GPH-International Journal of Educational Research
(Online ISSN 3050-9599)
Volume 9 | Issue 07 | July 2026
Published by: Global Publication House
DOI: https://doi.org/10.5281/zenodo.22229287
Available online at: www.gphjournal.org
Journal homepage: https://gphjournal.org/index.php/er/index

ARGUMENTATION FROM EVIDENCE: A BIBLIOMETRIC ANALYSIS OF THE KNOWLEDGE STRUCTURE OF AN INTERDISCIPLINARY RESEARCH FIELD

Nguyen Xuan Thanh,
Phu Duc High School, Hung Yen, Vietnam.
Nguyen Thi Thu Ha,
Thai Nguyen University of Education, Nguyen, Vietnam.
Nguyen Anh Thuan,
Hanoi National University of Education, Hanoi, Vietnam.
Corresponding author: Nguyen Xuan Thanh, xuanthanh2282@gmail.com

Abstract

Argumentation from evidence is treated as a core scientific practice in contemporary science education, yet the term argumentation is used with different meanings across several research communities. This creates difficulty in delimiting the literature when a review is undertaken. This study maps the knowledge structure of the literature on argumentation associated with data and evidence, using metadata from Scopus and four science mapping techniques implemented in VOSviewer. A high-sensitivity query on the TITLE-ABS-KEY field returned 14,938 records, with no restriction on year, document type, subject area or language. The keyword co-occurrence network retained 1,000 keywords in five clusters; the co-citation network retained 251 cited references in seven clusters; the co-authorship network retained 13 authors in three clusters; and the document citation network retained 21 documents in six clusters. The four analyses converge on a single observation: the field is a set of communities that share terminology rather than an integrated research area. The science education cluster in the co-citation network has the highest total link strength among the seven clusters and contains a stable set of foundational works published mainly between 2000 and 2010. The argumentation theory cluster is the smallest in size but has the highest mean link strength, indicating that it is cited across clusters. Keywords emerging in the final two years of the dataset are associated with generative artificial intelligence. Two implications are drawn, one for the design of search strategies in review studies and one for research on argumentation in school science.

Keywords:

Scientific argumentation; evidence-based reasoning; bibliometric analysis; co-citation analysis; VOSviewer; science education.

Introduction

Scientific argumentation competency refers to the ability to advance a claim about natural phenomena, to mobilise data and evidence in support of that claim, and to articulate the connection between evidence and claim through scientific principles. Over the past three decades this competency has moved from the position of a secondary objective to that of a core practice in science education. The framework for K-12 science education issued by the National Research Council places engaging in argument from evidence among the eight science and engineering practices that learners should experience at all grade levels (National Research Council, 2012). This positioning reflects an epistemological view: science is not only a body of established conclusions but also a process in which claims are proposed, tested against data and evaluated by a professional community.

From this view, science education research has developed a continuous line of work on the teaching and learning of argumentation in classrooms. Intervention studies indicate that the quality of student argumentation changes when teachers receive training on argument structure and are provided with learning tasks that leave room for scientific disagreement (Osborne et al., 2004). A parallel line of work concerns analytical tools, focusing on describing the components of an argument and establishing quality levels that can be observed in written products or in student talk. Both lines rest on a relatively stable theoretical foundation in which Toulmin's model of argument structure serves as a recurrent reference.

In Vietnam, the requirement for evidence-based argumentation appears in the 2018 General Education Curriculum. The inquiry component of natural science competency in the integrated Natural Science subject, and the corresponding components in Physics, Chemistry and Biology, include requirements concerning the collection of measurements, the processing of data, the drawing of conclusions and the presentation of conclusions together with their grounds (Ministry of Education and Training, 2018). The shift from a content-oriented curriculum to a competency-oriented one raises a need for assessment instruments, for procedures to design learning tasks and for teacher professional development materials connected to argumentation practice. This need is currently met at a level that does not match the pace of curriculum implementation.

Domestic publications related to argumentation are concentrated in the language disciplines and in argumentative writing, where argumentation is examined through rhetoric and critical thinking rather than through empirical evidence. A study by Nguyen and Nguyen (2020) with Vietnamese students of English reports an association between critical thinking and the quality of argumentative essays, and notes that argumentation remains a relatively new concept in the Vietnamese education system with a limited number of studies on the topic. Within science education, the number of Vietnamese-language publications that use an argumentation framework systematically is small, and most do not locate their position within the international research literature.

Locating that position encounters a terminological obstacle. The term argumentation and its cognates are used concurrently across several scholarly communities with different meanings. In generative linguistics, argument denotes a verbal argument in the sense of predicate structure. In computer science, argument mining denotes the automatic extraction of argument structure from text. In political science and law, argument denotes policy or legal reasoning. In science education research, the term denotes an epistemic practice tied to empirical data. This formal overlap with divergent content makes delimiting the literature difficult, particularly for researchers new to the field and for groups with limited access to subscription databases.

Existing reviews of argumentation in science education mostly use systematic review procedures applied to a corpus filtered manually against predefined inclusion criteria. This approach permits close analysis of individual studies but has two limits. First, corpus size is typically in the range of tens to a few hundred documents, which is not sufficient to describe the structure of an entire research field. Second, because inclusion criteria remove work outside science education at the outset, such reviews do not describe the relationship between the science education segment and other scholarly communities that use the same terminology.

Bibliometrics offers a complementary approach to this gap. Rather than reading and coding each document, the method exploits relations already present in bibliographic metadata, including co-occurrence among keywords, co-citation among cited references, co-authorship, and direct citation among documents. Each type of relation affords a different view: thematic structure, intellectual base, collaboration structure and lines of influence (Donthu et al., 2021). Applied to a large corpus not restricted in advance by subject area, this approach permits direct observation of the degree of fragmentation or integration of a research field.

The present study describes the knowledge structure of the scholarly literature on argumentation associated with data and evidence, drawing on the Scopus database and four science mapping techniques. Three research questions are posed: (1) What clusters constitute the thematic structure of this research field, and which topics have emerged in the most recent period? (2) What works constitute the intellectual base of the field, and how are these works distributed across clusters? (3) What degree of integration or fragmentation is reflected in the co-authorship network and the document citation network?

The results are addressed to two groups of readers. For science education researchers, the article provides an inventory of the foundational works of the argumentation segment and describes the position of that segment within the broader picture. For those conducting reviews, the article offers guidance on designing search strategies in a field with polysemous terminology.

Theoretical background

The structure of an argument and the role of data

Toulmin's (2003) model of argument structure is the most frequently invoked reference in research on argumentation. The model separates an argument into six components: the claim is the conclusion advanced; the data are the facts serving as grounds; the warrant is the proposition connecting data to claim; the backing consists of the theoretical assumptions or professional conventions that support the warrant; the qualifier indicates the degree of certainty attached to the claim; and the rebuttal specifies conditions under which the claim does not hold. A notable feature of the model is that it separates data from warrant, thereby showing that having data is not equivalent to having an argument.

In science education research the model is applied as an analytical tool for classroom discourse, commonly referred to as Toulmin's Argument Pattern. Erduran et al. (2004) describe two modes of application: one counts the frequency of each component in discourse, the other ranks argument quality according to the presence of rebuttals. The authors also note two limits. The boundary between data and warrant is difficult to determine in spoken discourse because it depends on contextual information that is not always available in transcripts, and the model describes the formal structure of an argument without evaluating the scientific correctness of its content. Subsequent studies therefore tend to combine structural analysis with an assessment of the scientific content of claims and evidence.

The nature of evidence varies with the context of the learning task, and this variation has consequences for the design of assessment instruments. In tasks framed around socio-scientific issues, evidence usually takes the form of qualitative information, expert opinion or secondary data, and the focus of argumentation lies in weighing competing values. In tasks framed around experimentation, evidence takes the form of measurements, data tables and graphs. In the second case, moving from raw data to evidence requires additional operations: selecting the quantities to be represented, processing the measurements, reading trends from graphs and estimating measurement reliability. These operations belong to data handling competency yet are a condition for the formation of evidence, so an assessment framework for argumentation in experimental contexts needs indicators for them alongside indicators of argument structure.

Argumentation as a practice in science education

The introduction of argumentation into science classrooms was grounded in work describing the gap between professional scientific discourse and classroom discourse. Driver et al. (2000) note that the traditional science classroom operates under norms in which knowledge is presented as a set of completed conclusions, offering few opportunities for students to advance competing claims and defend their own. The authors propose that establishing argumentation norms in the classroom requires simultaneous change along three dimensions: learning tasks must allow for reasonable disagreement, the teacher's role must shift from supplier of answers to facilitator of discussion, and students must receive explicit guidance on the criteria of a quality argument.

Kuhn (1991) provides a psychological foundation for this line of research by describing systematic difficulties in learners' argumentation, among them the tendency to conflate claim with evidence and the tendency to overlook competing explanations. On this basis, applied studies have developed simplified frameworks to support students, the most widespread being the three-component framework of claim, evidence and reasoning. This framework retains the core of the Toulmin model while reducing the number of components to be tracked, which makes it more workable with school students and with criterion-based scoring of written products.

Berland and Reiser (2009) add a distinction concerning purpose. In their account, scientific explanation and scientific argumentation share surface components but differ in function: explanation aims at making a phenomenon intelligible to the learner, whereas argumentation aims at persuading others and defending a claim against rebuttal. This distinction bears on instructional design, since a task that only asks students to explain a phenomenon does not automatically generate a need for argumentation in the absence of an alternative position. Berland and McNeill (2010) develop this line into a learning progression in which argument quality is described along three dimensions, namely the instructional context, the argumentative product and the argumentative process, each comprising several levels from simple to complex.

These three groups of work form the theoretical core that subsequent studies build on. What they share is a view of argumentation as both a learning goal and a learning means: students learn to argue and at the same time learn science content through arguing. What distinguishes them is the unit of analysis, which ranges from the structure of a written product to the dynamics of a group discussion and to the norms of an entire classroom community. This difference explains why studies in the field use different measurement instruments and why comparison of results across studies is not always feasible.

Bibliometrics and four science mapping techniques

Bibliometrics comprises quantitative methods that analyse the metadata of scholarly documents in order to describe the structure and dynamics of a research field. Zupic and Čater (2015) distinguish two families of application: evaluation of the productivity and influence of authors, institutions or journals; and science mapping, which describes relations among units of knowledge. The present study belongs to the second family. The four techniques used here correspond to four types of relation and four levels of meaning.

Keyword co-occurrence analysis rests on the assumption that two keywords appearing in the same document are related in content. The result is a thematic map reflecting the current state of the field. Co-citation analysis rests on the frequency with which two cited references are cited together in a document, thereby identifying the intellectual base on which the research community draws. Co-authorship analysis describes the collaboration structure among researchers. Document citation analysis describes direct lines of influence among works in the dataset. Because the four techniques exploit four different types of relation, their results complement one another and do not necessarily coincide in cluster structure (Donthu et al., 2021).

The software used is VOSviewer. The program constructs maps using the VOS technique, in which the position of each unit is determined so that the distance between two units reflects their relatedness, and clustering is performed by an algorithm based on a modularity measure with an adjustable resolution parameter (van Eck & Waltman, 2010). Cluster results depend on the selection threshold and the resolution parameter, so the cluster structure obtained is one of several technically acceptable partitions rather than a unique one. Naming clusters is an interpretive operation performed by the researcher on the basis of the units with the highest link strength in each cluster.

Method

The study uses a descriptive design, analyses secondary data in the form of bibliographic metadata, and involves no human participants, so no ethical approval was required. Data were retrieved from the Scopus database, selected for its coverage of social science and education journals and for providing the cited reference fields required for co-citation analysis.

The search string was built from two blocks combined with the AND operator and applied to the TITLE-ABS-KEY field. The first block comprised expressions denoting argumentation and scientific reasoning: argument, “scientific reasoning”, “evidence-based reasoning”, “evidentiary reasoning”, “scientific explanation”, “claim evidence reasoning”. The second block comprised expressions denoting data and evidence: data, dataset, graph, measurement, “experimental result”, “empirical data”, evidence. The search was conducted on 20 August 2026 and returned 14,938 records. No restriction was applied on publication year, document type, subject area or language.

The absence of restrictions was a deliberate choice. The aim of the study is to observe the structure of the entire body of literature that uses argumentation terminology in connection with data, including scholarly communities outside science education. Applying a subject area restriction at the query stage would make the third research question unanswerable, since the communities to be compared would already have been removed from the dataset. The consequence is a query with high sensitivity and low specificity, which is addressed at the interpretation stage rather than at the collection stage.

The dataset exported from Scopus was in CSV format and included the fields for authors, title, publication year, source, author keywords, index keywords, abstract and cited references. All 14,938 records were imported into VOSviewer version 1.6.21 to construct four networks.

The keyword co-occurrence network was built on the all-keywords field, combining author keywords and index keywords, using full counting and a minimum occurrence threshold of 10. The number of keywords meeting the threshold was 1,450; of these, the 1,000 keywords with the highest total link strength were retained for analysis and distributed across five clusters. The co-citation network was built on cited references with a minimum citation threshold of 20, yielding 251 cited references in seven clusters. The co-authorship network was built on authors with minimum thresholds of 5 documents and 20 citations per author, yielding 13 authors in three clusters. The document citation network was built on documents with a minimum citation threshold of 0, together with the largest set of connected items option, yielding 20 documents in six clusters. Full parameter settings are reported in Table 1.

Keywords were not normalised with a thesaurus file, and index terms originating from the biomedical indexing system in Scopus were not removed. These two decisions were taken in order to preserve the data structure as it appears in the database, so that the results can be reproduced by others from the same query without depending on preprocessing choices specific to the present authors. The consequences of these decisions for interpretation are addressed in the Limitations section.

The indicators used in the analysis are: occurrences for keywords; citations for documents and cited references; total link strength for all units, which represents the summed weight of a unit's links to the remaining units in the network; normalised citations, defined as the ratio of a document's actual citation count to the mean citation count of documents published in the same year within the dataset, used to compare influence across documents with different citation accumulation periods; and average publication year of the documents containing a given keyword, used to identify emerging topics.

Cluster naming followed three steps: listing the ten units with the highest total link strength in each cluster; two members of the research team proposing cluster names independently on the basis of that list together with the list of units with the highest frequency; and comparing the two sets of names and resolving differences through discussion. Cluster names are interpretive and are presented together with the list of representative units so that readers can compare them against the data.

The threshold for identifying emerging topics was set at an average publication year of 2024.3 or later combined with an occurrence count of 12 or more. The first threshold corresponds to approximately the final two years of the dataset; the second excludes keywords whose frequency is too low to support inference about trends.

Map data were exported from VOSviewer in tabular format, including label, coordinates, cluster identifier, number of links, total link strength and the weights corresponding to each analysis. The descriptive statistics reported in the Results section were computed directly from these export files, which are supplied with the article as an appendix.

Results

Thematic structure of the field

The keyword co-occurrence network retained 1,000 keywords distributed across five clusters of uneven size. Cluster 1 contains 456 keywords with 14,196 total occurrences; cluster 2 contains 251 keywords with 9,040 occurrences; cluster 3 contains 196 keywords with 6,903 occurrences; cluster 4 contains 96 keywords with 3,089 occurrences; and cluster 5 contains a single keyword, grounded theory, with 11 occurrences. This structure indicates that the body of literature is not organised around a single thematic centre but separates into blocks with distinct content (Figure 1).

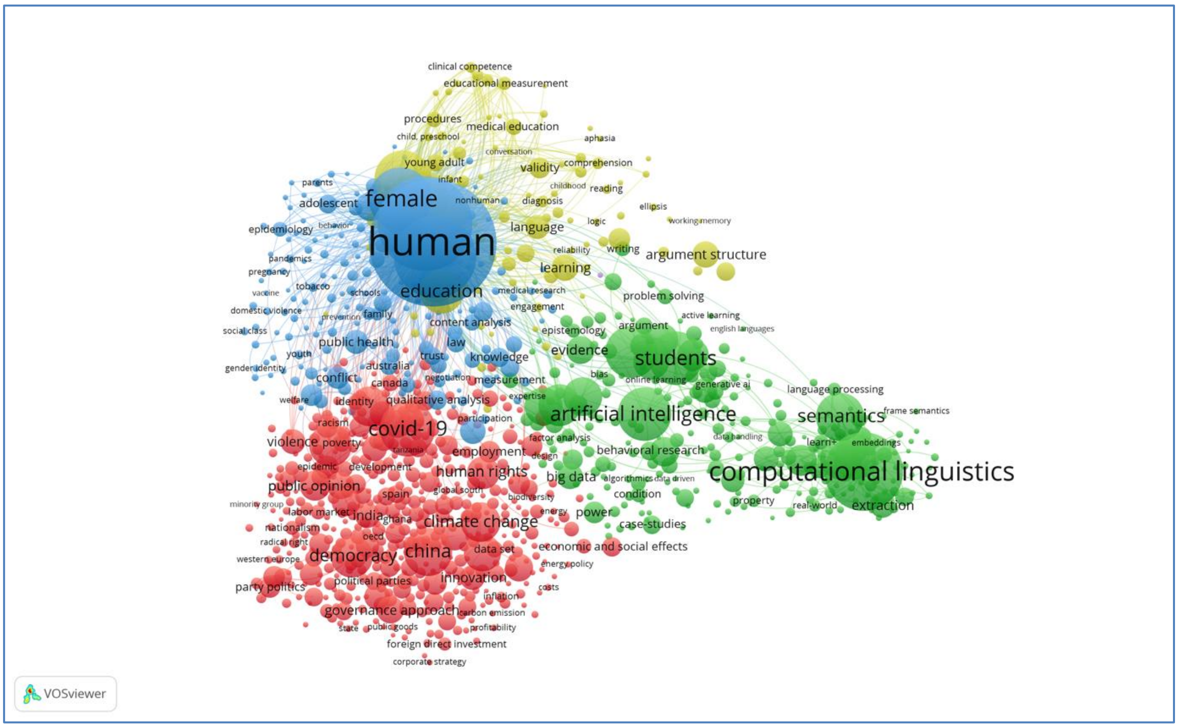

Figure 1. Keyword co-occurrence network of the literature on argumentation associated with data and evidence

Note. Network visualization, VOSviewer; n = 1,000 keywords, 5 clusters. Node size is proportional to occurrence count; distance between nodes reflects co-occurrence strength.

The ten most frequent keywords are human (773 occurrences), humans (521), article (431), computational linguistics (401), female (299), male (262), covid-19 (253), artificial intelligence (241), students (241) and argumentation (234). Six of these ten are index terms originating from the biomedical indexing system in Scopus and carry no conceptual content relating to the research topic. Those six terms also rank highest on total link strength, with human at 7,680 and humans at 5,317, which reflects their assignment to a large share of documents rather than a central thematic position.

Comparing frequency with link strength reveals a different pattern for the keyword argumentation. It ranks tenth on frequency but reaches a total link strength of only 370, considerably lower than computational linguistics (1,999), students (1,279) or natural language processing systems (1,090) despite comparable or lower frequency. This gap indicates that argumentation co-occurs with a more restricted set of partner keywords, that is, it is used in specific contexts rather than distributed evenly across the dataset.

Table 2. Characteristics of the five keyword clusters in the co-occurrence network

Cluster	Keywords	Occurrences	Proposed label	Representative keywords (occurrences)	Mean pub. year
1	456	14,196	Politics, policy and sustainable development	covid-19 (253), china (210), united states (178), democracy (176)	2022.78
2	251	9,040	Computation, artificial intelligence and education	computational linguistics (401), artificial intelligence (241), students (241), argumentation (234)	2023.36
3	196	6,903	Biomedical index terms and qualitative methods	human (773), humans (521), article (431), ethics (151)	2022.96
4	96	3,089	Demographics, linguistics and medical education	male (262), adult (222), education (187), argument structure (93)	2022.93
5	1	11	grounded theory	grounded theory (11)	2023.55

Cluster 2 contains most of the keywords directly related to scientific argumentation in education and has the latest mean publication year among the four main clusters. Two branches can be distinguished within it. The first is technical and comprises computational linguistics (401), natural language processing systems (177), language model (175), machine learning (108) and argument mining (28). The second is educational and comprises students (241), higher education (158), teaching (140), critical thinking (123), scientific reasoning (117), evidence (101), science education (82), argumentative writing (45), teacher education (44), socio-scientific issues (24) and secondary education (21). Link strength is higher in the technical branch: natural language processing systems reaches 1,090 while science education reaches 202 and scientific reasoning reaches 257.

Table 3. Keywords directly related to argumentation in science education

Keyword	Occurrences	Cluster	Total link strength	Mean pub. year	Mean norm. citations
students	241	2	1,279	2023.26	1.20
argumentation	234	2	370	2022.85	0.80
critical thinking	123	2	319	2024.13	1.11
scientific reasoning	117	2	257	2023.44	0.79
evidence	101	2	153	2022.61	0.52
science education	82	2	202	2023.20	1.20
argumentative writing	45	2	53	2023.56	1.11
teacher education	44	2	72	2022.82	0.83
epistemology	41	2	117	2022.88	1.05
socio-scientific issues	24	2	47	2023.17	0.84
secondary education	21	2	52	2023.24	1.21

Keyword	Occurrences	Cluster	Total link strength	Mean pub. year	Mean norm. citations
scientific evidence	20	2	59	2023.45	0.38

The mean publication year score of individual keywords permits identification of topics emerging at the end of the observation period. Applying the threshold stated in the Method section, 29 keywords meet the criteria, of which 25 belong to cluster 2. The leading group by time comprises generative ai (mean year 2025.57), large language models (2025.39), large language model (2025.35), chatgpt (2024.86) and ai (2024.41). Several keywords in this group have mean normalised citations above 1, specifically chatgpt at 3.67, educational technology at 3.66, large language models at 2.65 and ai at 2.22. At the other end of the time scale, internet of things (29 occurrences) has a mean year of 2021.03 and explanation (18 occurrences) a mean year of 2021.56.

Table 4. Ten keywords with the latest mean publication year (occurrences ≥ 12)

Keyword	Occurrences	Cluster	Mean pub. year	Mean norm. citations
generative ai	28	2	2025.57	1.96
large language models	28	2	2025.39	2.65
large language model	26	2	2025.35	1.06
ethical aspects	14	2	2025.14	0.35
human engineering	13	2	2025.00	1.11
chatgpt	28	2	2024.86	3.67
ethical technology	40	2	2024.85	0.29
personnel training	33	2	2024.64	0.60
educational technology	17	2	2024.53	3.66
critical thinking skills	30	2	2024.43	1.08

Across the full set of 1,000 keywords, no term denoting a specific school science subject meets the display threshold. Keywords denoting school-level contexts are present but at low frequency: stem (16 occurrences) and high school (10 occurrences). By comparison, medical education reaches 51 occurrences and engineering education reaches 86, both belonging to professional education at the tertiary level.

Intellectual base of the field

The co-citation network retained 251 cited references distributed across seven clusters (Figure 2). Unlike the keyword co-occurrence network, cluster size and link strength are not proportional here. Cluster 1 is the largest with 84 references but has a mean link strength of only 49.9. Cluster 7 contains only 8 references but has a mean link strength of 384.9, approximately seven times the overall mean. Cluster 2 contains 41 references with a total link strength of 6,174, the highest among the seven clusters, and a mean of 150.6.

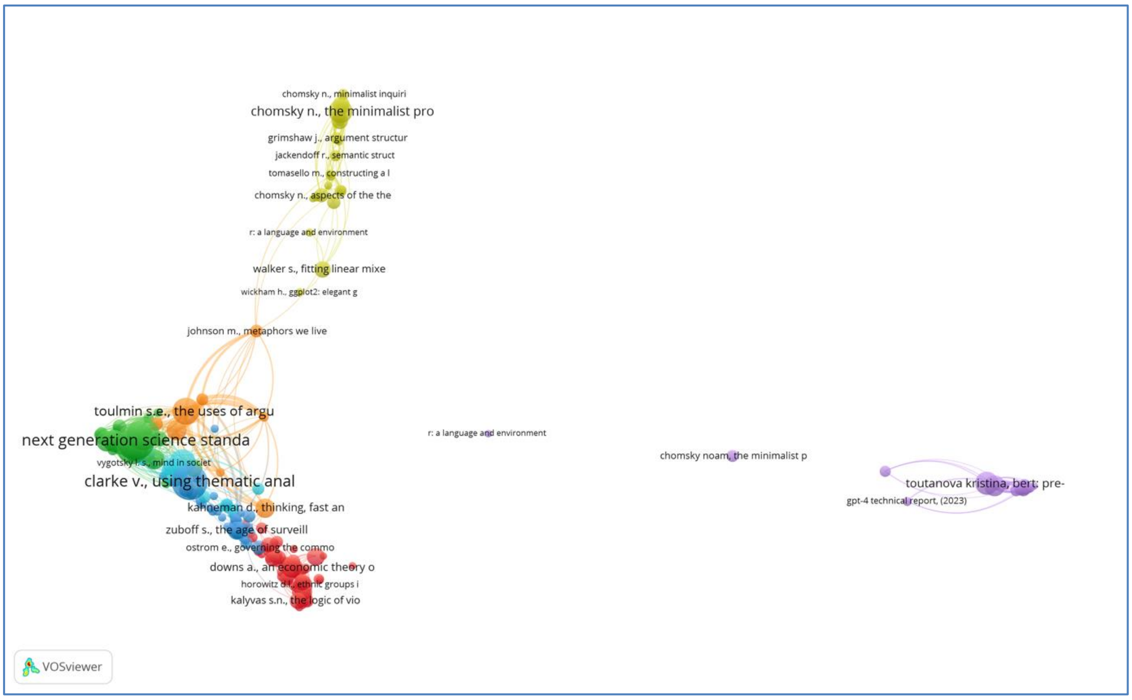

Figure 2. Co-citation network of cited references

Note. Network visualization, VOSviewer; n = 251 cited references, 7 clusters. Node size is proportional to the number of citations received within the dataset.

Table 5. Seven clusters in the co-citation network and representative references

Cluster	Refs	Mean link strength	Proposed label	References with the highest link strength
1	84	49.9	Political science, institutions and conflict	Signorino (2010, 144); Svolik (2012, 130); Frantz (2014, 113)
2	41	150.6	Argumentation in science education	NGSS (2013, 453); Osborne et al. (2004, 394); NRC (2012, 358)
3	40	46.0	Sociology of knowledge and critical data studies	Bourdieu (1977, 192); Braun and Clarke (2006, 108); O'Neil (2016, 89)
4	30	68.0	Generative grammar and argument structure	Chomsky (1995, 199); Chomsky (2001, 139); Kratzer (1996, 122)
5	25	58.2	Argument mining with deep learning	Devlin et al. (2019, 163); Liu et al. (2019, 137); Stab and Gurevych (2017, 109)
6	23	51.8	Measurement, psychometrics and assessment	Cohen (1988, 155); AERA et al. (2014, 144); Kane (2013, 110)
7	8	384.9	Argumentation theory	Toulmin (2003, 942); Perelman and Olbrechts-Tyteca (1969, 753); Walton et al. (2008, 345)

Cluster 2 assembles the foundational works of argumentation in science education. Besides the three leading references, the cluster contains Driver et al. (2000) with a link strength of 346, Toulmin (1958) with 313, Kuhn (1991) with 234, Berland and McNeill (2010) with 209, Berland and Reiser (2009) with 207 and Erduran et al. (2004) with 203. Works on the distinction between explanation and argument, on learning progressions and on research challenges in the field are also present with link strengths between 130 and 181. The composition of the cluster indicates a relatively stable theoretical base formed mainly between 2000 and 2010 that continues to be cited in the current literature.

Cluster 7 comprises works in argumentation theory belonging to the philosophical and rhetorical traditions. The distinctive position of this cluster lies in its small size combined with the large number of links each reference maintains, which indicates that these works are cited from several different thematic clusters. The 2003 edition of Toulmin's work has a link strength of 942, the highest in the network.

One observation deserves attention: Toulmin's work appears as three separate records belonging to three different clusters, owing to variation in how author names and publication years are recorded for the original and the reissued editions. The three records together received 225 citations. A similar pattern occurs for the work on Toulmin's Argument Pattern, which appears as two records differing only in the apostrophe character. This is a direct consequence of not normalising the cited reference data.

The remaining four clusters belong to fields outside science education. Cluster 4 comprises works on generative syntax, in which argument denotes a verbal argument in the sense of predicate structure. Cluster 5 comprises language models and works on the automatic extraction of argument structure. Cluster 1 comprises works on political institutions and armed conflict. Cluster 3 comprises works on social theory, qualitative methods and critical research on data and algorithms. Cluster 6 comprises works on measurement validity and statistics, cited across fields.

Collaboration structure among authors

At the selected threshold, the co-authorship network retained only 13 authors distributed across three clusters (Figure 3). All 13 belong to the research community working on natural language processing and argument mining. No author from science education met the threshold.

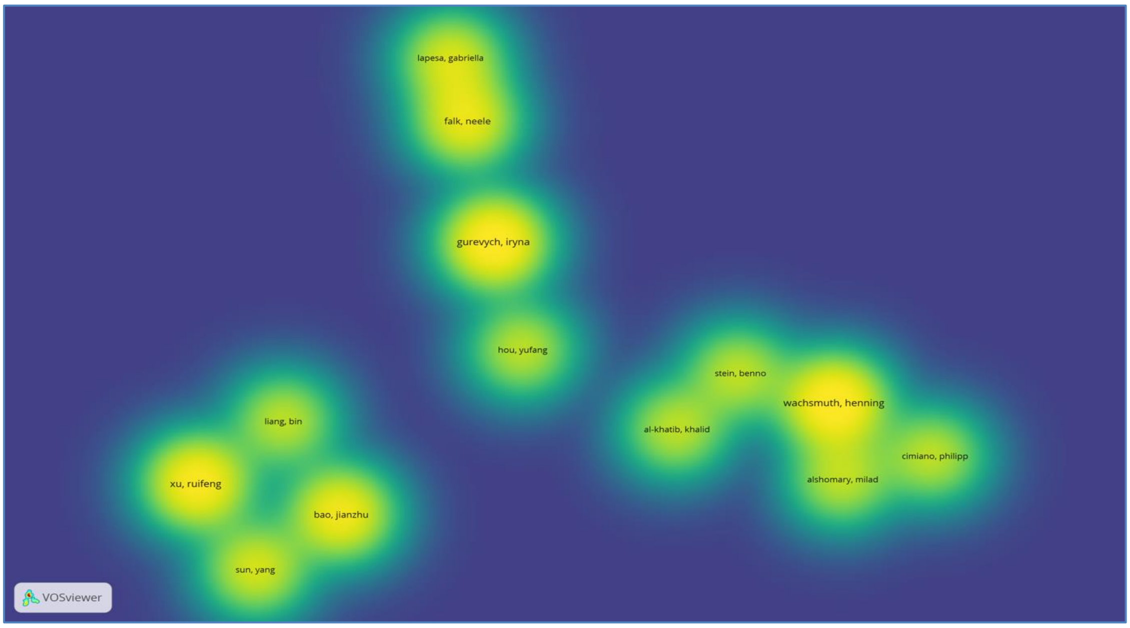

Figure 3. Co-authorship network

Note. Network visualization, VOSviewer; n = 13 authors, 3 clusters. Node size is proportional to the number of documents; link thickness is proportional to link strength.

Table 6. Authors in the co-authorship network

Cluster	Author	Docs	Cites	Link strength	Mean pub. year	Mean norm. citations
1	Wachsmuth, H.	10	237	12	2022.3	2.21
1	Stein, B.	5	163	10	2022.0	2.74
1	Alshomary, M.	5	152	9	2022.2	2.81
1	Al-Khatib, K.	5	95	9	2021.8	1.37
1	Cimiano, P.	5	39	2	2022.6	0.74
2	Gurevych, I.	10	257	3	2023.6	2.72
2	Falk, N.	7	73	7	2022.7	1.22
2	Lapesa, G.	6	66	6	2022.5	1.21
2	Hou, Y.	5	62	10	2023.0	1.27
3	Xu, R.	9	93	20	2023.8	1.29
3	Bao, J.	8	90	19	2023.6	1.28
3	Sun, Y.	6	20	17	2024.3	0.79
3	Liang, B.	5	20	14	2024.2	0.94

The three clusters differ in period of activity and in internal cohesion. Cluster 3 has the latest mean publication year, from 2023.6 to 2024.3, and the highest internal link strength, from 14 to 20 per author, but the lowest mean normalised citations, from 0.79 to 1.29. Cluster 1 has an earlier mean publication year and higher mean normalised citations, reaching 2.74 and 2.81 for two authors. This difference is consistent with more recent publications having had less time to accumulate citations, even though the indicator is normalised by publication year.

Document citation network

The document citation network retained 21 documents in the largest connected component, distributed across six clusters of two to five documents each (Figure 4). All 21 documents belong to political science and concern violence in civil war, nonviolent action and the long-term political consequences of conflict. The source journals are Journal of Peace Research (4 documents), Comparative Political Studies (4), Journal of Conflict Resolution (3) and nine further journals with one document each. No education journal appears.

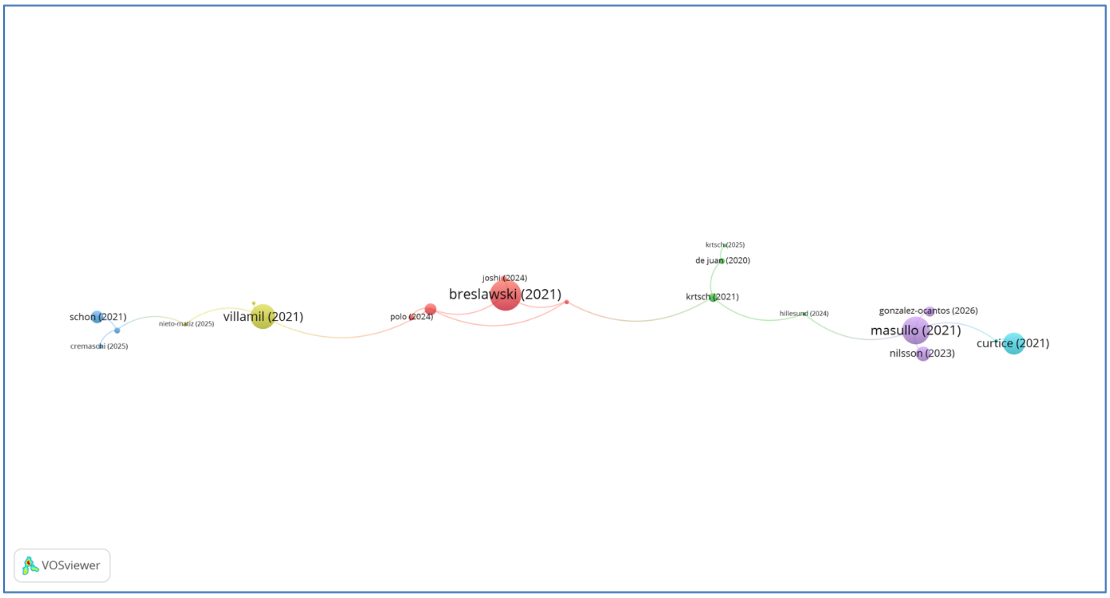

Figure 4. Document citation network

Note. Network visualization, VOSviewer; n = 21 documents in the largest connected component, 6 clusters. Node labels follow the first author and year format.

Table 7. Six clusters in the document citation network

Cluster	Docs	Documents
1	5	Bauer (2026), Joshi (2024), Polo (2024), Welsh (2023), Breslawski (2021)
2	4	Krtsch (2025), Hillesund (2024), Krtsch (2021), de Juan (2020)
3	3	Cremaschi (2025), Villamil (2022), Schon (2021)
4	3	Aguilar (2025), Nieto-Matiz (2025), Villamil (2021)
5	3	González-Ocantos (2026), Nilsson (2023), Masullo (2021)
6	2	Son (2025), Curtice (2021)

In terms of temporal distribution, the documents span 2020 to 2026, with concentrations in 2021 (6 documents) and 2025 (5 documents). In terms of influence, the two documents with the highest normalised citations were both published in 2026: González-Ocantos and Masullo at 25.98 with 12 citations, and Bauer and Cunningham at 8.66 with 4 citations. The high values of the normalised indicator in these two cases reflect a small denominator, since the mean citation count of documents published in 2026 within the dataset remains very low. Among documents published in 2021, Breslawski reaches 3.77 with 52 citations and Masullo reaches 3.19 with 44 citations.

The four analyses converge on one point. The keyword co-occurrence network shows thematic clusters separated by content. The co-citation network shows seven distinct intellectual bases, of which only cluster 7 is cited across clusters. The co-authorship network displays a single community. The document citation network displays a single connected component belonging to one field, and that field does not coincide with the community appearing in the co-authorship network.

Discussion

The results of the four analyses indicate that the literature retrieved by the query does not form a unified research field but a set of scholarly communities sharing a common terminology. Evidence for this comes from all four networks: thematic clusters separated by content, seven distinct intellectual bases in the co-citation network, a single collaboration community displayed in the co-authorship network, and a single connected component belonging to a different field in the document citation network. Those four techniques exploiting four different types of relation lead to the same conclusion about fragmentation increases confidence in that conclusion (Donthu et al., 2021).

The source of the fragmentation lies in the polysemy of the terminology. Cluster 4 of the co-citation network assembles works on generative syntax, in which argument denotes a verbal argument in the sense of predicate structure, a meaning unrelated to argumentation as understood in science education. Cluster 5 assembles language models used for the automatic extraction of argument structure from text, that is, a technical task rather than a learning objective. These two clusters account for 55 of the 251 cited references. This result indicates that using argument as a free-text keyword in the TITLE-ABS-KEY field generates a substantial proportion of noise, and that subsequent reviews should add subject area or learning context constraints at the query stage if the objective is content analysis.

Within this fragmentation, the science education segment has a clearly defined intellectual base. Cluster 2 of the co-citation network assembles policy documents such as the K-12 framework (National Research Council, 2012) together with research on classroom argumentation norms (Driver et al., 2000), on pedagogical intervention (Osborne et al., 2004), on discourse analysis tools (Erduran et al., 2004), on the distinction between explanation and argument (Berland & Reiser, 2009), on learning progressions (Berland & McNeill, 2010) and on the psychological basis of argumentation skills (Kuhn, 1991). The mean link strength of this cluster reaches 150.6, the second highest among the seven clusters. This level of cohesion indicates that studies within the segment share a common set of reference works, a condition favourable to the accumulation of knowledge.

The position of cluster 7 adds a further detail. With 8 references but a mean link strength of 384.9, this cluster comprises works in argumentation theory cited from several different clusters, among them Toulmin (2003) with the highest link strength in the network. Communities that differ in object and method nonetheless share a common theoretical foundation at a general level, and the framework for analysing argument structure serves as a bridge among them.

As regards recent dynamics, the keywords emerging in the final two years are all associated with generative artificial intelligence and all belong to cluster 2, the cluster that contains both the technical and the educational branch. Link strength differs considerably between the two branches, which indicates that they have not merged. That chatgpt and educational technology both have mean normalised citations above 3.6 indicates that publications at this intersection are attracting attention.

For the Vietnamese context, the results carry two implications. First, the requirement for evidence-based argumentation in the 2018 General Education Curriculum (Ministry of Education and Training, 2018) can be operationalised through frameworks already tested within cluster 2, in particular the three-dimensional learning progression and the criteria for assessing argumentative products. Second, the absence of keywords denoting school-level contexts from the list of 1,000 keywords, together with the concentration of domestic research on argumentation in the language disciplines (Nguyen & Nguyen, 2020), indicates scope for studies of argumentation in the teaching of school science subjects.

Finally, the results also reflect properties of the analytical tool itself. That biomedical index terms occupy six of the ten highest frequency positions, and that Toulmin's work is split into three records across three clusters, indicates that the quality of a science map depends on preprocessing operations more than on the clustering algorithm (van Eck & Waltman, 2010; Zupic & Čater, 2015). Bibliometric studies in education should report these decisions explicitly.

Limitations

Several limitations should be noted when interpreting the results. The data come from Scopus only, so publications not indexed in that database, including most Vietnamese journals, fall outside the scope of the survey. The search string has low specificity and generates a substantial proportion of noise, although this was a deliberate choice serving the third research question. The absence of any keyword denoting a specific school science subject above the display threshold reflects frequency below threshold and does not demonstrate that such works are absent from the dataset. Keywords and cited references were not normalised, so several variants of the same unit were counted separately. Cluster structure depends on selection thresholds and the resolution parameter, so the partition obtained is one of several acceptable partitions. The co-authorship network and the document citation network retained few units, so statements about these two networks apply to the portion above threshold rather than to the dataset as a whole.

Conclusion

This study describes the knowledge structure of 14,938 Scopus records on argumentation associated with data and evidence, using four science mapping techniques. As regards thematic structure, 1,000 keywords are distributed across five clusters, of which the cluster concerning computation, artificial intelligence and education is the most recent in time and contains most of the keywords on scientific argumentation. The topics emerging in the final two years are all associated with generative artificial intelligence.

As regards the intellectual base, 251 cited references are distributed across seven clusters corresponding to seven distinct research traditions. The cluster on argumentation in science education has the highest total link strength and assembles foundational works formed mainly between 2000 and 2010. The argumentation theory cluster is the smallest in size but has the highest mean link strength, indicating a bridging role among communities.

As regards collaboration structure and lines of influence, the co-authorship network displays only 13 authors from a single community, and the document citation network displays only 21 documents from a different field. Taken together, the four results characterise the field under study as a set of communities that share terminology but share little in the way of intellectual foundations and engage in little direct collaboration.

Two implications follow. For review methodology, studies of argumentation should add subject area or context constraints at the query stage and should report data preprocessing decisions explicitly. For science education research, the inventory of foundational works identified here can serve as a starting point for constructing a theoretical framework. A subsequent line of work is a systematic review of a corpus filtered by an improved query, focused on the teaching of school science subjects.

References

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. Science Education, 94(5), 765–793. https://doi.org/10.1002/sce.20402

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. Science Education, 93(1), 26–55. https://doi.org/10.1002/sce.20286

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84(3), 287–312. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. Science Education, 88(6), 915–933. https://doi.org/10.1002/sce.20012

Kuhn, D. (1991). The skills of argument. Cambridge University Press.

Ministry of Education and Training. (2018). General education curriculum: Master curriculum (Issued with Circular No. 32/2018/TT-BGDĐT of 26 December 2018). Hanoi. [In Vietnamese]

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. The National Academies Press. https://doi.org/10.17226/13165

Nguyen, T. S., & Nguyen, H. B. (2020). Unravelling Vietnamese students' critical thinking and its relationship with argumentative writing. Universal Journal of Educational Research, 8(11B), 5972–5985. https://doi.org/10.13189/ujer.2020.082233

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. Journal of Research in Science Teaching, 41(10), 994–1020. https://doi.org/10.1002/tea.20035.

Toulmin, S. E. (2003). The uses of argument (Updated ed.). Cambridge University Press.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. Organizational Research Methods, 18(3), 429–472. https://doi.org/10.1177/1094428114562629.