—concept mapping, mind mapping and argument mapping—and highlighted their distinct functions in organising and representing knowledge. Concept mapping and mind mapping are used to depict a wide range of semantic relationships between concepts. Argument mapping differs by focusing on logical reasoning, representing inferential connections between statements rather than causal or associative links. Crucially, each box or node in an argument map must contain at least a proposition, whereas a node in a concept map or mind map may contain a noun or noun phrase.

1.2 | Mapping to Learn

Argument-focused learning activities in postsecondary education tend to be justified with respect to two distinct curricular goals. They are presented as (a) developing students' general argumentative or critical thinking ability (Davies 2019; van Gelder 2005) or (b) providing a cognitively engaging vehicle for fostering student construction of domain concepts (Jonassen and Kim 2010). This is equally true of learning activities based on argument mapping.

From a cognitive perspective, the first goal of argument mapping, learning to argue, means internalising an argument schema and rehearsing the ability to apply it in a range of situations. As argument maps are a more tangible and systematic formalism for representing arguments than other verbal media such as argument essays, repeated practice in constructing them may be a particularly efficient way to develop argument schemas in long term memory.

The second goal, arguing to learn, means using argument to cognitively process and construct meaning from information about the domain. For instance, making an argument about the relationship between voltage, current and resistance using data gleaned from measurements of a circuit may be an effective way to develop an accurate and resilient mental model of Ohm's law. In this case, the argument map serves as a cognitive tool (Pakdaman-Savoji et al. 2019) that guides the learner to process important semantic relations and thus facilitates the construction of the mental model. Davies (2009) contended that mapping facilitates meaningful learning by structuring new information in a way that connects to existing knowledge, thereby enhancing comprehension and retention.

We recognise a third goal, not always made explicit, which is the intersection of argumentation ability (the first goal) and disciplinary expertise (the second goal). The third goal is to develop *domain-specific* argumentation abilities, which are the unique argumentative strategies and their associated terminology that emerge from the application of argument to domain-specific concepts. Almost every disciplinary domain in postsecondary education (law, medicine, philosophy, etc.) strives to develop students' domain-specific reasoning and critical thinking strategies. The goal of developing discipline-specific argumentation skills may have implications for the types of argument maps and mapping activities adopted in the varied disciplines of higher education.

1.3 | Themes and Causal Relationships

In this review of argument mapping in higher education, our purpose was to identify and categorise the research questions investigated in this field, and to present the findings of selected representative studies. One of the challenges in reviewing a

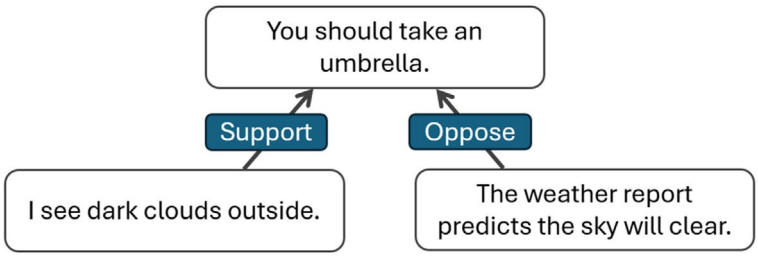

FIGURE 1 | Example of a simple argument map.

body of research with disparate goals and methods is to maintain a coherent narrative that selects the most representative and impactful examples and organises them into meaningful clusters. We have chosen to organise our review according to common research themes and causal relationships investigated by the reviewed studies. Studies that use qualitative methods or single-group quantitative methods are often not aiming to identify causal relationships. Instead, their contribution is to investigate what we refer to as *themes* such as student perceptions of a learning experience, the socio-cognitive processes or behaviours exhibited by learners, the products that result from those processes, or the instructional designs implemented by instructors. In contrast, studies that use quantitative methods to compare learners receiving different treatments seek to determine the existence and strength of *causal relationships* between independent and dependent variables. The crucial link between the two types of studies is that themes inform the construction and operationalisation of the independent and dependent variables that define causal relationships. Accordingly, our strategy in summarising research results is to first present the themes that researchers have explored and use them to inform our presentation of research investigating causal relationships.

A third purpose of this review is to survey the prospects for meta-analyses on argument mapping in higher education. In educational research, meta-analyses that investigate the effect of instructional treatments on learning outcomes are typically restricted to reviewing studies that compare the outcomes of treatment groups. Single-group, pretest–posttest studies are usually excluded because they pose obvious threats to validity. Students tend to progress intellectually as they go through university, and therefore any changes between the repeated measures cannot be solely attributable to the research intervention (Marsden and Torgerson 2012). To have enough statistical power to detect an effect, a meta-analysis requires sufficient studies and participants (Valentine et al. 2010). While any cutoff is contestable, we accept that at least 10 studies are desirable for a publishable meta-analysis that includes moderator analysis (Myung 2023). Because we are interested in assessing the viability of meta-analyses for research on argument mapping in postsecondary education, the coding system and analysis adopted for this review gives greater attention to experimental and quasi-experimental studies that compare treatment groups.

The fourth purpose of the review is to draw implications from the results for the design of learning activities in higher education. As we have discussed, critical thinking and argumentation are widely regarded as important goals at this educational level, and those goals permeate the curricula across most undergraduate programs. Therefore, research findings that bear on the efficacy or perceived value of instructional innovations aiming to develop those abilities should be of great interest to university and college instructors. In the final section of the discussion, we propose ways, informed by our findings, that instructors might feasibly and productively incorporate argument mapping in their courses.

2 | Methods

Selection criteria were developed to capture all available journal articles, conference proceedings and edited book chapters relevant to argument mapping in postsecondary education and published in English. Dissertations, theses and full-length books were excluded to ensure a focus on sources that have undergone formal peer review. We included research that:

- Reviewed theories, empirical studies, or instructional practices related to argument mapping in postsecondary contexts. Such studies provide foundational insights into how argument mapping is conceptualised, implemented and evaluated in higher education.
- Collected and analysed data from postsecondary students who either constructed argument maps or engaged with preconstructed maps as a learning activity. Although relatively few studies focused on learning from preconstructed maps, they were included to broaden our understanding of different instructional approaches. We recognise that the distinction between studying and constructing maps is instructionally important and clarify the mode of learning when describing individual studies.
- Employed quantitative or qualitative research designs. Including studies with diverse methodologies and research foci enabled us to develop a clearer understanding of the research landscape—what has been explored, which designs and methods have been used, and where gaps or opportunities for future inquiry remain.

Between 6 September and 7 October, 2024, we searched EBSCO (PsycInfo, Academic Search Complete, Education Source, ERIC), the Web of Science Core Collection, Scopus, ACM Digital Library and IEEE Xplore. Due to differences in database features, in EBSCO, Web of Science and Scopus, we searched the title, abstract and key words; and in ACM DL and IEEE Xplore, we searched all fields and full text. The following search query, or a near-as-possible proxy, was used for all databases:

("argument* map*" OR "argument* visual*" OR "argument* diagram*")
AND
(learn* OR educat* OR instruct*)

Figure 2 shows a PRISMA flowchart (Page et al. 2021) that details the identification, screening and inclusion of the reports. The publications excluded as 'not on topic' during screening tended to be computing science investigations of problems such as automatic extraction of argument maps from text, uses of argument maps for public policy decision making, and linguistic research on the representation of argumentation in language. 119 reports were retained for coding and indexed in an Excel spreadsheet. During the coding process, four articles were discovered that reported multiple studies. The studies were separately indexed, and the spreadsheet was expanded to list 124 studies.

2.1 | Code Development and Interrater Agreement

A set of codes (i.e., a codebook) was developed that reflected the goals of the review and our nascent understanding of common themes in the studies acquired through the selection process.

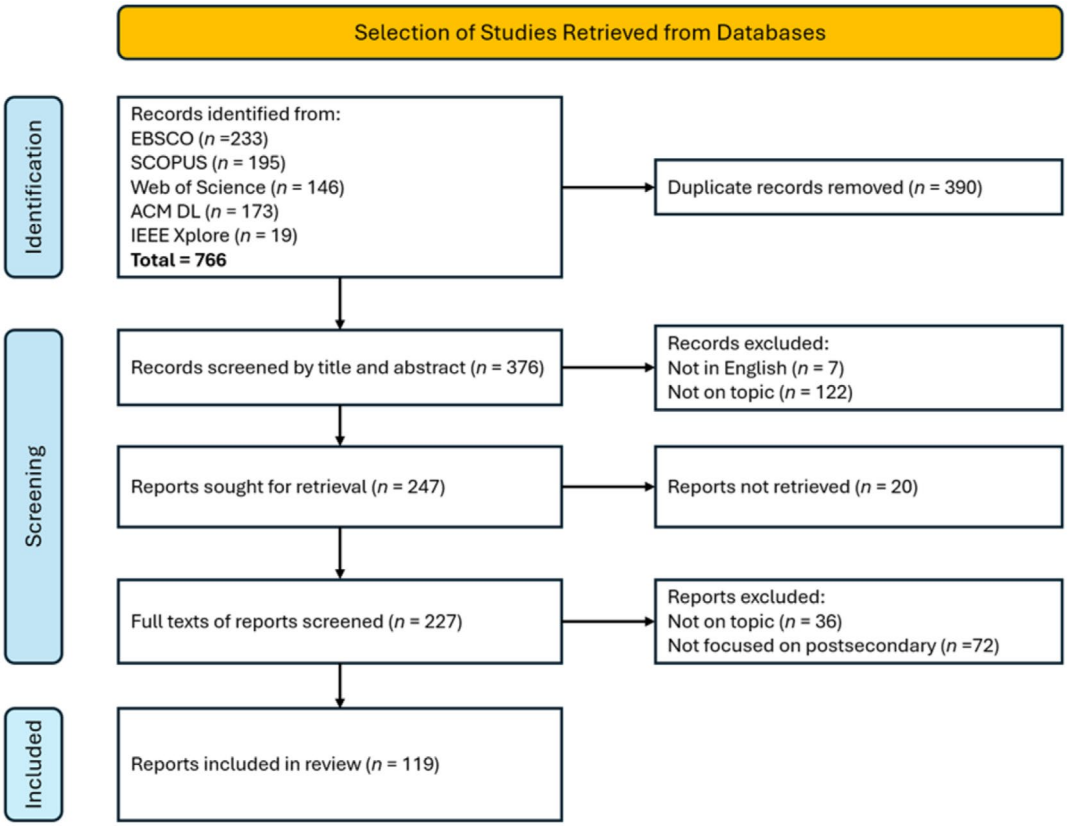

FIGURE 2 | PRISMA flowchart showing how studies were selected.

We created codes to identify whether the study was an empirical investigation; whether it analysed quantitative data, qualitative data, or both; and, if quantitative, what type of research design was used. Focusing on the themes investigated by the studies, we created codes to identify whether studies examined collaborative argument mapping, student attitudes toward argument mapping, use of technology to support argument mapping, the validity of argument maps for educational assessment, and the quality or structure of students' argument maps. Studies that gave different treatments to groups of participants and quantitatively assessed the effect of independent variables on dependent variables were of special interest. For such studies, we developed codes to mark the use of key independent variables (argument mapping, argument map type, learning context) and key dependent variables (thinking skills, writing skills, motivation to learn, domain concepts, argument map quality). A few other codes (country in which the research was conducted, name of software used for argument mapping, setting in which the argument mapping occurred) were used to collect data that were peripheral to the research questions.

Coding studies for systematic review involves many decisions that require interpretation of predefined codes (i.e., variables) and application of the codes to unanticipated study features. Although independent coding by multiple researchers raises the risk of inconsistent coding decisions, evaluation and improvement of agreement among multiple raters can substantially increase the transparency, consistency and replicability of the coding process (Belur et al. 2021). We randomly selected studies to form three batches that both authors coded as part of a code development and interrater agreement process. After independently applying the draft codes to the first batch of studies ($n = 13$), we met to discuss disagreements and modify, merge, add, or discard codes that had proven difficult to apply with consistent agreement. The resulting codebook consisted of the 19 codes shown in Appendix.

Two codes (Country, Mapping Tool) were exempted from interrater agreement evaluation because they were found to be easily and objectively codable. We used Cohen's kappa (k) to assess interrater agreement (Cohen 1960; McHugh 2012). Our goal for interrater agreement was $k \geq 0.81$ for each code, which Landis and Koch (1977, 165) identified as an arbitrary benchmark indicating "almost perfect" agreement. After using the codebook independently to code the second batch of studies ($n = 12$), we obtained a mean interrater agreement of $k = 0.77$ and a range across the codes from $k = 0.31$ to $k = 1.00$. After another meeting to discuss disagreements and further refine definitions in the codebook, we coded the third batch of studies ($n = 19$). This time we obtained a mean interrater agreement of $k = 0.91$. Only one code (Thinking Skills as DV, $k = 0.73$) fell below our criterion. Also, kappa could not be calculated for three codes (Argument Map Type as IV, Motivation as DV, Argument Mapping Ability as DV) due to insufficient instances of those features in our final batch. We decided to randomly divide the remaining studies between the two authors, who proceeded individually to code them using all 19 codes. Finally, we calculated post hoc kappa values for the four codes which had not yet met the reliability criterion. For the Thinking Skills as DV code, this was done by randomly selecting and double-coding five cases that had been coded 1 and six cases that had been coded 0, a procedure which returned $k = 0.82$. The relative lack of consistent agreement on

this code was due to the difficulty of defining the concept of thinking skills and the frequent occurrence of unanticipated borderline cases of the concept. But since the criterion level for agreement was reached, we retained the code for analysis. For the low frequency codes, we double-coded all the cases coded as 1 and a random selection of a sufficient number of those coded 0 to allow a kappa calculation based on a set of 11 cases. Each low frequency code returned $k = 1.00$.

In reporting results of individual studies, we present mean difference effect sizes whenever they can be obtained from or reliably derived from the study. Where there is a difference between Cohen's *d* and Hedge's *g* due to small sample size, we report *g*. When the two metrics produce the same value, we report it as *d*.

3 | Results

Of the 124 retained studies, the 102 coded as empirical were subject to further coding and analysis. The remainder were 7 reviews dealing with various aspects of argument mapping in higher education and 15 proposals for technical or instructional designs. Our systematic review focuses on the empirical studies but cites the identified reviews as needed to explain and support ideas emerging from the analysis.

3.1 | Argument Mapping Software Tools

Of the empirical studies, 91 had students work with computer-based tools to construct argument maps. While some studies had students use general graphical tools (e.g., PowerPoint) or pencil and paper, most deployed specialised argument mapping tools. The studies identified 42 specialised argument mapping tools, which varied widely in structure and functionality. The most widely used tools were Rationale (15 reports), LASAD (8), LARGO (7), Kialo Edu (3), GAIL (3), jMap (3), MindMup (3) and AGORA (3).

3.2 | Research Methodologies

We found 98 studies that collected, analysed and presented data from learners using an identifiable research methodology. Of these, 58 used solely quantitative analysis, 10 used solely qualitative analysis and 30 combined quantitative and qualitative analysis.

Among the 40 studies which used qualitative analysis, there were three main approaches. By far, the most common approach was to ask students open-ended questions in an interview or survey after they had participated in an argument mapping activity and then categorise their expressed beliefs about and attitudes toward the activity. Kaeppel (2021), for example, conducted semi-structured interviews with students after they engaged with argument mapping. The students' statements were analysed to categorise their attitudes toward the argument mapping experience. A second approach was to use qualitative methods to analyse participants' sequences of cognitive and behavioural processes while constructing argument maps. The data were obtained from a variety of sources, which included think-aloud protocols, video, chat logs and retrospective interviews. Jeong and Kim (2022) exemplified this approach by recording think-aloud protocols during a mapping activity and participants' post-activity recollections in interviews. Through this method, they were able to find differences in the way experts and novices constructed argument maps. A third and remarkably rare approach was to qualitatively analyse the structures or semantics of argument maps created by students (e.g., Hoffmann and Lingle 2015; Muller Mirza et al. 2007). Researchers interested in analysing argument maps were more likely to adopt a quantitative approach in which they counted features such as the number of counterarguments or instances of supportive evidence (e.g., Liu et al. 2024).

Among the 88 studies which used quantitative analysis, 43 were limited to single-group or observational research designs, 24 used a quasi-experimental design that compared pre-existing groups who received different treatments, and 26 used experimental designs that compared randomly assigned treatment groups. Paralleling the qualitative research, the most common application of quantitative single-group research designs was to have students engage in one or more argument mapping activities and then respond to a self-report questionnaire which measured their attitudes toward the experience. Another common approach was to use a pretest and posttest to measure changes over time that were presumed to be due to the argument mapping experience. Rather than administer repeated tests, some researchers gave two or more argument mapping assignments and used rubrics to score the quality of the maps. Wu et al. (2014) used this approach to show that the quality of students' maps was greater on the second mapping assignment. Repeated measures were sometimes extended to more than two time points. Crudele and Raffaghelli (2022) had participants construct a series of three argument maps, and after each mapping activity they were given a test evaluating critical thinking ability and a separate test evaluating the ability to comprehend argumentative text. A quite different approach in some of the single-group quantitative studies was to correlate features of argument maps with characteristics of the students who constructed them.

For reasons discussed in the introduction, the 50 studies which quantitatively compared treatment groups were of particular interest. The research questions investigated in these studies, represented as pairs of independent and dependent variables, are described in a later section of the results. In the immediately following sections, we report on major research themes investigated in the studies.

3.3 | Attitudes Toward Argument Mapping

Forty-six studies explored student attitudes toward argument visualisation through surveys or interviews. Overall, students expressed favourable views on integrating argument mapping into their learning and remarked on the convenience of computer-aided argument visualisation tools. Many students described argument mapping as an engaging approach to learning argumentation and advanced thinking skills, emphasising its role in helping them visualise, organise, and evaluate their thinking (Pinho et al. 2009; Robillos and Thongpai 2022), maintain

learning interest (Kaya 2024), consider problems from different perspectives (Ngajie et al. 2020), promote understanding and appropriate use of different argumentation components (Robillos and Art-in 2023), and use time more efficiently in argument construction (Pinho et al. 2009; Robillos and Art-in 2023; Yıldızlı and Şimşek 2020). Argument mapping was perceived to be effective in directing learners' attention to the structure of arguments, helping them avoid the tendency to express ideas haphazardly or without clear focus (Kaeppel 2021). Furthermore, students shared that argument mapping facilitated collaboration, enhancing both the quality and frequency of idea sharing among group members. They saw it as fostering inclusivity by encouraging participation from students who were often hesitant to share their ideas in group discussions (Kaya 2024; Ouyang et al. 2024).

Several lessons emerged from the negative experiences shared by participants. For learners new to argument mapping, it is essential that the interface be intuitive and user-friendly to minimise extraneous cognitive load. Additionally, clear and detailed guidance is crucial to help students navigate both the tool and the associated learning process (Nesbit et al. 2024; Wu et al. 2013, 2014). A commonly reported concern across studies was the insufficient time allocated for completing argument mapping exercises. For those unfamiliar with argument mapping, completing the tasks successfully required more time than initially anticipated (Piwek 2013; Uçar and Demiraslan Çevik 2021). Some students noted that argument mapping initially felt messy and that acquiring sufficient familiarity took time, but they encouraged persistence and underscored the importance of instructor and peer support during this phase. The provision of sample maps was also cited as a helpful resource for guiding students through the initial learning curve (Kaya 2024; Nesbit et al. 2024; Ngajie et al. 2020). These insights inform how to effectively structure student training in argument mapping and better prepare learners for successful engagement with the argument visualisation tool and the activities it supports.

3.4 | Critical Thinking and Argumentation Ability

By far the most common thinking skills investigated by the reviewed studies were critical thinking and argumentation ability, and only a few investigated skills such as problem solving, scientific inquiry and self-regulation. Across studies, the concepts of argumentation and critical thinking seemed to be used interchangeably, and we found no attempt to make conceptual distinctions between them beyond general comments that argumentation is a component or foundation for critical thinking. The ability to think critically and argue was assessed by a variety of methods—researcher-constructed tests (e.g., Harrell 2008), standardised instruments such as the California Critical Thinking Skills Test (CCTST, Eftekhari et al. 2016) or the Halpern Critical Thinking Assessment (HCTA, Dwyer et al. 2012), general argument items selected from the Law School Admissions Test (LSAT, Pinkwart et al. 2008), analysis of oral arguments using Toulmin's Argumentative Pattern model (TAP; Darmawansah, Hwang, and Lin 2024), and qualitative analysis of argument maps (Muller Mirza et al. 2007). Following a research design similar to other single-group quantitative studies, Crudele and Raffaghelli (2022) developed a 4-day course about argumentation that featured lectures, exercises, and argument mapping. The students were given a pretest, midpoint test, and posttest on critical thinking based on the Holistic Critical Thinking Scoring Rubric (HCTSR, Facione and Facione 2014). The researchers found a significant gain in critical thinking ability over the 4 days ($d = 0.87$).

3.5 | Evaluation of Argument Maps

Researchers evaluated student-constructed argument maps to gather research data in 49 of the studies. Various methods were employed to examine the structure and quality of the maps. For example, some studies assessed task completeness by comparing maps with assigned criteria, such as requiring a given number of supporting or opposing reasons (e.g., Fysaraki et al. 2016; Nesbit et al. 2024). Others, often drawing on Toulmin's model of argumentation (Toulmin 2003), examined structural complexity by counting the number of nodes and connections and categorising them based on their functions, such as supporting arguments, evidence, counterarguments and rebuttals (e.g., Gao et al. 2023; Liu et al. 2024). Some researchers developed grading rubrics focusing on criteria such as clarity, relevance and logical coherence (e.g., Archila, Barbosa, et al. 2022; Archila, Gravier, et al. 2022; Chen et al. 2022). Rapanta and Walton (2016b) proposed a fallacy-focused method for evaluating argument maps, examining why and where the flaws of reasoning emerge. Beyond argumentation-related features, some studies analysed the accuracy and coverage of domain knowledge represented in the maps, assessing how well they captured key concepts and relationships within a given subject area (e.g., Iandoli et al. 2014). Many studies applied a combination of two or more types of analysis to gain a comprehensive understanding of the argument maps' structure, quality and knowledge representation.

While manual coding was commonly employed, with inter-rater reliability checks ensuring validity, some studies utilised software tools like *jMap* to log and facilitate the analysis of the processes involved in constructing argument maps, which may not be easily captured by human coders (e.g., Jeong and Seok-Shin 2023).

3.6 | Collaborative Mapping

Students collaborated to construct argument maps in 41 of the reviewed studies. Collaborative argument mapping could potentially represent students' perspectives at the individual, collaborating group and whole-class levels (Rapanta and Walton 2016b; Ouyang et al. 2024) and may facilitate the exchange of perspectives, peer negotiation, reflection on others' viewpoints, and the joint construction of collective knowledge (Kimmerle et al. 2021; Onrubia and Engel 2009; Zheng et al. 2023). Loll and Pinkwart (2013) employed a within-subjects design to investigate the effects of collaboration on argument mapping, with each participant engaging in both individual and group phases. Participants were required to argue on one topic alone and on another topic in a group of three. To control for potential confounding variables, the researchers employed counterbalancing, so half of the participants started with the group phase, and the other half started with the individual phase. The study found

that collaboration led to argument maps with a higher number of duplicate elements, suggesting that the expected peer review process within groups did not occur. However, groups also provided a greater variety of perspectives and made more original contributions (i.e., contributions not derived from given material), resulting in more elaborate argument maps.

Almost all of the studies involving collaborative argument mapping used computer-aided tools to facilitate argument map construction ($n = 38$). Even though collaborative argument mapping can be implemented without computers, students can benefit from using argument mapping tools (e.g., Kialo Edu and LASAD) that mediate collaboration via synchronous chat or simultaneous co-editing of maps.

Collaborative argument mapping has been implemented across various learning environments, including in-class sessions ($n = 34$), out-of-class activities ($n = 11$), and laboratory research settings ($n = 23$), demonstrating its adaptability to different instructional contexts. In most cases, participants worked in pairs or in small groups of up to five members. Larger groups may negatively impact overall outcomes due to the increased need for coordination (Loll and Pinkwart 2013). To enhance student engagement and interaction, researchers have proposed several strategies, such as collaboration scripts (Scheuer et al. 2014), role-play (Darmawansah, Hwang, and Lin 2024; Darmawansah, Rachman, et al. 2024), clearly specified peer-review processes (Loll and Pinkwart 2013), and collective reflection (Darmawansah et al. 2022), which have been found effective in encouraging diverse perspectives, promoting dialectical thinking, and facilitating the integration of arguments from different sides.

3.7 | Causal Relationships

In this section we identify the most frequently investigated research questions—as defined by reoccurring pairs of independent and dependent variables—in studies which compared treatment groups. We then explore the methods and results in representative studies. Of the 50 between-group intervention studies, 33 used argument mapping as an independent variable, meaning that in each study at least one treatment group participated in an argument mapping learning activity and at least one group did not.

3.7.1 | What Effect Does Argument Mapping Have on Critical Thinking/Argumentation Ability? (14 Studies)

The most frequently examined research question among the experimental studies was the effect of argument mapping on critical thinking or argumentation skills. We found 11 studies that seemed eligible for a meta-analysis on that topic. For example, Pinkwart et al. (2008) used a randomised matched pairs design to assign 70 students to use either the LARGO argument mapping feature or a text-based argument representation. They gave pretests and posttests assessing reasoning in legal contexts and 'everyday argumentation' and found no significant advantages for the argument mapping treatment. A quite different result was found by Cullen et al. (2018) who gave intensive training in argument mapping and argument analysis to students who attended an undergraduate philosophy seminar and used LSAT logical reasoning questions as pretest and posttest to compare with students attending another philosophy course which did not involve argument mapping. The students attending the argument mapping seminar improved their logical reasoning by the end of the 12-week seminar ($d = 0.77$) and obtained significantly higher gain scores than the control group ($d = 0.71$). It is likely, though, that the argument analysis components in the seminar that were unrelated to mapping contributed to much of this difference, and therefore the benefit solely attributable to the argument mapping component is not quantifiable. Harrell (2011) found that students taught argument mapping in one section of an introductory philosophy course showed significantly greater gains on a critical thinking test than students in other sections who were not taught argument mapping. Notably, Harrell (p. 380) found argument mapping was much more advantageous for students with low pretest scores ($g = 0.91$) than students with high pretest scores ($g = 0.07$).

3.8 | What Effect Does Argument Mapping Have on Acquisition of Domain Concepts? (18 Studies)

We defined domain concepts as knowledge specific to the subject of a course from which students were recruited for participation or the factual and conceptual content of readings or learning activities assigned in the research. The studies in our dataset usually evaluated domain concepts via a posttest administered after a learning activity. Dwyer et al. (2010) randomly assigned a large sample of 400 students taking an introductory psychology course to study preconstructed argument maps or content-equivalent texts about psychology. Immediately after the study activity, the students were given a comprehension test that involved reasoning with the studied information. One month later, they were given a cued-recall test that assessed memory for the information. Although there was no statistically detectable difference in the comprehension test, the argument map groups significantly outperformed the text groups in the retention test.

Moving to a consideration of argument map construction, Dwyer et al. (2013) recruited 136 first-year Arts students to compare the effects of three active studying techniques—argument mapping, outlining and summarization—on recall of concepts from an argumentative text. The argument mapping group significantly outperformed the summarization group ($d = 1.30$), and though they also outperformed the outlining group ($d = 0.40$), the difference was not statistically detected. A study by Çoban (2013) investigated the effect of having science education students individually construct argument maps and then use the maps to discuss their laboratory observations with peers. Students' understanding of scientific processes and the epistemological foundations of science was measured by separate posttests. The control group performed the same scientific inquiry activities as the experimental group but without argument mapping and without the opportunity to discuss their laboratory observations with classmates. The experimental group outperformed the control group on both the scientific processes posttest ($g = 0.69$) and the epistemology of science posttest ($g = 0.70$). For our purposes, interpreting this result is challenging because the research

design does not allow a distinction between the effects of argument mapping and peer-to-peer discussion.

Due to the way we defined thinking skills and domain concepts, a study conducted within a course which had critical thinking or argument structure as a primary goal was tagged with both codes. For example, Kunsch et al. (2014) investigated the effects of argument mapping in a master-level course on critical thinking in business and used the standardised Business Critical Thinking Test as a posttest. Because the dependent variable met our definition for both a thinking skill and a domain concept, it was counted as featuring both the present research question and the previous one. A substantial number of thinking skill studies were 'double-counted' in this way, and we found very few studies that investigated the effects of argument mapping on the acquisition of declarative knowledge like that assessed in Dwyer et al. (2010) and Dwyer et al. (2013).

3.8.1 | What Effect Does Argument Mapping Have on Writing Skills? (6 Studies)

Undergraduate students are known to have difficulty deploying key argumentative components such as evidence and counterarguments in their writing (Jumariati et al. 2021). Researchers have theorised that visualising arguments and using argument mapping as part of the writing process can help students develop more advanced argument schemas and apply them in their writing (Liu et al. 2024; Nesbit et al. 2024).

The previously mentioned study by Cullen et al. (2018) investigated the effects of argument mapping on argument essay writing scores as well as LSAT scores. Students who took the argument mapping seminar significantly outperformed the control group on all essay elements scored by rubric: structure ($d = 0.76$), accuracy ($d = 0.74$) and understanding ($d = 0.97$). But again, these differences may be due to the argument analysis components of the seminar and not argument mapping per se. Working with 201 students learning English as a foreign language, Liu et al. (2024) investigated the direct and transfer effects of argument mapping as a pre-writing activity. In the first week, all participants received an instructor-led tutorial on argument structure and analysis. The tutorial received by the experimental group introduced the argument mapping software (DMap), and the explanation of argument components was illustrated with reference to the software interface. The tutorial received by the control group explained the same argument terms and components but without argument visualisation. Then, while the argument map group constructed an argument map using information from a source text, the control group completed a vocabulary exercise using information from the same source text. In the second week, both groups were asked to write an argument essay based on the source text. In the third week, without access to the argument mapping tool, all participants wrote a transfer argument essay about a different topic. The experimental group wrote significantly more counterarguments in both the initial essay and transfer essay. In a study of 81 students taking a first-year course on critical thinking and writing, Harrell and Wetzel (2013) compared the argument essays of students taught to analyse arguments by argument mapping with those who had been taught argument analysis without visualisation methods. On two writing tasks, those who received training in argument mapping significantly outperformed the control participants on several key elements such as the use of counterarguments, evidence and internal connections (i.e., coherence).

3.8.2 | What Effect Does Argument Mapping Have on Motivation to Learn (2 Studies)

There are a variety of self-report measures devised by psychologists to assess aspects of students' motivation to learn. Different from attitudes toward or expressed preference for a recently experienced learning activity, these are psychometrically validated instruments which tap psychological constructs such as self-efficacy (MSLQ; Pintrich et al. 1993), need for cognition (Liu and Nesbit 2024) and intrinsic motivation (AMS; Vallerand et al. 1992). We found only two between-subject studies which assessed the effects of argument mapping on such constructs. Dwyer et al. (2012) gathered data from 74 first-year psychology students. The experimental group was enrolled in an 8-week course on critical thinking that made extensive use of argument mapping. The control group received no critical thinking intervention beyond their regular coursework. Although the experimental group significantly outperformed the control group on a critical thinking posttest, there was no effect of treatment on need for cognition or motivation to learn as measured by eight subscales of the Motivated Strategies for Learning Questionnaire.

3.8.3 | What Effect Do Different Instructional Designs for Argument Mapping Have on Thinking Skills? (10 Studies)

Every argument mapping activity is embedded within a multifeatured instructional design. The mapping can be done individually or collaboratively, with pen and paper or with software tools, in-class or for homework, and so on. Several argument mapping studies investigated the effects of collaborative argument mapping on thinking skills. Memiş et al. (2023) had 50 science education students engage in argument-based inquiry learning every week for 8 weeks. Each weekly activity featured individual argument mapping with Rationale. All participants took the Test of Logical Thinking (TOLT) as a pretest and posttest. An ANCOVA using the pretest as a covariate found that students randomly assigned to participate in additional weekly collaborative argument mapping activities each week scored significantly higher on the posttest ($d = 0.97$). Using a similar research design, Uçar and Demiraslan Çevik (2021) administered an argument skills instrument to pre-service teachers as pretest and posttest. All students engaged with individual argument mapping using Argunet in a series of 4 weekly classes. Students who received peer feedback on their maps in each class attained significantly higher argument skill gain scores ($d = 0.86$). Chounta et al. (2017) assigned students taking an introductory philosophy course to collaborate synchronously in groups of 2–4. Each collaborating group discussed an assigned theme and co-constructed maps using LASAD. Compared with students who worked individually, students who collaborated achieved greater gains in conceptual knowledge about argument mapping as measured by pre- and post-tests. The collaborating students took slightly more time in the activity but constructed maps that were more complex and of higher quality.

Researchers have proposed that argument mapping with specially designed software tools may be more efficient in developing students' argument schemas than pencil and paper because they support more frequent and extensive revision of arguments and allow the creation of larger and more complex maps. Eftekhari et al. (2016), for example, deployed the CCTST as a pretest and posttest and found that students who used the Rationale software tool for an extended series of argument mapping assignments gained significantly more in critical thinking ability than a group that used paper and pencil ($d = 0.45$).

3.8.4 | What Effect Do Different Instructional Designs for Argument Mapping Have on Domain Concepts? (11 Studies)

The instructional designs in which argument mapping is embedded can potentially determine the extent to which students acquire key concepts that were the topic of the constructed maps. We found very few studies which measured, as a dependent variable, gains in factual or declarative domain knowledge distinct from understanding of argumentative structures (Dwyer et al. 2013; Eftekhari and Sotoudehnama 2018; Scheuer et al. 2014). Fewer still investigated the effects of instructional design on that type of dependent variable. Dwyer et al. found that students who individually studied an argument map in a lecture hall at the same time as others recalled more information from the map but scored no higher on comprehension than students who studied the map in an isolated booth. They speculated that the lecture hall setting promoted performance goal motivation, which activated rote learning strategies more strongly than meaningful learning strategies.

3.8.5 | What Effect Do Different Instructional Designs Have on Mapping Processes and Products? (8 Studies)

The instructional designs in which argument mapping is embedded can affect students' mapping behaviours and the resulting map structures. Scheuer et al. (2014) developed a protocol for collaborative argument mapping in LASAD in which dyads first studied texts individually and then collaborated via chat to produce an argument map and a joint conclusion. While dyads in a control group followed the basic protocol, dyads in an experimental group were asked to engage in constructive criticism of their partner's position using sentence openers like 'An argument against that point is...' The researchers found no significant difference between the two groups on a posttest of factual knowledge of the discussed topics. The experimental group did, however, produce discussions with a significantly higher proportion of logically elaborated statements ($d = 0.82$), and they had significantly stronger beliefs in the efficacy of the activity for learning argumentation. Also seeking to enhance collaborative argument mapping, Iandoli et al. (2014) developed, as an addition to the Cohere mapping tool, a social augmentation dashboard outside the map window which provided learners with information about other participants and the interaction process. An experimental group and a control group both engaged in collaborative mapping, but only the experimental group was provided with the dashboard. The study found no significant difference in the quality of arguments between the two groups, but the experimental group showed significantly wider coverage of the discussion topics as well as fewer off-topic posts and fewer repeated statements. The experimental group also had more positive attitudes about the usability of the tool and the quality of the collaboration.

3.8.6 | Does the Type of Argument Map Matter? (6 Studies)

As we have noted, argument maps come in a variety of graphical types or formats. So, does the type of map have any effect on instructional outcomes? When it comes to studying argument maps, the previously mentioned study by Dwyer et al. (2010) found that using colour to redundantly distinguish between supporting and opposing reasons (objections) had no effect on memory or comprehension of the map content. Turning to argument construction, Zumbach et al. (2008) investigated the effects of argument map type on domain knowledge and argument quality. They had each learner study a multimedia genetics tutorial by constructing one of three types of argument representation based on the information presented in the tutorial: (a) a node-link type argument map, (b) a node-link argument map augmented with images from the tutorial, (c) two columns of text with supporting reasons listed on the left and opposing reasons listed on the right. As discussed in our introduction, we consider the spatial division of pro and con reasons as a key strategy for argument visualisation, and therefore we treat the research design of Zumbach et al. as a comparison of three types of argument maps. Although the students who constructed the two-column text representation produced the most reasons and obtained the highest scores on the knowledge posttest, the differences were not statistically significant. A similar result was obtained by Loll and Pinkwart (2013) who randomly assigned each participant to construct one of three types of node-link argument maps: (a) a simple map allowing only three types of relationships between statements (pro, con, undefined), (b) a highly structured map based on the multiple statement and relationship types in Toulmin's model (Toulmin 2003) and (c) a domain-specific map adapted from Belevedere (Suthers 2003). An argumentation ability posttest based on LSAT items found no significant difference between the three treatment groups, and the same for a knowledge posttest based on the argued domain.

While improving learning outcomes may be the overarching goal for researchers who evaluate argument map formats and tools, it is also important to investigate facilitating factors such as usability and perceived usefulness. Le et al. (2013) examined the benefits of adding a mini-map feature to LASAD. Mini-maps are commonly used as a navigational feature in games and online geographical maps. The researchers found that introducing mini-maps decreased the time to construct an argument by 10% and improved user ratings of the interface's efficiency and aesthetic appeal.

4 | Discussion

4.1 | Student Attitudes

There is plentiful evidence in the reviewed studies that students have positive attitudes toward argument mapping and recognise

its value for learning how to construct good arguments. Many said that argument mapping should be assigned more often by instructors (Nesbit et al. 2024). The reasons may lie in comparisons they make with writing argument essays, which demand nearly simultaneous attention to the challenges of argumentation, writing mechanics and writing style.

At this stage, the most important goal for attitudinal research on argument mapping may be the development of a standardised instrument that could be adapted or localised for situational requirements. The availability of a standardised questionnaire would greatly facilitate comparison of results from individual studies. The instrument would go beyond the holistic question of whether students like argument mapping by assessing attitudes toward specific features of the task (collaboration, use of software, grading rubric, etc.) and would be evaluated for psychometric reliability and validity. Researchers should, of course, adopt best practices for instrument development and the analysis of attitudinal data (Lovelace and Brickman 2013).

4.2 | Effects on Argumentation Ability, Critical Thinking and Argumentative Writing

We found that, with notable exceptions, the single-group studies and experimental studies tended to find that argument mapping is effective in developing general argumentation and critical thinking skills (e.g., Harrell 2011). Argument mapping was also found to improve the quality of written arguments (Cullen et al. 2018; Harrell and Wetzel 2013; Liu et al. 2024). In our view, the weight of evidence supports a recommendation that instructors use argument mapping to develop critical thinking and argumentation skills.

The effect of argument mapping on these skills was the only causal relationship investigated by research of sufficient quantity and quality to support a conventional meta-analysis. The 11 experimental and quasi-experimental studies that investigated this research question had mixed results, and a meta-analysis is needed to determine the strength and confidence level of the effect. Considering the small number of eligible studies from post-secondary education, the meta-analysis should expand the scope to cover other levels of education and introduce level of education as a moderator. We found the studies varied greatly in how well the research design isolated argument mapping as an independent variable. In several of them, the experimental treatment provided additional instruction beyond argument mapping that was not replicated in the control treatment. Therefore, any meta-analysis should capture the quality of the research design as another moderator.

4.3 | Collaborative Versus Individual Argument Mapping

Given the firmly established efficacy of collaborative and cooperative learning in postsecondary education (Swanson et al. 2019; Johnson et al. 2014), there is a well-founded expectation that collaboration will enhance the benefits of argument mapping. There are also theoretical reasons to imagine that collaboration and argument mapping could work synergistically (Kaeppel 2021; Ouyang et al. 2024). As mentioned, several studies reported that students who collaborated outperformed those who did not on various outcome measures. The problem, though, is that there are insufficient studies that directly compare collaborative and individual mapping on equivalent learning tasks, and the very few that make that comparison do not examine the same or similar dependent variables. Most often a collaborative mapping treatment is combined with other features not available in the individual mapping treatment, such as a group discussion or an additional opportunity for learning. Consequently, the research base has not developed to the stage where a meta-analysis of collaborative versus individual argument mapping is feasible.

4.4 | Gaps in the Research

4.4.1 | Arguing to Learn?

There is a surprising dearth of research on the effects argument mapping has on conceptual change and the acquisition of domain knowledge beyond argumentation and critical thinking. Although several studies investigated the effects of argument mapping in the context of inquiry learning activities (e.g., Çoban 2013), the research designs only rarely examined its effects on content knowledge, and never the effects on conceptual change in the subject domain. Science education research has strongly established that reading refutational texts, texts that describe misconceptions and use evidence to rebut them, is effective in helping learners overcome common and persistent misconceptions (Guzzetti 2000). There is one study (Liu and Nesbit 2018) which found that studying a 'refutational map' (i.e., an argument map) produced conceptual change in postsecondary learners' understanding of the motion of objects, but there is no equivalent research for constructing argument maps.

4.4.2 | Assessment Studies

Although plenty of the reviewed studies scored or otherwise assessed student argument maps, there were only five empirical studies (Kubosawa et al. 2013; Lynch et al. 2009a, 2009b, 2014; Rapanta and Walton 2016b) primarily concerned with methods for assessing student maps. More research is needed on reliable and valid argument map assessment. Not only are assessment studies needed to inform instructors in designing and scoring rubrics, but they are also crucial for researchers who investigate argument map quality and structure as dependent variables.

4.4.3 | Argument Mapping Tools

There is a need for comparative research on argument mapping tools and their features. While previous reviews (e.g., Davies 2019; Harrell 2005; Scheuer et al. 2010) have catalogued and described the available tools, primary studies that vary features and provide evidence of their impact are essential for promoting progress in the development of argument mapping tools. Design-based research (Hoadley and Campos 2022) may be the most efficient methodology for realising this goal.

4.4.4 | Mapping Tools That Track and Adapt to Learning Trajectories

There is some evidence (Scheuer et al. 2010) that complex mapping features can hinder learning, but it may be that more advanced learners could benefit from those same features due to an expertise reversal effect (Kalyuga 2021). More research is needed on mapping tools that track student progress in argument mapping and add next-level features as an adaptation to demonstrated growth in expertise.

4.4.5 | Reporting Issues

We observed several recurring problems with reporting the methods and results of the research. Some studies reported means but not standard deviations, creating challenges for calculating effect sizes. Some were unclear on whether participants were randomly assigned to treatment groups. Others did not report basic information about the procedure, such as which software tool was used by students and the physical and social setting in which argument mapping activities were performed. Another observed shortcoming was the failure to distinguish between self-report instruments which assess learners' self-perceptions and performance instruments which assess abilities. In one case, for example, the authors identified a questionnaire asking students about their problem-solving abilities as a test of problem-solving ability.

4.4.6 | Research Design Issues

A common research design issue in argument mapping studies is the failure to isolate independent variables, a design flaw which can obscure the specific effects of argument mapping and other factors on learning outcomes. For example, several studies investigated the effects of collaboration while argument mapping without distinguishing its impact from individual argument mapping, making it difficult to determine whether observed benefits stem from the mapping process or from collaboration. Confounded independent variables are a barrier to establishing clear causal relationships and assessing the unique contributions of argument mapping.

4.5 | Limitations of This Review

The results reported in any systematic review are directly determined by, and therefore limited to, the variables or codes chosen by the researchers; and that choice is a subjective decision. Where one researcher may choose to delve into a feature reported in the research by defining multiple relevant variables, each with a range of values, another researcher might choose only a single binary variable to code the feature. For example, while we used a single binary value to report whether the research assessed student argument maps, other reviewers might have developed multiple codes indicating the type of assessment (e.g., qualitative or quantitative). In hindsight, we should not have used Thinking Skills, broadly defined, as a coded variable. It reliably captured key studies dealing with critical thinking or argumentation skill as a dependent variable but was much less reliable when applied to other abilities that researchers identified as science process skills, metacognitive skills, self-regulatory skills, dialogic skills, and so on. This led to challenges in establishing sufficient interrater reliability for the code. A better approach would have defined a code that only covered the use of critical thinking and argumentation ability as a dependent variable.

4.6 | Implications for the Design of Learning Activities

Taken together, the results of this review support the adoption of argument mapping as an effective method for teaching and learning in higher education. In translating from research to practice, and from the science of learning to the craft of teaching, we must consider argument mapping in the context of a postsecondary classroom with all its constraints and opportunities. With these in mind, we present six principles that instructors can adopt to successfully incorporate argument mapping in their teaching. While evidential support varies across the principles, each has grounding in educational research and theory.

1. *Describe for students the benefits of argument mapping.* When introducing students to argument mapping, emphasise its known advantages. Tell them they are likely to enjoy argument mapping or at least prefer it over other methods you could have adopted. As we have seen, students generally have positive attitudes toward argument mapping and prefer it to argument writing assignments (Nesbit et al. 2024). There is some evidence that their positive attitudes are based on an understanding of the functional value of the maps for guiding them as they create and assemble argument components (Kaeppel 2021). Advise students that repeated argument mapping will advance their critical thinking skills and academic writing, especially if those abilities are underdeveloped (Harrell 2011).
2. *Use argument mapping software.* Rather than have students draw maps on paper or use generic graphical software, have them use software specifically designed for argument mapping. Although there is only limited research supporting the use of specialised software (e.g., Eftekhari et al. 2016), the greater ease it allows for editing maps probably means that students are more likely to engage in repeated revision of their maps and hence have more opportunities for rehearsing the structure of arguments. This principle should not preclude the use of hand-drawn maps when first introducing the concept of argument mapping or when specialised software is unavailable.
3. *Don't skimp on pretraining.* Provide sufficient prior instruction in the theory and practice of argument mapping, including guided practice in the use of software. If one of the instructional goals is to develop students' critical thinking, then the time dedicated to this preparatory phase should not be considered an add-on cost because learning about the structure and function of argument maps is learning about argumentation itself. Indeed, several of the studies we reviewed obtained large effects solely as a result of argument map training (e.g., Harrell and Wetzel 2013). Despite its well-documented benefits, pretraining is often neglected as a component of multimedia learning

(Mayer 2020). As part of an introduction to argument mapping, we recommend showing samples of completed maps and explaining the function of each of their components (e.g., Kaya 2024).

4. *Ensure sufficient engagement to develop mastery*. Provide multiple, progressive argument mapping activities with feedback for each activity. Like any worthwhile skill, argument mapping requires practice to achieve mastery (van Gelder 2005, 2015), as does argumentation in general. Researchers have found that courses which provided "lots of argument mapping practice" (LAMP), about one exercise per week for a semester, obtained significantly greater gains in critical thinking ability than courses which provided only a few practice activities (Alvarez 2007; Rider and Thomason 2014).

5. *Adopt both collaborative and individual argument mapping*. Although individual argument mapping with feedback is beneficial, instructors should also provide students with multiple collaborative argument mapping activities. There is evidence that adding collaborative argument mapping activities to individual mapping activities can produce large learning gains (Memiş et al. 2023). We recommend consulting the extensive literature on collaborative argumentation (e.g., Asterhan and Schwarz 2016) for ideas on developing collaborative argument mapping activities.

6. *Integrate argument mapping within writing assignments*. If developing argumentative writing is one of the instructional goals, initiate a writing assignment with an argument mapping activity on the same topic (Liu et al. 2024). Students will be able to develop the logic of their argument before working on its style and emotional appeal via the writing. This approach manages the intrinsic cognitive load of the task so that students are not overwhelmed by having to deal simultaneously with the argument structure and the writing.

Although there is much more that can be said about the pedagogy of argument mapping (e.g., Davies et al. 2021), the six principles presented here summarise the subset of that pedagogy supported by currently available research. A clearer and more complex picture will no doubt emerge as more well-designed studies and meta-analyses building on them are published.

5 | Conclusion

The purpose of this review was to comprehensively survey research on argument mapping in higher education to understand its scope, methods and findings. We found that researchers deployed many different argument mapping tools, most frequently Rationale (Van Gelder 2007). The studies investigated a wide range of research questions relating to student attitudes and motivation, critical thinking and argumentation skills, writing skills, the argument mapping process, features of argument maps, collaborative argument mapping and acquisition of domain concepts. Many studies examined the effects of argument mapping on critical thinking, and a substantial number of those had a control group. Although the results from these were mixed, they tended to find that argument mapping is an effective method for developing critical thinking. Meta-analysis is needed to establish the size and reliability of the effect. In contrast, few studies examined the effect of argument mapping on acquisition of domain-specific factual or conceptual knowledge. These offered, however, some evidence for the efficacy of argument mapping as a study technique (Dwyer et al. 2013). Notably, no studies examined the effects of map construction on conceptual change. Argument mapping appears to enhance essay quality, particularly in structure, use of evidence and incorporation of counterarguments. Further, the benefits of argument mapping may be enhanced by collaborative mapping, peer feedback, and use of mapping software.

Considering the central role of critical thinking and argumentation in undergraduate education, these findings underscore the potential of argument mapping as a valuable instructional innovation. For instructors, its integration into higher education curricula offers an evidence-based means of strengthening student learning, while for researchers, important opportunities remain to investigate underexplored applications such as promoting conceptual change.

Author Contributions

John C. Nesbit: conceptualization, data curation, formal analysis, visualisation, writing – original draft, methodology, investigation, writing – review and editing, supervision, funding acquisition. **Qing Liu:** conceptualization, data curation, formal analysis, writing – original draft, methodology, investigation, writing – review and editing, funding acquisition.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The article is a systematic review. We have a spreadsheet consisting of a list of the reviewed studies (rows) and codes we assigned to the studies (columns). It is available upon request by contacting the first author.

References

*References marked with an asterisk indicate studies included in the systematic review.

*Alsmari, N. A. 2022. "Integrating Computer-Aided Argument Mapping Into EFL Learners' Argumentative Writing: Evidence From Saudi Arabia." *International Journal of Advanced Computer Science and Applications* 13, no. 10.

Alvarez, C. 2007. "Does Philosophy Improve Reasoning Skills?" Master's Thesis, University of Melbourne.

*Andiliou, A., and P. K. Murphy. 2014. "Creative Solutions and Their Evaluation: Comparing the Effects of Explanation and Argumentation Tasks on Student Reflections." *Frontline Learning Research* 2, no. 3: 92–114.

*Archila, P. A., V. Barbosa, G. Gravier, et al. 2022. "Integrating Peer Feedback and Instructor Feedback to Support the Construction of Bilingual Scientific Argument Maps." *International Journal of Science Education* 44: 2283–2305.

*Archila, P. A., G. Gravier, L. Levy, et al. 2022. "Using Formal and Informal Formative Assessment to Support Bilingual Argument Mapping in University Bilingual Science Courses." *International Journal of Bilingual Education and Bilingualism* 25, no. 9: 3420–3435.

*Ashley, K., N. Pinkwart, C. Lynch, and V. Aleven. 2007. "Learning by Diagramming Supreme Court Oral Arguments." In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, 271–275. ACM.

*Ashley, K. D. 2009. "Teaching a Process Model of Legal Argument With Hypotheticals." *Artificial Intelligence and Law* 17: 321–370.

*Ashley, K. D., and C. Lynch. 2010. "Instructional Explanations in a Legal Classroom: Are Students' Argument Diagrams of Hypothetical Reasoning Diagnostic?" In *Instructional Explanations in the Disciplines*, edited by M. K. Stein and L. Kucan, 171–188. Springer.

Asterhan, C. S. C., and B. B. Schwarz. 2016. "Argumentation for Learning: Well-Trodden Paths and Unexplored Territories." *Educational Psychologist* 51, no. 2: 164–187.

*Baker, D. P. 2012. "Making Good Better: A Proposal for Teaching Ethics at the Service Academies." *Journal of Military Ethics* 11, no. 3: 208–222.

*Barstow, B., L. Fazio, J. Lippman, M. Falakmasir, C. D. Schunn, and K. D. Ashley. 2017. "The Impacts of Domain-General vs. Domain-Specific Diagramming Tools on Writing." *International Journal of Artificial Intelligence in Education* 27: 671–693.

*Barstow, B., L. Fazio, C. Schunn, and K. Ashley. 2017. "Experimental Evidence for Diagramming Benefits in Science Writing." *Instructional Science* 45: 537–556.

*Barstow, B. J., C. D. Schunn, L. K. Fazio, M. H. Falakmasir, and K. Ashley. 2015. "Improving Science Writing in Research Methods Classes Through Computerized Argument Diagramming." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 37. https://escholarship.org/uc/item/1b28v9cz.

Belur, J., L. Tompson, A. Thornton, and M. Simon. 2021. "Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making." *Sociological Methods & Research* 50, no. 2: 837–865. https://doi.org/10.1177/0049124118799372.

*Butchart, S., D. Forster, I. Gold, et al. 2009. "Improving Critical Thinking Using Web-Based Argument Mapping Exercises With Automated Feedback." *Australasian Journal of Educational Technology* 25, no. 2: 268–291.

*Carrington, M., R. Chen, M. Davies, J. Kaur, and B. Neville. 2011. "The Effectiveness of a Single Intervention of Computer-Aided Argument Mapping in a Marketing and a Financial Accounting Subject." *Higher Education Research and Development* 30, no. 3: 387–403.

*Chen, X., L. Wang, X. Zhai, and Y. Li. 2022. "Exploring the Effects of Argument Map-Supported Online Group Debate Activities on College Students' Critical Thinking." *Frontiers in Psychology* 13: 856462. https://doi.org/10.3389/fpsyg.2022.856462.

*Chen, X., H. Zhao, H. Jin, and Y. Li. 2024. "Exploring College Students' Depth and Processing Patterns of Critical Thinking Skills and Their Perception in Argument Map (AM)-Supported Online Group Debate Activities." *Thinking Skills and Creativity* 51: 101467.

*Chounta, I. A., B. M. McLaren, and M. Harrell. 2017. "Building Arguments Together or Alone? Using Learning Analytics to Study the Collaborative Construction of Argument Diagrams." In *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer-Supported Collaborative Learning (CSCL) 2017*, edited by B. K. Smith, M. Borge, E. Mercier, and K. Y. Lim, 589–592. International Society of the Learning Sciences.

*Çoban, G. Ü. 2013. "The Effects of Inquiry Supported by Argument Maps on Science Process Skills and Epistemological Views of Prospective Science Teachers." *Journal of Baltic Science Education* 12, no. 3: 271–288.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20: 37–46. https://doi.org/10.1177/001316446002000104.

*Crudele, F., and J. Raffaghelli. 2022. "Promoting Critical Thinking Through Argument Mapping: A Lab for Undergraduate Students." *Journal of Information Technology Education: Research* 22, no. 1: 497–525.

*Cullen, S., J. Fan, E. van der Brugge, and A. Elga. 2018. "Improving Analytical Reasoning and Argument Understanding: A Quasi-Experimental Field Study of Argument Visualization." *Npj Science of Learning* 3, no. 1: 21.

*Darmawansah, D., G. J. Hwang, and C. J. Lin. 2024. "Integrating Dialectical Constructivist Scaffolding-Based Argumentation Mapping to Support Students' Dialectical Thinking, Oral and Dialogical Argumentation Complexity." *Educational Technology Research and Development* 72: 1–29.

*Darmawansah, D., C. J. Lin, and G. J. Hwang. 2022. "Empowering the Collective Reflection-Based Argumentation Mapping Strategy to Enhance Students' Argumentative Speaking." *Computers & Education* 184: 104516.

*Darmawansah, D., D. Rachman, F. Febiyani, and G. J. Hwang. 2024. "ChatGPT-Supported Collaborative Argumentation: Integrating Collaboration Script and Argument Mapping to Enhance EFL Students' Argumentation Skills." *Education and Information Technologies* 30: 1–25.

*Davies, M. 2008. "'Not Quite Right': Helping Students to Make Better Arguments." *Teaching in Higher Education* 13, no. 3: 327–340. https://doi.org/10.1080/13562510802045352.

*Davies, M. 2009. "Computer-Assisted Argument Mapping: A Rationale Approach." *Higher Education* 58: 799–820.

*Davies, M. 2011. "Concept Mapping, Mind Mapping and Argument Mapping: What Are the Differences and Do They Matter?" *Higher Education* 62: 279–301.

*Davies, M. 2019. "New Directions in the Teaching of Critical Thinking." *Change: The Magazine of Higher Learning* 51, no. 5: 18–27.

Davies, M., A. Barnett, and T. van Gelder. 2021. "Using Computer-Assisted Argument Mapping to Teach Reasoning to Students." In *Studies in Critical Thinking*, edited by J. A. Blair, vol. 8, 131–176. Windsor Studies in Argumentation.

*Diehl, C. 1995. "Multiple Representations for Improving Scientific Thinking." ERIC. https://eric.ed.gov/?id=ED392842.

*Dowell, J., and T. Gladisch. 2007. "Design of Argument Diagramming for Case-Based Group Learning." In *Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!*, 99–105. ACM.

*Dowell, J., M. Tscholl, T. Gladisch, and M. Asgari-Targhi. 2009. "Argumentation Scheme and Shared Online Diagramming in Case-Based Collaborative Learning." In *Designing for CSCL Practices: Proceedings of the 8th International Conference on Computer-Supported Collaborative Learning (CSCL) 2009*, 567–575. International Society of the Learning Sciences.

*Dwyer, C. P., M. J. Hogan, and I. Stewart. 2010. "The Evaluation of Argument Mapping as a Learning Tool: Comparing the Effects of Map Reading Versus Text Reading on Comprehension and Recall of Arguments." *Thinking Skills and Creativity* 5, no. 1: 16–22.

*Dwyer, C. P., M. J. Hogan, and I. Stewart. 2012. "An Evaluation of Argument Mapping as a Method of Enhancing Critical Thinking Performance in e-Learning Environments." *Metacognition and Learning* 7: 219–244.

*Dwyer, C. P., M. J. Hogan, and I. Stewart. 2013. "An Examination of the Effects of Argument Mapping on Students' Memory and Comprehension Performance." *Thinking Skills and Creativity* 8: 11–24.

*Dwyer, C. P., M. J. Hogan, and I. Stewart. 2015. "The Effects of Argument Mapping-Infused Critical Thinking Instruction on

Reflective Judgement Performance." *Thinking Skills and Creativity* 16: 11–26.

*Eftekhari, M., and E. Sotoudehnama. 2018. "Effectiveness of Computer-Assisted Argument Mapping for Comprehension, Recall, and Retention." *ReCALL* 30, no. 3: 337–354.

*Eftekhari, M., E. Sotoudehnama, and S. S. Marandi. 2016. "Computer-Aided Argument Mapping in an EFL Setting: Does Technology Precede Traditional Paper and Pencil Approach in Developing Critical Thinking?" *Educational Technology Research and Development* 64: 339–357.

Facione, P. A., and N. C. Facione. 2014. *The Holistic Critical Thinking Scoring Rubric—HCTSR*. California Academic Press/Insight Assessment.

*Ferrero, F., A. Gewerc, and D. Letzen. 2018. "Activity System Transformation With ICT: The Case of Argument Diagramming in Logic." In *2018 XIII Latin American Conference on Learning Technologies (LACLO)*, 248–255. IEEE.

*Fysaraki, M., F. Fischer, H. Hußmann, and K. Stegmann. 2016. "Team Awareness Support for Collaborative Argumentation in Higher Education: A Qualitative Multiple-Case Study." In *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016*, edited by C. K. Looi, C. K. Polman, J. L. Cress, and P. Reimann, vol. 2, 1227–1228. International Society of the Learning Sciences.

*Fysaraki, M., and H. Hußmann. 2016. "Towards Tool Support for Team Awareness in Collaborative Argumentation." In *International Conference on Computer Supported Education*, vol. 2, 507–512. Scitepress.

*Gao, L., X. Li, Y. Li, and W. Hu. 2023. "Capturing Temporal and Sequential Patterns of Socio-Emotional Interaction in High-and Low-Performing Collaborative Argumentation Groups." *Asia-Pacific Education Researcher* 32, no. 6: 817–831.

*Garcia Romano, L., M. Occelli, and A. Adúriz-Bravo. 2021. "School Scientific Argumentation Enriched by Digital Technologies: Results With Pre-and in-Service Science Teachers." *Eurasia Journal of Mathematics, Science and Technology Education* 17, no. 7: 1–16.

*Gargouri, C., and M. K. Naatus. 2017. "An Experiment in Mind-Mapping and Argument-Mapping: Tools for Assessing Outcomes in the Business Curriculum." *E-Journal of Business Education & Scholarship of Teaching* 11, no. 2: 39–78.

*Green, N. L. 2013. "Towards Automated Analysis of Student Arguments." In *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9–13, 2013. Proceedings*, edited by H. C. Lane, K. Yacef, J. Mostow, and P. Pavlik, vol. 16, 591–594. Springer.

*Green, N. L. 2017. "Argumentation Scheme-Based Argument Generation to Support Feedback in Educational Argument Modeling Systems." *International Journal of Artificial Intelligence in Education* 27: 515–533.

*Green, N. L., and L. J. Crotts. 2020. "Argument Schemes for AI Ethics Education." In *Proceedings of CMNA'20*, edited by F. Grasso, N. Green, J. Schneider, and S. Wells, 41–50. IOS Press.

*Green, N. L., M. Hinshaw, C. Martensen, M. Narasimhan, and T. Tobgay. 2015. "GAIL: A Genetics Argumentation Inquiry Learning System." In *Proceedings of the 26th Modern Artificial Intelligence and Cognitive Science Conference (MAICS)*, edited by M. Glass and J. H. Kim, 97–101. Curran Associates.

Guzzetti, B. J. 2000. "Learning Counter-Intuitive Science Concepts: What Have We Learned From Over a Decade of Research?" *Reading and Writing Quarterly* 16: 89–98.

*Harrell, M. 2005. "Using Argument Diagramming Software in the Classroom." *Teaching Philosophy* 28, no. 2: 163–177.

*Harrell, M. 2007. "Using Argument Diagramming Software to Teach Critical Thinking Skills." In *Proceedings of IMSCI '07: International Multi-Conference on Society, Cybernetics and Informatics*, vol. 1, 315–320. International Institute of Informatics and Systemics.

*Harrell, M. 2008. "No Computer Program Required: Even Pencil-And-Paper Argument Mapping Improves Critical Thinking Skills." *Teaching Philosophy* 31, no. 4: 351–374.

*Harrell, M. 2011. "Argument Diagramming and Critical Thinking in Introductory Philosophy." *Higher Education Research and Development* 30, no. 3: 371–385.

*Harrell, M., and D. Wetzel. 2013. "Improving First-Year Writing Using Argument Diagramming." *Proceedings of the Annual Meeting of the Cognitive Science Society* 35, no. 35: 2488–2493.

*Hitchcock, D. 2004. "The Effectiveness of Computer-Assisted Instruction in Critical Thinking." *Informal Logic* 24: 183–221.

Hoadley, C., and F. C. Campos. 2022. "Design-Based Research: What It Is and Why It Matters to Studying Online Learning." *Educational Psychologist* 57, no. 3: 207–220. https://doi.org/10.1080/00461520.2022.2079128.

*Hoffmann, M., and J. Borenstein. 2014. "Understanding Ill-Structured Engineering Ethics Problems Through a Collaborative Learning and Argument Visualization Approach." *Science and Engineering Ethics* 20: 261–276.

*Hoffmann, M. H., and R. Catrambone. 2023. "Bad Arguments and Objectively Bad Arguments." *Informal Logic* 43, no. 1: 23–90.

*Hoffmann, M. H., and J. A. Lingle. 2015. "Facilitating Problem-Based Learning by Means of Collaborative Argument Visualization Software." *Teaching Philosophy* 38, no. 4: 371–398.

*Hogan, M., O. Harney, and B. Broome. 2014. "Integrating Argument Mapping With Systems Thinking Tools: Advancing Applied Systems Science." In *Knowledge Cartography: Software Tools and Mapping Techniques*, edited by A. Okada, S. J. Buckingham Shum, and T. Sherborne, 401–421. Springer.

*Hogan, M. J., C. P. Dwyer, O. M. Harney, C. Noone, and R. J. Conway. 2015. "Metacognitive Skill Development and Applied Systems Science: A Framework of Metacognitive Skills, Self-Regulatory Functions, and Real-World Applications." In *Metacognition: Fundamentals, Applications, and Trends*, edited by A. Peña-Ayala, 75–106. Springer.

*Hsi, S., and C. M. Hoadley. 1994. "An Interactive Multimedia Kiosk as a Tool for Collaborative Discourse, Reflection, and Assessment." American Educational Research Association. ERIC ED375814.

*Iandoli, L., I. Quinto, A. De Liddo, and S. B. Shum. 2014. "Socially Augmented Argumentation Tools: Rationale, Design and Evaluation of a Debate Dashboard." *International Journal of Human-Computer Studies* 72, no. 3: 298–319.

*Jeong, A., and H. Y. Kim. 2022. "Identifying Critical Thinking Skills Used by Experts Versus Novices to Construct Argument Maps in a Computer-Aided Mapping Tool." *Knowledge Management & E-Learning* 14, no. 2: 125–149.

*Jeong, A., and H. Seok-Shin. 2023. "Mining, Analyzing, and Modeling the Cognitive Strategies Students Use to Construct Higher Quality Causal Maps." Paper Presented at the IADIS International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA), Madeira Island, Portugal. International Association for Development of the Information Society.

*Jeong, A. C. 2020. "Developing Computer-Aided Diagramming Tools to Mine, Model and Support Students' Reasoning Processes." *Educational Technology Research and Development* 68, no. 6: 3353–3369.

Johnson, D. W., R. T. Johnson, and K. A. Smith. 2014. "Cooperative Learning: Improving University Instruction by Basing Practice on

Validated Theory." *Journal on Excellence in College Teaching* 25, no. 3–4: 85–118.

Jonassen, D. H., and B. Kim. 2010. "Arguing to Learn and Learning to Argue: Design Justifications and Guidelines." *Educational Technology Research and Development* 58: 439–457. https://doi.org/10.1007/s11423-009-9143-8.

Jumariati, J., E. Febriyanti, and M. Rizki. 2021. "Argumentation Skills: An Analysis on EFL Students' Essays Based on Toulmin's Model of Argument." In *Proceedings of the 2nd International Conference on Education, Language, Literature, and Arts*, 86–95. Atlantis Press.

*JURIX Corporate Author. 2012. "Comparing Argument Diagrams." In *Legal Knowledge and Information Systems: JURIX 2012: The Twenty-Fifth Annual Conference*, vol. 250, 81–90. IOS Press. https://doi.org/10.3233/978-1-61499-167-0-81.

*Kaeppel, K. 2021. "The Influence of Collaborative Argument Mapping on College Students' Critical Thinking About Contentious Arguments." *Thinking Skills and Creativity* 40: 100809.

Kalyuga, S. 2021. "The Expertise Reversal Principle in Multimedia Learning." In *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer and L. Fiorella, 171–182. Cambridge University Press. https://doi.org/10.1017/9781108894333.017.

*Katsura, Y., K. Nishina, S. Okada, and K. Nitta. 2016. "Argumentation Support Tool With Modularization Function and Its Evaluation." In *Advances in Social Computing and Digital Education: 7th International Workshop on Collaborative Agents Research and Development, CARE 2016, Singapore, May 9, 2016 and Second International Workshop on Social Computing in Digital Education, SocialEdu 2016, Zagreb, Croatia, June 6, 2016, Revised Selected Papers*, vol. 7, 117–135. Springer International Publishing.

*Kaya, S. 2024. "One Intervention, Several Benefits: Deliberate Computer-Assisted Argument Mapping Practices in an Online Teacher Education Course." *Journal of Computer Assisted Learning* 40: 1–19.

*Khartabil, D., C. Collins, S. Wells, B. Bach, and J. Kennedy. 2021. "Design and Evaluation of Visualization Techniques to Facilitate Argument Exploration." *Computer Graphics Forum* 40, no. 6: 447–465.

Kimmerle, J., F. Fischer, and U. Cress. 2021. "Argumentation and Knowledge Construction." In *International Handbook of Computer-Supported Collaborative Learning*, edited by U. Cress, J. Oshima, C. Rosé, and A. Wise, 183–198. Springer International Publishing.

Kleemola, K., H. Hyytinen, and A. Toom. 2022. "The Challenge of Position-Taking in Novice Higher Education Students' Argumentative Writing." *Frontiers in Education* 7: 885987. https://doi.org/10.3389/feduc.2022.885987.

*Klekovkina, V., and L. Denié-Higney. 2022. "Machine Translation: Friend or Foe in the Language Classroom?" *L2 Journal* 14, no. 1: 105–135.

*Kubosawa, S., K. Nishina, M. Sugimoto, S. Okada, and K. Nitta. 2013. "A Discussion Training Support System and Its Evaluation." In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, 197–201. Association for Computing Machinery.

*Kunsch, D. W., K. Schnarr, and R. van Tyle. 2014. "The Use of Argument Mapping to Enhance Critical Thinking Skills in Business Education." *Journal of Education for Business* 89, no. 8: 403–410.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33, no. 1: 159–174. https://doi.org/10.2307/2529310.

*Le, N. T., S. Niebuhr, D. Drexler, and N. Pinkwart. 2013. "Scaffolding Computer-Supported Argumentation Processes Through Mini-Map Based Interaction Techniques." In *Design, User Experience, and Usability. Web, Mobile, and Product Design: Second International Conference, DUXU 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21–26, 2013, Proceedings, Part IV 2*, 533–542. Springer.

*Li, X. 2023. "Exploring the Relationship Between Individual Learners' Physiological Arousal and Learning Outcomes in Collaborative Learning." In *IEEE International Conference on Advanced Learning Technologies (ICALT)*, 139–143. https://doi.org/10.1109/ICALT58122.2023.00047. IEEE.

*Liu, Q., and J. C. Nesbit. 2018. "Conceptual Change With Refutational Maps." *International Journal of Science Education* 40, no. 16: 1980–1998.

Liu, Q., and J. C. Nesbit. 2024. "The Relation Between Need for Cognition and Academic Achievement: A Meta-Analysis." *Review of Educational Research* 94, no. 2: 155–192. https://doi.org/10.3102/00346543231160474.

*Liu, Q., Z. Zhong, and J. C. Nesbit. 2024. "Argument Mapping as a Pre-Writing Activity: Does It Promote Writing Skills of EFL Learners?" *Education and Information Technologies* 29, no. 7: 7895–7925.

*Loll, F., and N. Pinkwart. 2013. "LASAD: Flexible Representations for Computer-Based Collaborative Argumentation." *International Journal of Human-Computer Studies* 71, no. 1: 91–109.

*Loureiro, M. J., F. N. de Souza, A. Bezerra, and A. Rodrigues. 2014. "Collaboration, Knowledge Sharing, and Digital Environments: What About Argumentation and Questioning Skills?" In *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration: First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part II*, 440–449. Springer International Publishing.

Lovelace, M., and P. Brickman. 2013. "Best Practices for Measuring Students' Attitudes Toward Learning Science." *CBE Life Sciences Education* 12: 507–617.

*Lynch, C., K. Ashley, N. Pinkwart, and V. Aleven. 2009b. "Toward Assessing Law Students' Argument Diagrams." In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, 222–223. Association for Computing Machinery.

*Lynch, C., and K. D. Ashley. 2012. "Modeling Student Arguments in Research Reports." In *Proceedings of the 4th AHFE Conference*, 191–201. AHFE International.

*Lynch, C., K. D. Ashley, N. Pinkwart, and V. Aleven. 2009a. "Argument Diagramming and Diagnostic Reliability. Frontiers in Artificial Intelligence and Applications." In *Legal Knowledge and Information Systems*, vol. 205, 106–115. IOS Press.

*Lynch, C. F., K. D. Ashley, and M. Chi. 2014. "Can Diagrams Predict Essay Grades?" In *Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, June 5–9, 2014. Proceedings 12*, 260–265. Springer International Publishing.

*Ma, Z. Q., L. Y. Kong, Y. F. Tu, G. J. Hwang, and Z. Y. Lyu. 2024. "Strengthening Collaborative Argumentation With Interactive Guidance: A Dialogic Peer Feedback Approach Based on the Six Thinking Hats Strategy." *Interactive Learning Environments* 33, no. 1: 296–320.

*Macagno, F., G. Rowe, C. Reed, and D. Walton. 2006. "Araucaria as a Tool for Diagramming Arguments in Teaching and Studying Philosophy." *Teaching Philosophy* 29, no. 2: 111–124.

*Maftoon, P., P. Birjandi, and P. Pahlavani. 2014. "The Impact of Using Computer-Aided Argument Mapping (CAAM) on the Improvement of Writing Achievement of Iranian Learners of English." *Theory and Practice in Language Studies* 4, no. 5: 982–988.

*Mahmood, S. 2016. "Effectiveness of an Instructional Intervention in Developing Critical Thinking Skills: Role of Argument Mapping in Facilitating Learning of Critical Thinking Skills." In *Proceedings of the 8th International Conference on Computer Supported Education*, 330–336. Springer.

Marsden, E., and C. J. Torgerson. 2012. "Single Group, Pre- and Post-Test Research Designs: Some Methodological Concerns." *Oxford Review of Education* 38, no. 5: 583–616. https://doi.org/10.1080/03054985.2012.731208.

Mayer, R. E. 2020. "Pre-Training Principle." In *Multimedia Learning*, 265–280. Cambridge University Press.

*McGinn, J., R. Picking, L. Picking, and V. Grout. 2007. "The Use of Interactive Visual Metaphors to Enhance Group Discussions Using Mobile Devices." In *Universal Access in Human-Computer Interaction: Ambient Interaction: 4th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2007*, Held as Part of HCI International 2007, Beijing, China, July 22–27, 2007, Proceedings, Part II, 952–960. Springer.

McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22, no. 3: 276–282. https://doi.org/10.11613/BM.2012.031.

*Memiş, E. K., E. Sönmez, and B. N. Ç. Akkaş. 2023. "Computer-Aided Argument Mapping: Effect of Individual and Collaborative Practices on Reasoning Skills." *Malaysian Online Journal of Educational Sciences* 11, no. 1: 25–33.

*Morgan, A. 2006. "Argumentation, Geography Education and ICT." *Geography* 91, no. 2: 126–140.

*Mubarok, H., C. J. Lin, and G. J. Hwang. 2023. "A Virtual Reality-Based Collaborative Argument Mapping Approach in the EFL Classroom." *Interactive Learning Environments* 32, no. 9: 1–4996.

*Muller Mirza, N., V. Tartas, A. N. Perret-Clermont, and J. F. de Pietro. 2007. "Using Graphical Tools in a Phased Activity for Enhancing Dialogical Skills: An Example With Digalo." *International Journal of Computer-Supported Collaborative Learning* 2: 247–272.

Myung, S. K. 2023. "How to Review and Assess a Systematic Review and Meta-Analysis Article: A Methodological Study (Secondary Publication)." *Journal of Educational Evaluation for Health Professions* 20, no. 24. https://doi.org/10.3352/jeehp.2023.20.24.

*Nesbit, J., Q. Liu, J. Sharp, et al. 2024. "Argument Visualization With DMaps: Cases From Postsecondary Learning." *Journal of Interactive Learning Research* 35, no. 2: 223–253.

*Ngajie, B. N., Y. Li, D. T. Tiruneh, and M. Cheng. 2020. "Investigating the Effects of a Systematic and Model-Based Design of Computer-Supported Argument Visualization on Critical Thinking." *Thinking Skills and Creativity* 38: 100742.

Onrubia, J., and A. Engel. 2009. "Strategies for Collaborative Writing and Phases of Knowledge Construction in CSCL Environments." *Computers & Education* 53, no. 4: 1256–1265. https://doi.org/10.1016/j.compedu.2009.06.008.

*Ouyang, F., L. Zhang, M. Wu, and P. Jiao. 2024. "Empowering Collaborative Knowledge Construction Through the Implementation of a Collaborative Argument Map Tool." *Internet and Higher Education* 62: 100946.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *Systematic Reviews* 10, no. 1: 89. https://doi.org/10.1186/s13643-021-01626-4.

Pakdaman-Savoji, A., J. Nesbit, and N. Gajdamaschko. 2019. "The Conceptualisation of Cognitive Tools in Learning and Technology: A Review." *Australasian Journal of Educational Technology* 35, no. 2: 1–24. https://doi.org/10.14742/ajet.4704.

*Pinho, A., L. Á. Pereira, A. Moreira, and M. J. Loureiro. 2009. "Reading, Annotation and Production of an Argumentative Text Written by Higher Education Students in a Blended-Learning Situation." *L1—Educational Studies in Language and Literature* 09: 179–205.

*Pinkwart, N., V. Aleven, K. Ashley, and C. Lynch. 2006. "Toward Legal Argument Instruction With Graph Grammars and Collaborative Filtering Techniques." In *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26–30, 2006. Proceedings 8*, 227–236. Springer.

*Pinkwart, N., K. Ashley, C. Lynch, and V. Aleven. 2009. "Evaluating an Intelligent Tutoring System for Making Legal Arguments With Hypotheticals." *International Journal of Artificial Intelligence in Education* 19, no. 4: 401–424.

*Pinkwart, N., C. Lynch, K. Ashley, and V. Aleven. 2008. "Re-Evaluating LARGO in the Classroom: Are Diagrams Better Than Text for Teaching Argumentation Skills?" In *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23–27, 2008. Proceedings 9*, 90–100. Springer.

Pintrich, P. R., D. A. Smith, T. Garcia, and W. J. McKeachie. 1993. "Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ)." *Educational and Psychological Measurement* 53, no. 3: 801–813. https://doi.org/10.1177/0013164493053003024.

*Piwek, P. 2013. "Supporting Computing and Technology Distance Learning Students With Developing Argumentation Skills." In *2013 IEEE Global Engineering Education Conference (EDUCON)*, 258–267. IEEE.

*Rapanta, C., and D. Walton. 2016a. "Identifying Paralogisms in Two Ethnically Different Contexts at University Level/Identificación de Paralogismos en dos Contextos Universitarios Diferenciados Étnicamente." *Journal for the Study of Education and Development* 39, no. 1: 119–149. https://doi.org/10.1080/02103702.2015.1111610.

*Rapanta, C., and D. Walton. 2016b. "The Use of Argument Maps as an Assessment Tool in Higher Education." *International Journal of Educational Research* 79: 211–221.

Rider, Y., and N. Thomason. 2014. "Cognitive and Pedagogical Benefits of Argument Mapping: LAMP Guides the Way to Better Thinking." In *Knowledge Cartography*, edited by A. Okada, S. Buckingham Shum, and T. Sherborne, 113–134. Springer.

*Robillos, R. J. 2021. "Learners' Writing Skill and Self-Regulation of Learning Awareness Using Computer-Assisted Argument Mapping (CAAM)." *Teaching English With Technology* 21, no. 4: 76–93.

*Robillos, R. J., and S. Art-in. 2023. "Argument Mapping With Translanguaging Pedagogy: A Panacea for EFL Students' Challenges in Writing Argumentative Essays." *International Journal of Instruction* 16, no. 4: 651–672.

*Robillos, R. J., and J. Thongpai. 2022. "Computer-Aided Argument Mapping Within Metacognitive Approach: Its Impact on Students' Argumentative Writing Performance and Selfregulated Learning." *LEARN Journal: Language Education and Acquisition Research Network* 15, no. 2: 160–186.

*Rocke, S., C. A. Radix, J. Persad, and D. Ringis. 2014. "Use of Argument Maps to Promote Critical Thinking in Engineering Education." In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 1–4. IEEE.

*Rousseau, D. L., and T. van Gelder. 2024. "Teaching Critical Thinking With Argument Mapping." *Journal of Political Science Education* 21: 1–17.

Scheuer, O., F. Loll, N. Pinkwart, and B. M. McLaren. 2010. "Computer-Supported Argumentation: A Review of the State of the Art." *International Journal of Computer-Supported Collaborative Learning* 5, no. 1: 43–102. https://doi.org/10.1007/s11412-009-9080-x.

*Scheuer, O., and B. M. McLaren. 2013. "CASE: A Configurable Argumentation Support Engine." *IEEE Transactions on Learning Technologies* 6, no. 2: 144–157.

*Scheuer, O., B. M. McLaren, A. Weinberger, and S. Niebuhr. 2014. "Promoting Critical, Elaborative Discussions Through a Collaboration Script and Argument Diagrams." *Instructional Science* 42: 127–157.

*Sönmez, E., B. N. Ç. Akkas, and E. K. Memis. 2020. "Computer-Aided Argument Mapping for Improving Critical Thinking: Think Better! Discuss Better! Write Better!" *International Journal of Contemporary Educational Research* 7, no. 2: 291–306.

*Su, F., and D. Zou. 2022. "Applying Shimo-Supported Concept Maps, Mind Maps, and Argument Maps to Assist Students' Argumentative Writing." In *2022 9th International Conference on Behavioural and Social Computing (BESC)*, 1–5. IEEE.

Suthers, D. D. 2003. "Representational Guidance for Collaborative Inquiry." In *Arguing to Learn: Computer-Supported Collaborative Learning*, edited by J. Andriessen, M. Baker, and D. Suthers, vol. 1, 1–26. Springer. https://doi.org/10.1007/978-94-017-0781-7_2.

Swanson, E., L. V. McCulley, D. J. Osman, N. Scammacca Lewis, and M. Solis. 2019. "The Effect of Team-Based Learning on Content Knowledge: A Meta-Analysis." *Active Learning in Higher Education* 20, no. 1: 39–50. https://doi.org/10.1177/1469787417731201.

*ter Berg, T., E. van der Brugge, and S. Cullen. 2013. "Teaching Critical Thinking With Rationale." In *INTED2013 Proceedings*, 5145–5150. IATED.

Toulmin, S. E. 2003. *The Uses of Argument*. Updated ed. Cambridge University Press.

*Tsovaltzi, D., A. Weinberger, O. Scheuer, T. Dragon, and B. M. McLaren. 2013. "Collaborative Learning in Facebook: Can Argument Structure Facilitate Academic Opinion Change?" In *To See the World and a Grain of Sand: Learning Across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings*, edited by N. Rummel, M. Kapur, M. Nathan, and S. Puntambekar, vol. 2, 177–180. International Society of the Learning Sciences.

*Uçar, B., and Y. Demiraslan Çevik. 2021. "The Effect of Argument Mapping Supported With Peer Feedback on Pre-Service Teachers' Argumentation Skills." *Journal of Digital Learning in Teacher Education* 37, no. 1: 6–29.

Valentine, J. C., T. D. Pigott, and H. R. Rothstein. 2010. "How Many Studies Do You Need? A Primer on Statistical Power for Meta-Analysis." *Journal of Educational and Behavioral Statistics* 35, no. 2: 215–247. https://doi.org/10.3102/1076998609346961.

Vallerand, R. J., L. G. Pelletier, M. R. Blais, N. M. Brière, C. Senecal, and E. F. Vallieres. 1992. "The Academic Motivation Scale: A Measure of Intrinsic, Extrinsic, and Amotivation in Education." *Educational and Psychological Measurement* 52, no. 4: 1003–1017. https://doi.org/10.1177/0013164492052004025.

*van Amelsvoort, M., and A. Maes. 2016. "Show Me Your Opinion. Perceptual Cues in Creating and Reading Argument Diagrams." *Instructional Science* 44: 335–357.

*van Gelder, T. 2005. "Teaching Critical Thinking: Some Lessons From Cognitive Science." *College Teaching* 53, no. 1: 41–48.

Van Gelder, T. 2007. "The Rationale for Rationale." *Law, Probability, and Risk* 6, no. 1–4: 23–42. https://doi.org/10.1093/lpr/mgm032.

van Gelder, T. 2015. "Using Argument Mapping to Improve Critical Thinking Skills." In *The Palgrave Handbook of Critical Thinking in Higher Education*, edited by M. Davies and R. Barnett. Palgrave Macmillan. https://doi.org/10.1057/9781137378057_12.

Wolfe, C. R. 2011. "Argumentation Across the Curriculum." *Written Communication* 28, no. 2: 193–219. https://doi.org/10.1177/0741088311399236.

*Wu, B., M. Wang, J. M. Johnson, and T. A. Grotzer. 2014. "Improving the Learning of Clinical Reasoning Through Computer-Based Cognitive Representation." *Medical Education Online* 19, no. 1: 25940. https://doi.org/10.3402/meo.v19.25940.

*Wu, B., M. Wang, J. M. Spector, and S. J. Yang. 2013. "Design of a Dual-Mapping Learning Approach for Problem Solving and Knowledge Construction in Ill-Structured Domains." *Journal of Educational Technology & Society* 16, no. 4: 71–84.

*Yilmaz-Na, E., and E. Sönmez. 2023a. "Having Qualified Arguments: Promoting Pre-Service Teachers' Critical Thinking Through Deliberate Computer-Assisted Argument Mapping Practices." *Thinking Skills and Creativity* 47: 101216.

*Yilmaz-Na, E., and E. Sönmez. 2023b. "Unfolding the Potential of Computer-Assisted Argument Mapping Practices for Promoting Self-Regulation of Learning and Problem-Solving Skills of Pre-Service Teachers and Their Relationship." *Computers & Education* 193: 104683.

*Yıldızlı, H., and İ. Şimşek. 2020. "The Effects of Software-Aided Mind and Argument Mapping on Learning in Higher Education." *International Journal of Contemporary Educational Research* 7, no. 2: 187–201.

Zheng, L., M. Long, J. Niu, and L. Zhong. 2023. "An Automated Group Learning Engagement Analysis and Feedback Approach to Promoting Collaborative Knowledge Building, Group Performance, and Socially Shared Regulation in CSCL." *International Journal of Computer-Supported Collaborative Learning* 18, no. 1: 101–133. https://doi.org/10.1007/s11412-023-09386-0.

*Zumbach, J., M. Ramsauer, N. H. Schwartz, and S. C. Koch. 2008. "Fostering Hypermedia Learning With Different Argumentation Tools: The Role of Argument Visualisation." In *Proceedings of the International Conference on Computer-Supported Collaborative Learning*, 267–274. Springer.

Appendix

Codes Used to Categorise the Studies

Code	Definition
Country	In what country was the research conducted? If not an empirical study, in which country is the first author's institutional affiliation?
Mapping tool	If reported, what is the name of the argument mapping tool(s) used by participants? Do not identify generic tools such as PowerPoint but do identify any node-link mapping tools such as Cmap or Mind Map
Empirical [0, 1]	Code 0 for literature reviews or other non-empirical work Code 1 empirical studies
Tech [0, 1]	Did participants use a digital application to construct maps? Code 0 if non-empirical, participants did not construct maps, or constructed maps on paper, etc. Code 1 if participants constructed maps using technology such as argument mapping tools, node-link tools, or any graphical software
Methodology [0, 1, 2, 3]	What research method was used? Code 0 if not empirical Code 1 if quantitative Code 2 if qualitative Code 3 if mixed quant/qual
Design [0, 1, 2, 3]	What type of quantitative research design was used? If a combination of designs was used, code the highest number Code 0 if qualitative or N/A Code 1 if single group intervention or observational design Code 2 if quasi-experimental between-subjects intervention Code 3 if randomised trial intervention
Attitude [0, 1]	Were participant attitudes toward argument mapping investigated? (Include quantitative and qualitative analysis) Code 0 if no Code 1 if yes
Setting [0, 1, 2, 3]	Where did the argument mapping learning activity take place? Code 0 if N/A or not reported Code 1 if in class or instructional lab Code 2 if for homework Code 3 if in research lab
Eval [0, 1]	Were participants' argument maps evaluated to produce research data? (Include quantitative and qualitative data) Code 0 if no or N/A Code 1 if yes
Assess [0, 1]	Was it a major goal of the research to investigate argument mapping as a way of assessing learning? For example, was it correlated with posttest data to assess its validity as an assessment tool Code 0 if no or N/A Code 1 if yes
Collab [0, 1]	Did any of the participants engage in collaborative argument mapping? This would not include collaboration before or after the actual AM phases Code 0 if no or N/A Code 1 if yes
The following codes identified independent and dependent variables investigated in experiments or quasi-experiments that compared outcomes of groups that received different instructional treatments. Reviews, qualitative studies, and observational quantitative studies that did not compare treatment groups were all coded as 0. The prefix IV is used for independent variables. The prefix DV is used for dependent variables	
IV ArgMap [0, 1]	Was argument mapping (including studying preconstructed maps) treated as an independent variable? Code 0 if no Code 1 if yes
IV learning context [0, 1]	Was the designed learning context for argument mapping treated as an independent variable (e.g., individual vs. collaborative argument mapping)? Code 0 if no Code 1 if yes

Code	Definition
IV ArgMap type [0, 1]	Did the study compare the effects of different argument map types or visualisation tools (e.g., Rationale versus LASAD or argument map versus concept map)? Code 0 if no Code 1 if yes
DV thinking skills [0, 1]	Did the study treat thinking skills (e.g., critical thinking, argumentation, self-regulated learning) as a dependent variable? Code 0 if no Code 1 if yes
DV writing skills [0, 1]	Did the study treat writing skills (including argumentative writing) as a dependent variable? Code 0 if no Code 1 if yes
DV motivation [0, 1]	Did the study treat motivation to learn (e.g., need for cognition, self-efficacy, interest in a curricular domain) as a dependent variable? Code 0 if no. Code 1 if yes.
DV domain concepts [0, 1]	Did the study treat acquisition of domain concepts (e.g., chemistry concepts in a chemistry course) as a dependent variable? Domain concepts are usually assessed via a posttest Code 0 if no Code 1 if yes
DV task performance [0, 1]	Did the study treat task performance during the learning task (e.g., quality of online discussion) as a dependent variable? Code 0 if no Code 1 if yes
DV ArgMap quality [0, 1]	Did the study treat the quality or structure of student argument maps (constructed either within the learning task or in a posttest) as a dependent variable? Code 0 if no Code 1 if yes