



Pedagogical Agents in Mathematics Education (1999–2025): A Bibliometric Analysis of a Field Shifting Toward Large Language Models

Fadhil Zil Ikram¹, Kusnandi^{2*}

^{1,2}Universitas Pendidikan Indonesia, Bandung, Indonesia

ARTICLE INFO	ABSTRACT
Article history: Submitted: May 14, 2026 Final Revised: July 29, 2026 Accepted: August 16, 2026 Published: August 28, 2026	Purpose: This study maps the scientific landscape of pedagogical agents in mathematics education (1999–2025) to characterize its developmental trajectory, key actors and intellectual foundations, and thematic structure, and to locate mathematics learning within the broader pedagogical-agents paradigm. Methods: A corpus of 133 Scopus-indexed documents was analyzed with Biblioshiny and VOSviewer, combining performance analysis (production, citation impact, sources, authors, affiliations, countries) and science mapping (keyword co-occurrence at a minimum threshold of three, thematic mapping, and overlay visualization). Findings: The field shows an annual growth rate of 11.25%, far above the global-science baseline of 4%. Knowledge production concentrates in three institutional clusters (Lund University, Carnegie Mellon, UMass Amherst), with international co-authorship at 19.55%, below the cross-disciplinary average. Pedagogical agents anchor the field, with an intellectual base in cognitive principles of learning-by-teaching, while Large Language Models emerged as a converging keyword after 2022. Author-keyword co-occurrence resolved eight clusters dominated by general instructional-technology and AI motor themes, with mathematics learning fragmented across three clusters in a marginal, basic-theme position an application context rather than an epistemic core. Research Implications: Findings highlight open positioning opportunities for new entrants, guide developers integrating LLMs into established pedagogical-agent architectures, and signal to policymakers the need for broader institutional participation. Originality: This is the first bibliometric mapping specifically targeting the pedagogical-agents and mathematics-education intersection, documenting continuity in the paradigm shift from classical to LLM-based agents. As the corpus is drawn from a single database, it should be read as a Scopus-bounded, English-leaning view rather than an exhaustive census.
	
 	Doi: https://doi.org/10.61255/jupiter.v4i4.1200

INTRODUCTION

Mathematics is a foundational human competence: it hones transferable problem-solving skills used across academic, daily, and professional settings, and cultivates the critical, systematic thinking needed to meet contemporary challenges (Maass & Engeln, 2019; Wulandari et al., 2024). In the contemporary era, there has been an increase in the utilisation of artificial intelligence (AI) in educational settings, including those dedicated to mathematics instruction. This is due to the fact that AI is not merely a tool but rather a partner in education and learning (Li, 2025). The utilisation of AI has been demonstrated to facilitate enhanced personalisation and adaptivity in learning, thereby ensuring accessibility to mathematics learning materials and catering to the diverse requirements of learners (Koedinger & Alevan, 2007; Lyu et al., 2025a). This enhancement encompasses the improvement of learners' mathematical knowledge.

Although AI holds clear potential to improve mathematics learning, many applications still position learners as passive recipients, limiting active participation (Lyu et al., 2025b). In education, AI is increasingly embodied as a pedagogical agent (PA) an autonomous on-screen character that interacts with learners within a learning environment (Johnson & Lester, 2018). Beyond delivering content, PAs provide affective support and personalised, needs-adaptive

feedback that can raise engagement and learning outcomes (Schlimbach et al., 2022; Schroeder et al., 2013; Tao et al., 2022). The paradigm has a three-decade trajectory: it gained traction in the mid-1990s with Lester et al.'s (1997) 'persona effect', whereby lifelike characters improved learners' perceptions of their experience; Johnson et al. (2000) then articulated a PA research agenda integrating animated interfaces with intelligent learning environments; and Baylor and Kim (2005) validated three instructional roles (Expert, Motivator, Mentor) with distinct effects on motivation and learning. Meta-analyses subsequently confirmed that PAs produce positive but generally small learning effects, moderated by design and instructional context (Castro-Alonso et al., 2021; Schroeder et al., 2013). Most recently, the release of ChatGPT in late 2022 has opened an LLM-based phase marked by more natural, adaptive generative dialogue while raising challenges of factual accuracy, learner overreliance, and pedagogical validation (Yan et al., 2024).

Research on pedagogical agents in mathematics education has increased both in terms of the number of publications and technological complexity, particularly since the development and release of various generative AI-powered pedagogical agents. Adjacent reviews circle, but do not occupy this space: In the field of educational technology, a seminal review of educational chatbots was conducted by Kuhail et al. (2023). Building upon this, Martha and Santoso (2019) undertook a comprehensive review of pedagogical agents, with a particular focus on those not dedicated to mathematics. The study by Samuel et al. (2025) is a pioneering investigation into the utilisation of educational robots in the domain of mathematics, employing a SWOT analysis to assess their effectiveness and potential impact. Existing bibliometric studies of AI in mathematics education (Daulay et al., 2025; Nguyen & Pham, 2025; Subroto et al., 2024; Yıldız & Körpeoğlu, 2025) treat AI as a broad construct rather than the pedagogical-agent paradigm. No study has bibliometrically mapped the pedagogical agents and mathematics education intersection itself – the explicit gap this study aims to fill. The present studies address the topic of artificial intelligence in mathematics education in general, rather than focusing on the specific application of AI agents. Furthermore, the significant increase in publications during the 2023–2025 period concerning pedagogical agents in mathematics education indicates that this field has reached a tipping point, yet no bibliometric map exists for the pedagogical-agents × mathematics-education intersection specifically; its intellectual architecture, collaboration patterns, and thematic trajectory thus remain unexamined at the macro level. A bibliometric approach rather than a narrative or systematic review is warranted precisely at this scale: it aggregates 133 documents into a reproducible structural map of authors, venues, and concepts that qualitative synthesis cannot deliver (Donthu et al., 2021). The present study addresses this gap by triangulating citation analysis, keyword co-occurrence networks, thematic mapping, and thematic evolution analysis (Aria & Cuccurullo, 2017; Cobo et al., 2011) to characterise how pedagogical agents in mathematics education evolved. The bibliometric analysis is of paramount importance for a field such as PA in mathematics education, which is situated at the intersection of artificial intelligence, learning sciences, and mathematics education, and whose evolution is being shaped by rapid technological transitions.

The findings of this study are of direct benefit to three stakeholder groups: Thematic cluster analysis can be utilised by educational software developers to identify under-explored design dimensions; the findings on the publication ecosystem can be used by mathematics curriculum designers to identify gaps between pedagogical practices and technology-oriented research venues; and citation velocity analysis can be used by AI researchers to identify foundational and high-impact frontier works for the development of generative AI-powered PA.

This study will address the following research questions.

1. What are the trends in scientific output and citation intensity in the field of pedagogical agents in mathematics education over the period 1999–2025?
2. Who are the key academic actors authors, institutional affiliations and publication channels that form the basis of contributions to this field?
3. What is the global geographical distribution of scientific output and citation impact in this field, and what are the implications for the representation of researchers from different regional contexts?
4. What documents and references constitute the intellectual foundation and most influential outputs of this field, both globally and within the corpus?
5. How is the thematic structure of this field currently configured, and how has it evolved over time?

METHOD

This study uses bibliometric analysis to examine the existing literature on pedagogical agents in mathematics education. The analysis is employed when researchers must manage substantial scientific data sets to unravel thematic progression within a specific research domain (Aria & Cuccurullo, 2017; Cobo et al., 2011; Donthu et al., 2021). Zupic and Čater (2015) furthermore posit that this method is used to map research specialisations and has the potential to

increase rigour and reduce reviewer bias through evidence aggregation. Adapting the stages proposed by Büyükkıdık (2022) and Zupic and Čater (2015), the research procedures are presented in Figure 1.

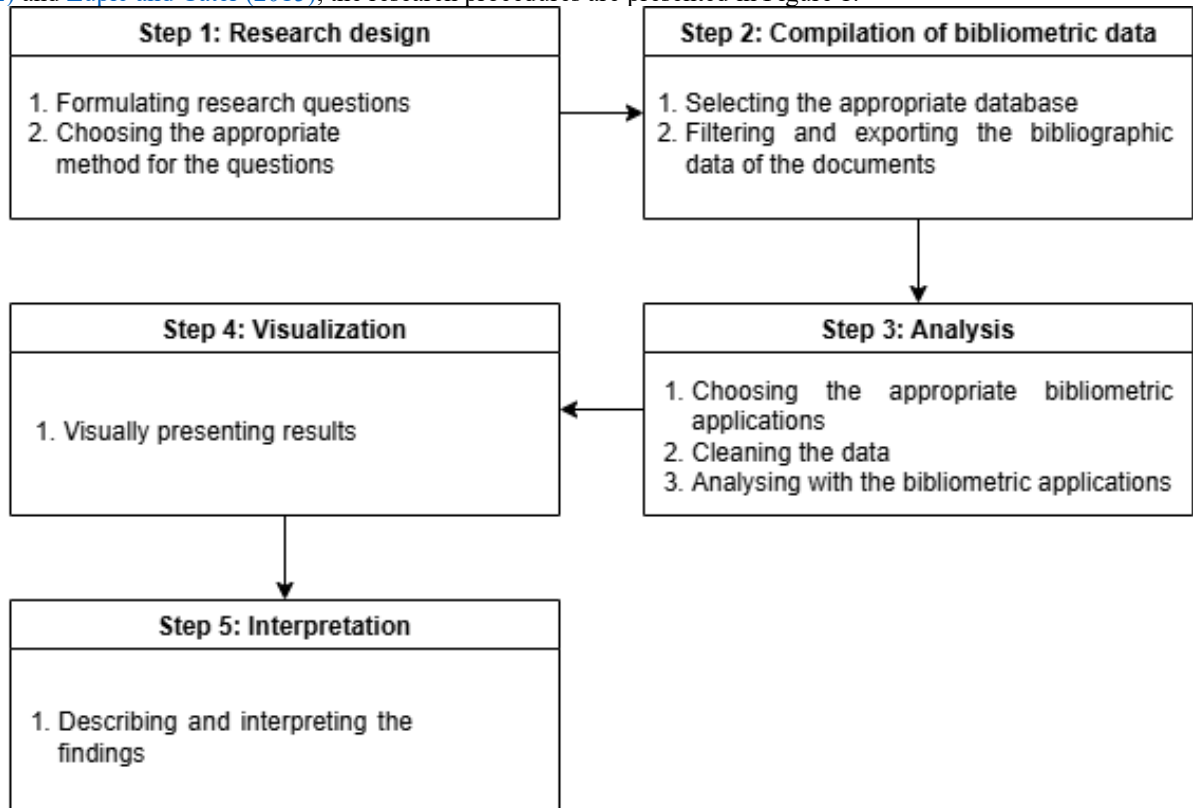


Figure 1. Research Procedure

Scopus was selected as the primary database for its comprehensive, curated metadata, its coverage of the peer-reviewed conference proceedings central to this field (AIED, ITS, CHI) alongside journals, and its advanced search and extraction functions. The search was run on 7 March 2026 at 9.00 PM in the TITLE-ABS-KEY field, combining two concept blocks pedagogical-agent terms and mathematics terms with the Boolean operator AND and truncation for term variants; the full query is reported verbatim in Figure 2. Following PRISMA-ScR (Tricco et al., 2018) and the workflow of Donthu et al. (2021), retrieved records were screened against explicit criteria journal articles and conference papers only, English-language only, and the 1999–2025 window excluding records that did not substantively address pedagogical agents in mathematics education on abstract screening. The complete identification-to-inclusion flow, with the count and exclusion reason at each stage, is shown in Figure 2 (177 identified; 133 included).

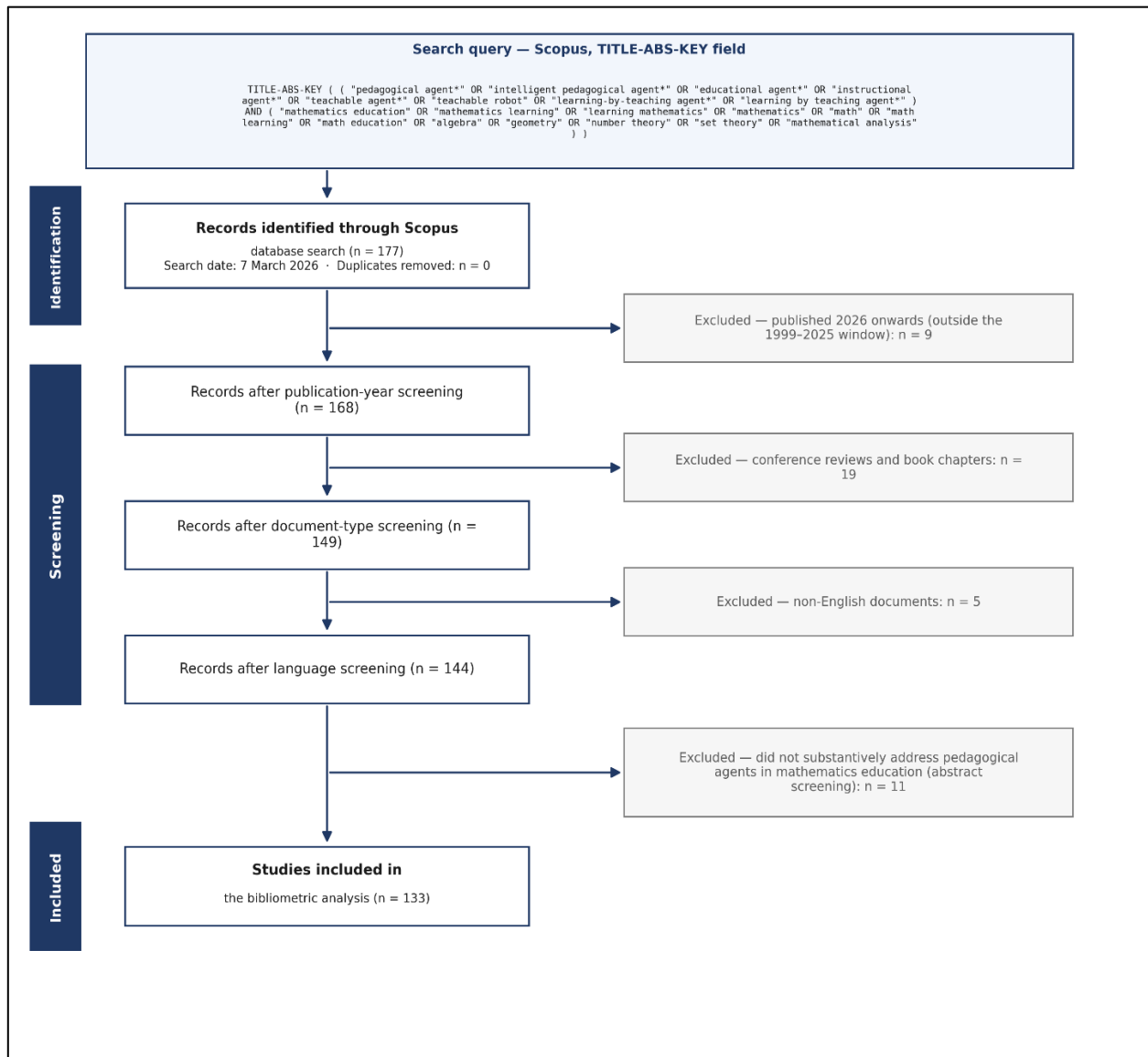


Figure 2. Metadata Extraction Procedure from Scopus

Metadata were cleaned in OpenRefine in three documented stages. (1) Normalisation: trimming whitespace, collapsing internal spaces, and standardising case across fields. (2) Keyword consolidation by clustering followed by manual review of every proposed merge, governed by an explicit rule set—singular/plural unification ('agent' → 'agents'), hyphenation variants ('learning-by-teaching' → 'learning by teaching'), orthographic variants ('math education' / 'maths education' → 'mathematics education'), and a controlled synonym map to a single preferred term. (3) Removal of within-record duplicate keywords. The clustering in step (2) was performed using OpenRefine's key-collision (fingerprint) and nearest-neighbour methods, with every proposed merge reviewed manually before acceptance. For the conceptual structure analysis, Author Keywords were used as the primary source of terms. For documents in which Author Keywords were unavailable, Index Keywords provided by Scopus were used as a substitute to ensure the inclusion of all relevant publications in the keyword-based analysis.

The analysis used the Bibliometrix R package (Biblioshiny) for comprehensive scientific mapping (Aria & Cuccurullo, 2017) and VOSviewer for graphical bibliometric mapping (van Eck & Waltman, 2010). Regarding conceptual structure, a minimum keyword co-occurrence threshold of three was applied, with full counting (so that each occurrence contributed equally to the network) and association strength normalisation method (representing the relative strength of keyword relationships); automatic clustering based on the co-occurrence algorithm was used to

generate thematic clusters. The threshold of three was selected to balance network interpretability and thematic representativeness within the relatively limited corpus: lower thresholds introduce excessive noise, whereas higher thresholds exclude relevant but less frequent keywords (Donthu et al., 2021; van Eck & Waltman, 2010; Zupic & Čater, 2015). Because bibliometric research analyses publicly available metadata and publication patterns rather than evaluating the methodological quality of individual studies, risk-of-bias assessment was not conducted.

RESULTS

This section reports the results in two parts performance analysis followed by science mapping beginning with an overview of the corpus.

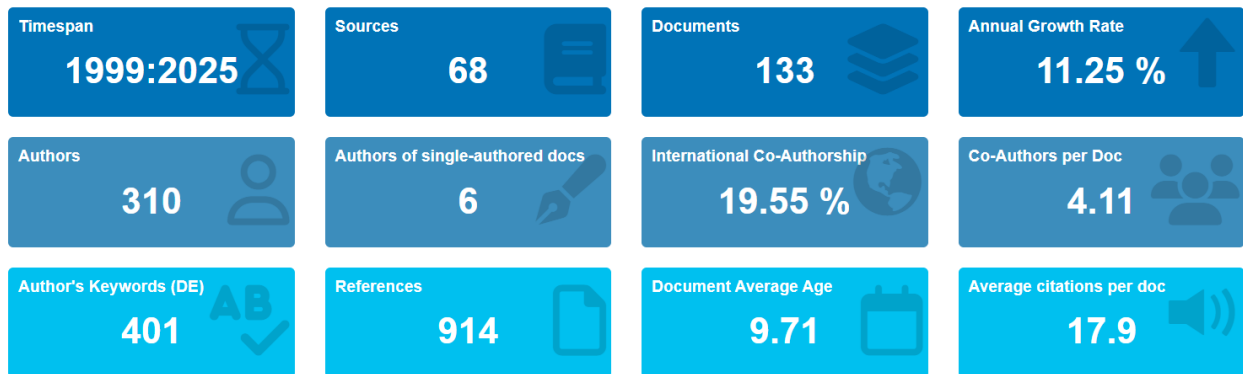


Figure 3. Corpus Metadata Summary from Scopus

The corpus extracted from Scopus covers a 26-year period (1999–2025), comprising 133 documents from 68 sources and involving 310 authors. The annual growth rate of 11.25% exceeds the estimated average growth rate of global scientific publications, which was approximately 4% per year for the period 2007–2017 (National Science Board [NSB], 2019) and remains above the estimated 8–9% per year for the modern phase of science up to 2012 (Bornmann & Mutz, 2015). This acceleration aligns with the mapping by Liu et al. (2024), which documents a surge in publications on pedagogical agents (PA) not specifically in mathematics education following 2016 alongside advancements in artificial intelligence. The authorship patterns reveal a prominent collaborative culture only six documents (4.5%) were single-authored, with an average of 4.11 co-authors per document consistent with the multidisciplinary nature of PAs identified in the co-cited source distribution of Liu et al. (2024), which spans journals in education, psychology, educational technology, computer science, and human–computer interaction. The international co-authorship rate of 19.55% is comparable to the internationalisation rate in educational research reported by Sezgin et al. (2022), yet remains below the global interdisciplinary average of 25.7% in 2021 (Aksnes & Sivertsen, 2023), indicating that collaboration remains predominantly domestic a traditional pattern described by Aman and Botte (2017) as a “nationally oriented tradition” in educational research. The ratio of 401 author keywords across 133 documents (≈ 3.01 per document) indicates moderate thematic diversification. The document average age of 9.71 years should be interpreted with caution, as without a reported temporal distribution, this value may signify an accumulation of mature literature or a slowdown in recent output. An average of 17.9 citations per document equivalent to ≈ 1.84 citations/document/year when normalised against the corpus’s average age indicates moderate impact. It is worth noting that this corpus substantially expands the empirical base: a recent meta-analysis of PA effectiveness by Castro-Alonso et al. (2021) identified only a small number of studies in the field of mathematics; thus, the 133 documents in the PA $\times$ mathematics education intersection mapped here represent a significant contribution to mapping the literature landscape of this field.

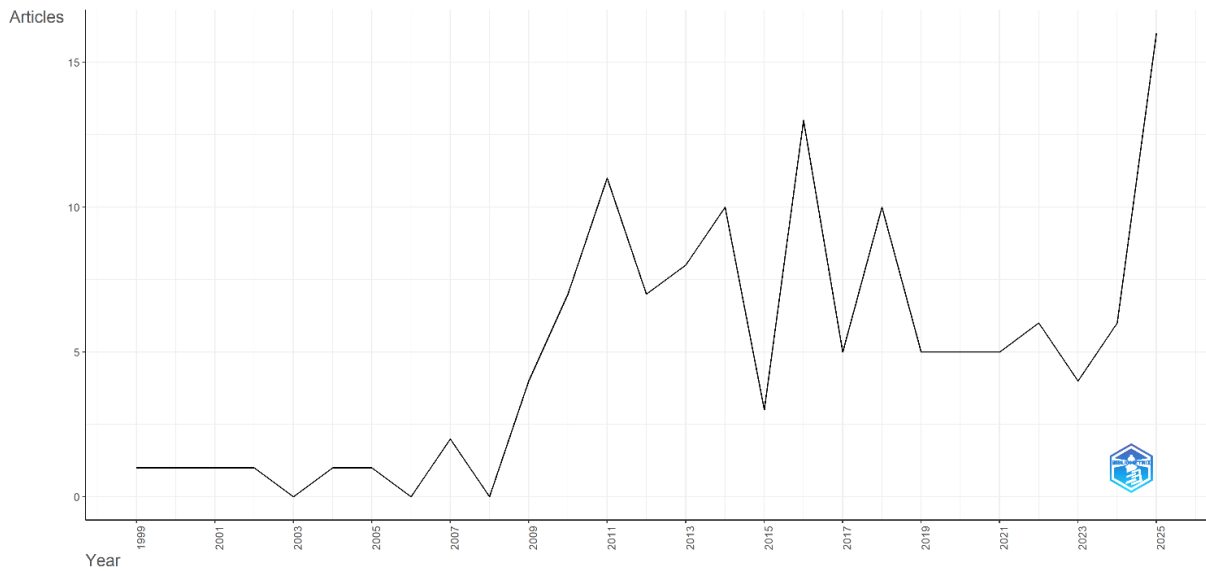


Figure 4. Annual Scientific Production

To examine this pattern of scholarly output in greater detail, Figure 4 breaks down the temporal distribution of the articles comprising the corpus. Figure 4 presents the annual scientific output for the field of pedagogical agents in mathematics education over the period 1999–2025, which can be mapped into four temporal segments. During the period 1999–2008, publications were very limited (0–2 articles per year, with no publications in 2003, 2006, and 2008). This period coincided with the emergence of the first generation of animated pedagogical agents such as AutoTutor, designed for computer literacy (Graesser et al., 1999), and Steve for procedural training in virtual reality environments (Johnson & Rickel, 1997) which had not yet been applied extensively in the field of mathematics. The period 2009–2018 shows fluctuating expansion (3–13 articles/year) with peaks in 2011 (≈ 11) and 2016 (≈ 13); this period coincides with two decades of maturation in research on pedagogical agents in general following their initial introduction in the late 1990s (Johnson & Lester, 2016). The period 2019–2023 shows stabilisation at around 4–6 articles per year, whilst the period 2024–2025 is marked by a sharp surge to ≈ 16 articles in 2025, coinciding with the integration of Large Language Model–based agents into mathematics learning research, as empirically demonstrated by a study comparing the learning effects of ChatGPT-generated hints with human tutor-authored hints on topics in algebra and statistics (Pardos & Bhandari, 2024). The Compound Annual Growth Rate (CAGR) from 1 article in 1999 to 16 articles in 2025 is approximately 11.3% per year.

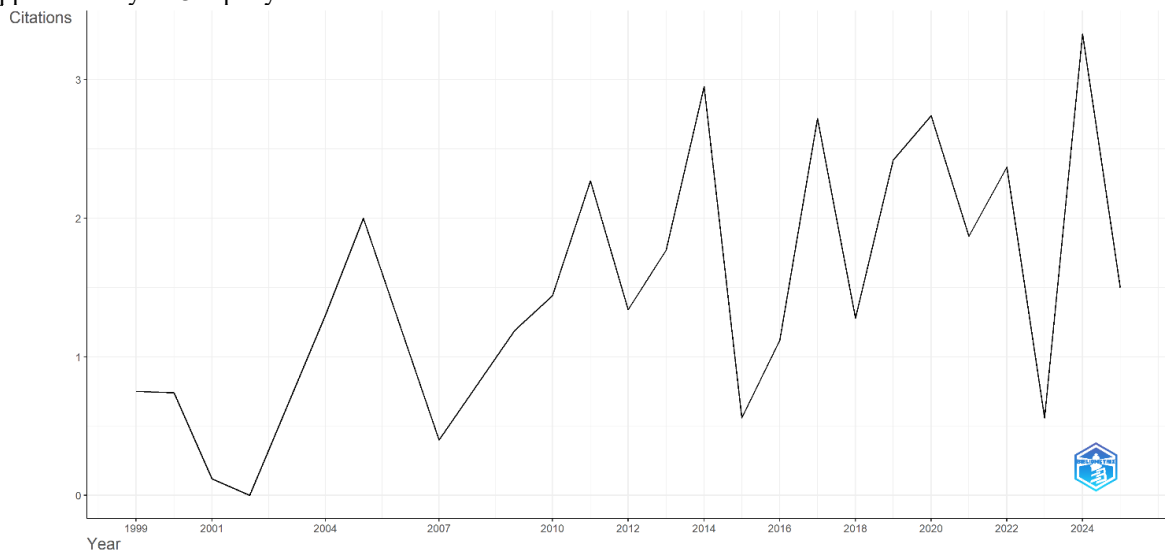


Figure 5. Average Citations per Year

Citation intensity mean citations per article per year, normalised for document age so that recent and older cohorts are comparable tells a different story from raw output. Across 1999–2025 the metric stays within a modest 0.0–3.3 band (Figure 5), but its trajectory is the analytically important feature: the decade mean rises roughly fivefold, from ≈ 0.4 in 1999–2008 to ≈ 2.0 in 2015–2024, confirming that the field’s volumetric growth (Figure 4) is matched by growing scholarly uptake rather than mere proliferation. Two qualifications follow directly from the small corpus. First, the year-to-year volatility including the ≈ 2.9 peak in 2014 and the adjacent ≈ 0.6 trough in 2015 is largely an artefact of low annual denominators, in which a single highly cited paper (e.g., the Wayang Outpost study of Arroyo et al., 2014) can dominate a cohort. Second, the ≈ 3.3 maximum in 2024 should not be read as a durable high: it reflects a short, intense citation window for very recent LLM-related work and is expected to regress as those cohorts age. Citation intensity in this field is therefore best characterised as bursty and concentration-driven rather than smoothly cumulative.

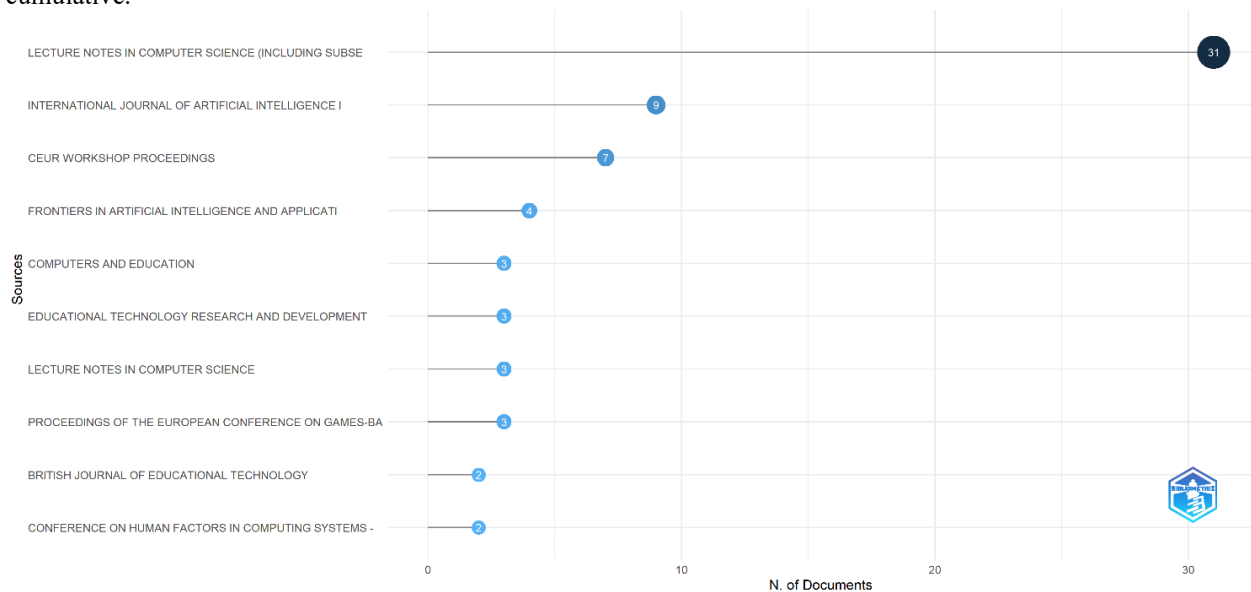


Figure 6. Most Relevant Sources

Having outlined the temporal dimensions and aggregate impact, attention now turns to the academic actors who make up the corpus, starting with the publication channels. Figure 6 displays the ten most frequent publication sources in the corpus. Lecture Notes in Computer Science (LNCS), including its sub-series, dominates significantly with 31 documents, followed by the International Journal of Artificial Intelligence in Education (IJAIED) with 9 documents, CEUR Workshop Proceedings (7), and Frontiers in Artificial Intelligence and Applications (4). Four sources share the next ranking with 3 documents each: Computers and Education, Educational Technology Research and Development, Lecture Notes in Computer Science (separate entries from the sub-series), and Proceedings of the European Conference on Games-Based Learning. The final two positions are occupied by the British Journal of Educational Technology and the Conference on Human Factors in Computing Systems (CHI), each with 2 documents. This distribution reveals a high concentration on computer science conference proceedings particularly LNCS volumes hosting core conferences in this field such as Artificial Intelligence in Education (AIED), Intelligent Tutoring Systems (ITS), and Intelligent Virtual Agents (IVA) compared to mainstream mathematics education journals, which do not appear in the top ten. This pattern is consistent with Bradford’s Law, which predicts a concentrated distribution across a small number of core sources.

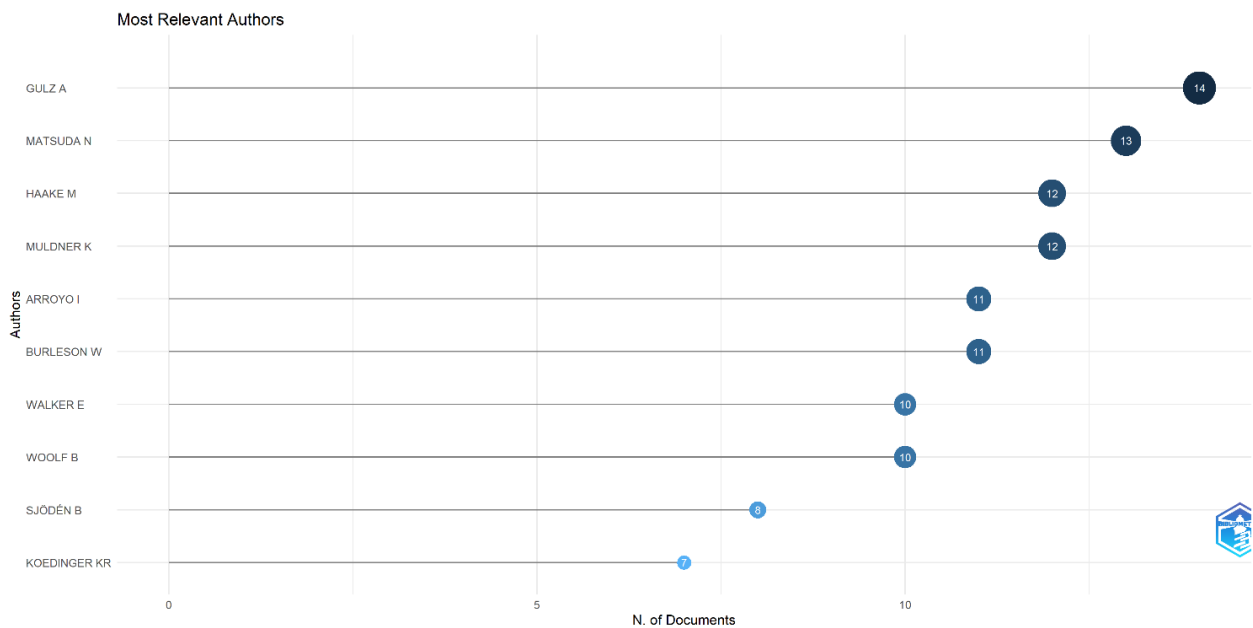


Figure 7. Most Relevant Authors

This mapping of publication channels needs to be combined with the identification of the most productive researchers in this field. Figure 7 shows the ten most productive researchers in the corpus, led by A. Gulz (14 documents), N. Matsuda (13), M. Haake (12), and K. Muldner (12), followed by I. Arroyo and W. Burleson (11 each), E. Walker and B. Woolf (10 each), B. Sjödén (8), and K. R. Koedinger (7). These ten researchers can be grouped into three dominant institutional clusters. The first cluster is centred at Lund University, Sweden, with Gulz, Haake, and Sjödén as the main contributors through the development of a game-based teachable agent, the impact of which on students' mathematical understanding and motivation was evaluated (Pareto et al., 2012). The second cluster is centred at Carnegie Mellon University, with Matsuda and Koedinger as the main drivers of the SimStudent project a machine learning-based teachable agent designed to help students learn to solve algebraic equations through the 'learning by teaching' paradigm (Matsuda et al., 2013). The third cluster is centred at the University of Massachusetts Amherst, bringing together Arroyo, Woolf, Burleson, and Muldner in the development of Wayang Outpost an Intelligent Tutoring System with an affective agent component for high-stakes mathematics testing (Arroyo et al., 2014).

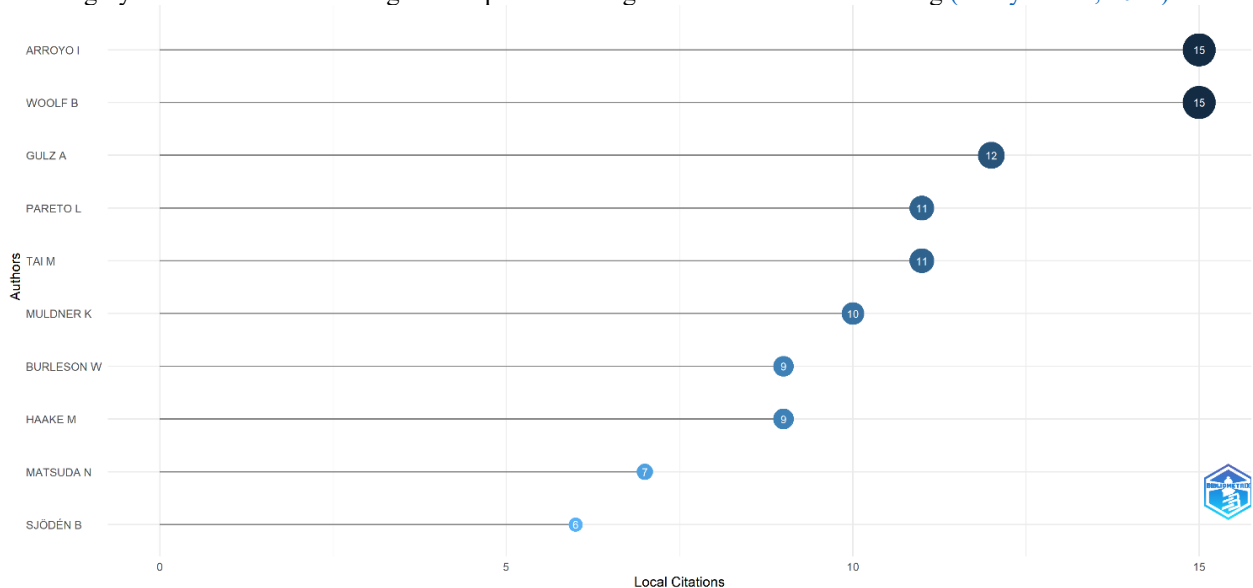


Figure 8. Most Local Cited Authors

Absolute productivity, as shown in Figure 7, does not automatically correlate with intellectual influence; Figure 8 presents a complementary metric that measures the extent to which a researcher's work is absorbed by the internal research community within the corpus. Figure 8 displays the ten researchers with the highest number of local citations, i.e. the number of citations received from other articles within the corpus 'pedagogical agents in mathematics education'. I. Arroyo and B. Woolf occupy the top positions with 15 local citations each, followed by A. Gulz (12), L. Pareto (11), M. Tai (11), K. Muldner (10), W. Burleson (9), M. Haake (9), N. Matsuda (7), and B. Sjöden (6). The ranking composition in Figure 8 shows a marked difference when compared to the productivity rankings in Figure 7: researchers associated with the Wayang Outpost project at the University of Massachusetts Amherst namely Arroyo, Woolf, Tai, Muldner, and Burleson (Arroyo et al., 2014) all feature within the top seven for local citations, whilst N. Matsuda, who ranked second in productivity (Figure 7), has shifted to ninth place in local citations. The Lund University cluster led by Gulz, Pareto, Haake, and Sjöden researchers who developed a game-based teachable agent paradigm for mathematics learning (Pareto et al., 2012) also consistently appears in both rankings. L. Pareto, who does not appear in the top ten for productivity in Figure 7, appears in fourth place for local citations in Figure 8 with 11 citations.

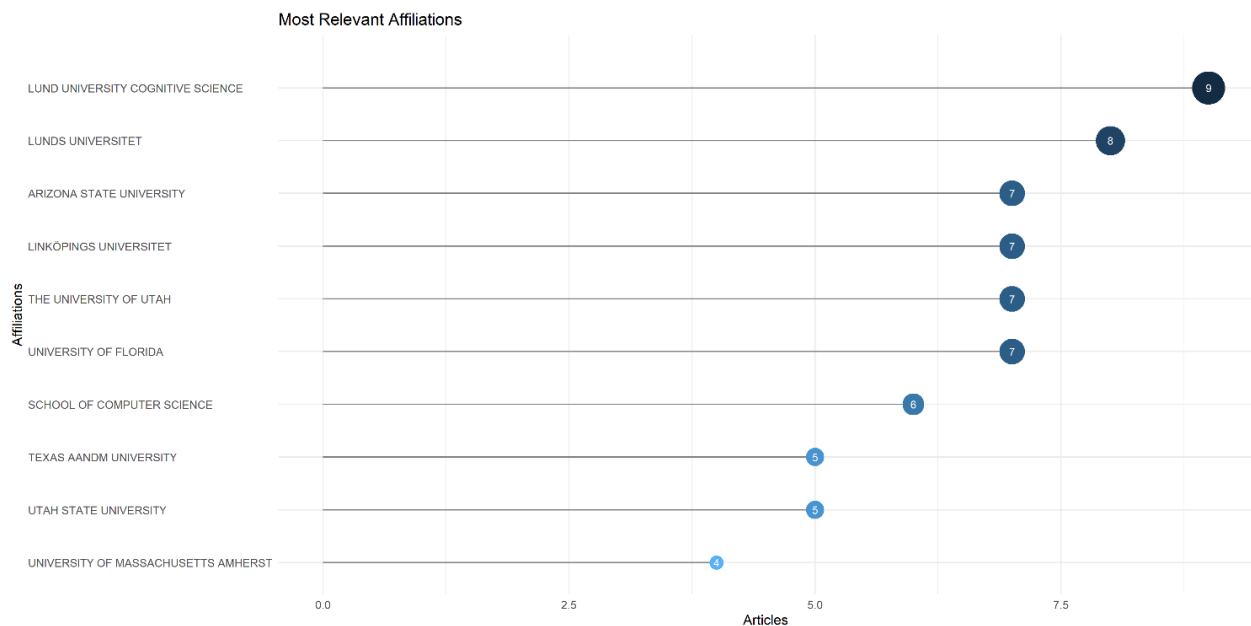


Figure 9. Most Relevant Affiliations

This mapping of individual contributions needs to be scaled up to the institutional level to map the organisational base of the field. Figure 9 shows the ten institutional affiliations with the highest number of articles in the corpus. Lund University Cognitive Science ranks first with 9 articles, and “Lunds Universitet” ranks second with 8 articles. These two entries actually refer to the same institution (Lund University, Sweden) which appears separately in the results one as the division name in English and one as the university name in Swedish thereby strengthening the institution's position with 17 articles, far surpassing other institutions. In the next tier, four affiliations appear to be competing with 7 articles each: Arizona State University, Linköpings Universitet (Linköping University, Sweden), The University of Utah, and the University of Florida. “School of Computer Science” a department name without an identified parent institution ranks seventh with 6 articles, followed by Texas A&M University (listed as “Texas AANDM University” in the results due to an ampersand character encoding artefact) and Utah State University with 5 articles, and the University of Massachusetts Amherst with 4 articles. Geographically, the top ten rankings are dominated by institutions in the United States (six affiliates) and Sweden (three affiliates).

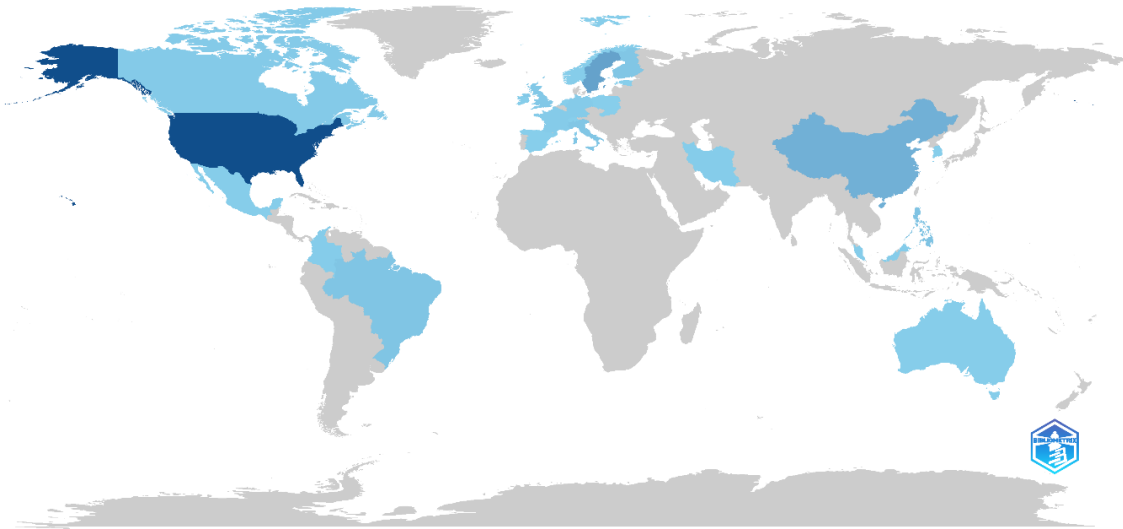


Figure 10. Country Scientific Production

Viewed geographically (Figure 10), output in the field is concentrated in the Global North and a handful of national systems rather than evenly distributed. The United States leads decisively, with Sweden and China forming a second tier and the remaining contributions clustered in Western Europe (notably the UK, Germany, the Netherlands, and the Nordic countries) plus a thin scatter elsewhere. The more telling feature is the map's silence: most of Africa, South and Central Asia, and Southeast Asia register little or no detectable activity. This is substantive rather than incidental for mathematics education much of the world's population, including regions with the most pressing needs in mathematics learning, is largely absent from the production of the very pedagogical-agent technologies proposed to serve those needs, mirroring the institutional concentration in Figure 9 (six of the top ten affiliations in the United States, three in Sweden). The geography of this field is thus one of supply concentrated far from much of its potential demand.

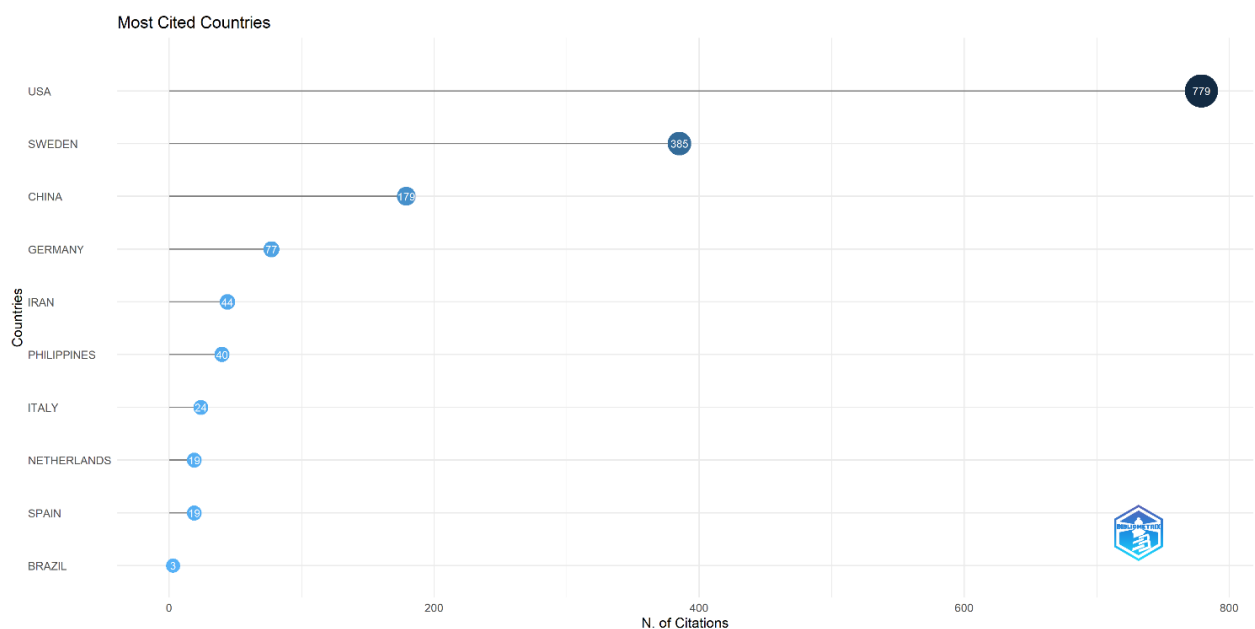


Figure 11. Most Cited Countries

A country's research output is not automatically proportional to the number of citations it receives; therefore, this dimension needs to be mapped separately. Figure 11 shows the ten countries with the highest number of citations in the corpus, calculated based on the country of affiliation of the corresponding author of each article. The United States ranks first with 779 citations, followed by Sweden (385), China (179), Germany (77), Iran (44), the Philippines (40), Italy (24), the Netherlands (19), Spain (19), and Brazil (3). This distribution reveals a highly concentrated distribution of citations at the top of the rankings: the United States alone accounts for approximately 50% of the total 1,569 citations in the top ten ($CR1 \approx 49.6\%$), whilst the top three countries (the United States, Sweden, and China) collectively account for around 86% of the total citations in the top ten ($CR3 \approx 85.6\%$). The citations-per-article ratio reveals an interesting asymmetry between countries: Sweden achieved 385 citations with institutional contributions largely originating from a single university (Lund University, as identified in Figure 9), whilst China, which appears to have substantial output on the map in Figure 10, received only 179 citations. A pattern of disproportionality also emerges in Brazil, which appears as an active producer on the map but is recorded as having received only 3 citations. Conversely, the Philippines and Iran two countries with relatively limited production visibility on the map are recorded as having received 40 and 44 citations respectively, indicating that certain articles from these countries have attracted high citation attention despite their modest volumetric contribution.

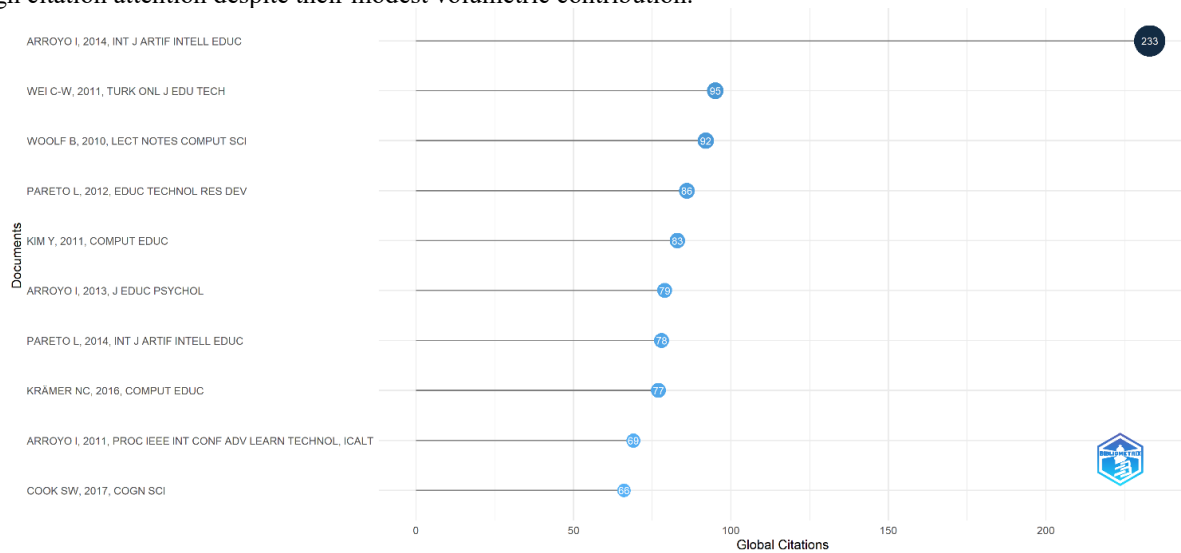


Figure 12. Most Global Cited Documents

The concentration of citations at the country level needs to be broken down to the level of individual documents in order to identify the articles that drive the field's academic impact. The most globally cited documents (Figure 12) reinforce the field's two-cluster structure rather than diversifying it. One paper stands apart [Arroyo et al. \(2014\)](#), with 233 citations, roughly two-and-a-half times the next-ranked work while the remaining nine sit in a tight 66–95 band, signalling a single breakout study atop an otherwise flat impact distribution. The list is dominated by the same groups identified in the authorship analysis (Figures 7–8): five of the top ten come from the Wayang Outpost group at UMass Amherst (four with Arroyo, one with Woolf) and two from the Lund teachable-agent group (Pareto), with only three [Wei et al. \(2011\)](#), [Kim & Wei \(2011\)](#), and [Krämer et al. \(2016\)](#) originating outside the productivity rankings. High citation impact in this field is thus concentrated in the same institutional canon that drives output, indicating a tightly held rather than broadly distributed evidence base. (Full ranked list with counts in Figure 12: [Arroyo et al., 2013](#); [Arroyo et al., 2011](#); [Arroyo et al., 2014](#); [Kim & Wei, 2011](#); [Krämer et al., 2016](#); [Pareto et al., 2012](#); [Pareto, 2014](#); [Wei et al., 2011](#); [Woolf et al., 2010](#))

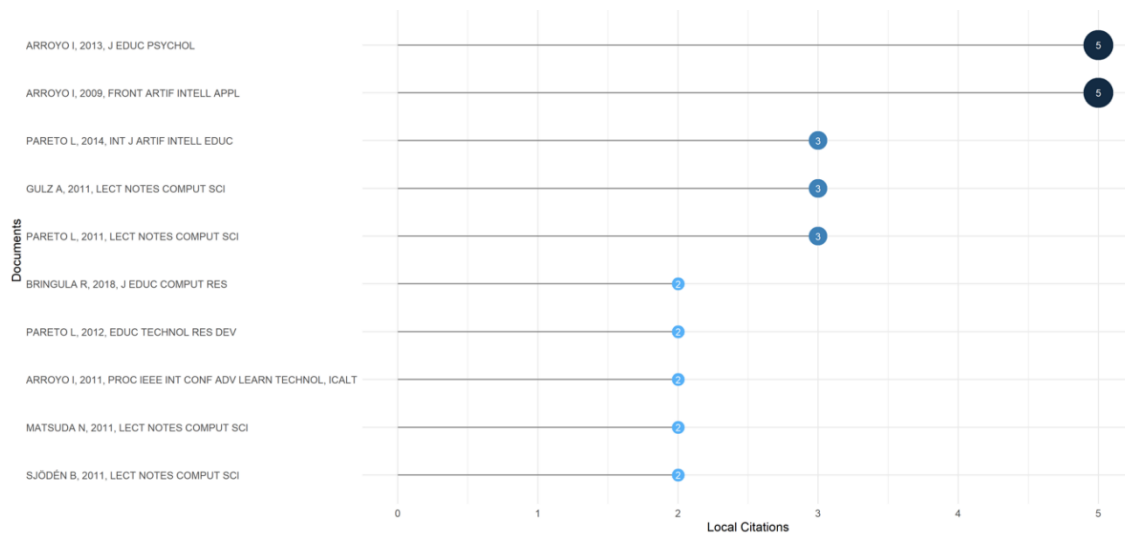


Figure 13. Most Local Cited Documents

This global impact must be considered alongside the local impact—namely, the frequency with which these documents are cited by other articles within the corpus itself. Figure 13 shows the ten documents with the highest number of local citations. Two articles rank highest with 5 local citations each: [Arroyo et al. \(2013, *Journal of Educational Psychology*\)](#) and [Arroyo et al. \(2009, *Frontiers in Artificial Intelligence and Applications*\)](#). In third to fifth place, three articles share 3 local citations: [Pareto \(2014, *IJAIED*\)](#), [Gulz et al. \(2011, *LNCS*\)](#), and [Pareto \(2011, *LNCS*\)](#). The five articles ranked sixth to tenth each received 2 local citations, namely [Bringula et al. \(2018, *Journal of Educational Computing Research*\)](#), [Pareto et al. \(2012, *ETR&D*\)](#), [Arroyo et al. \(2011, *ICALT*\)](#), [Matsuda et al. \(2011, *LNCS*\)](#), and [Sjödén et al. \(2011, *LNCS*\)](#). A comparison between Figure 12 and Figure 13 reveals two structurally distinct patterns. First, a stark contrast in absolute scale: the highest global citation count reaches 233 whilst the highest local citation count is only 5, meaning that even the most cited document in the corpus is, on average, cited by only a small fraction of other articles within the corpus. Second, the partial overlap between the two rankings: only four articles appear in both top ten lists, namely [Arroyo et al. \(2013, *J Educ Psychol*\)](#), [Pareto \(2014, *IJAIED*\)](#), [Pareto et al. \(2012, *ETR&D*\)](#), and [Arroyo et al. \(2011, *ICALT*\)](#). The article by [Arroyo et al. \(2014\)](#), which ranks first in global citations with 233 citations, does not appear in the top ten local citations in Figure 13, indicating that this article is cited more frequently by the broader research community in intelligent tutoring systems and pedagogical agents than by studies specifically addressing mathematics learning.

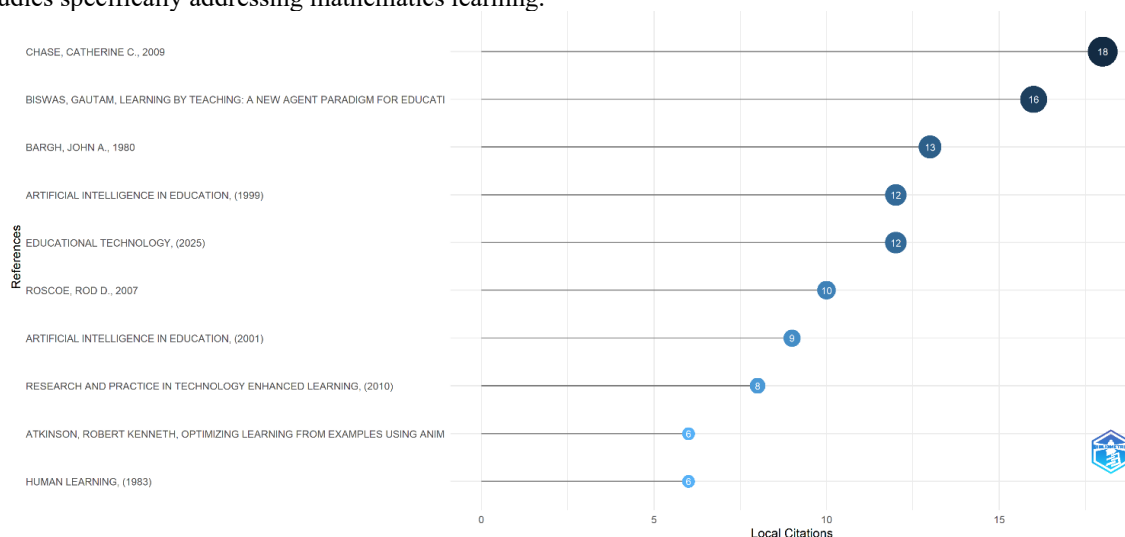


Figure 14. Most Local Cited References

Whilst Figures 12 and 13 identify the most frequently cited downstream articles, Figure 14 shifts the focus upstream that is, to the references that form the intellectual foundation of the field. The most locally cited references (Figure 14) reveal the field's intellectual bedrock. The three most-cited foundational works are conceptually coherent: Chase et al. (2009), which introduced the 'protégé effect'; Biswas et al. (2005), which articulated the 'learning by teaching' paradigm through Betty's Brain; and Bargh and Schul (1980), which supplied the underlying cognitive evidence that teaching benefits the teacher. Together they show the corpus is anchored in the learning-by-teaching/teachable-agent tradition directly consistent with the dominance of the Lund and Carnegie Mellon clusters (Figure 9). The lower-ranked entries (counts in Figure 14) comprise frequently-cited works on tutor learning and animated-agent worked examples. A methodological caveat applies to this figure: several high-frequency entries are recorded by source title rather than author (e.g., journal-level strings such as "Artificial Intelligence in Education" and "Educational Technology"), an artefact of incomplete first-author metadata in the cited-reference field; these should be disambiguated to author–year form before publication so the intellectual-base claim rests on unambiguous sources.

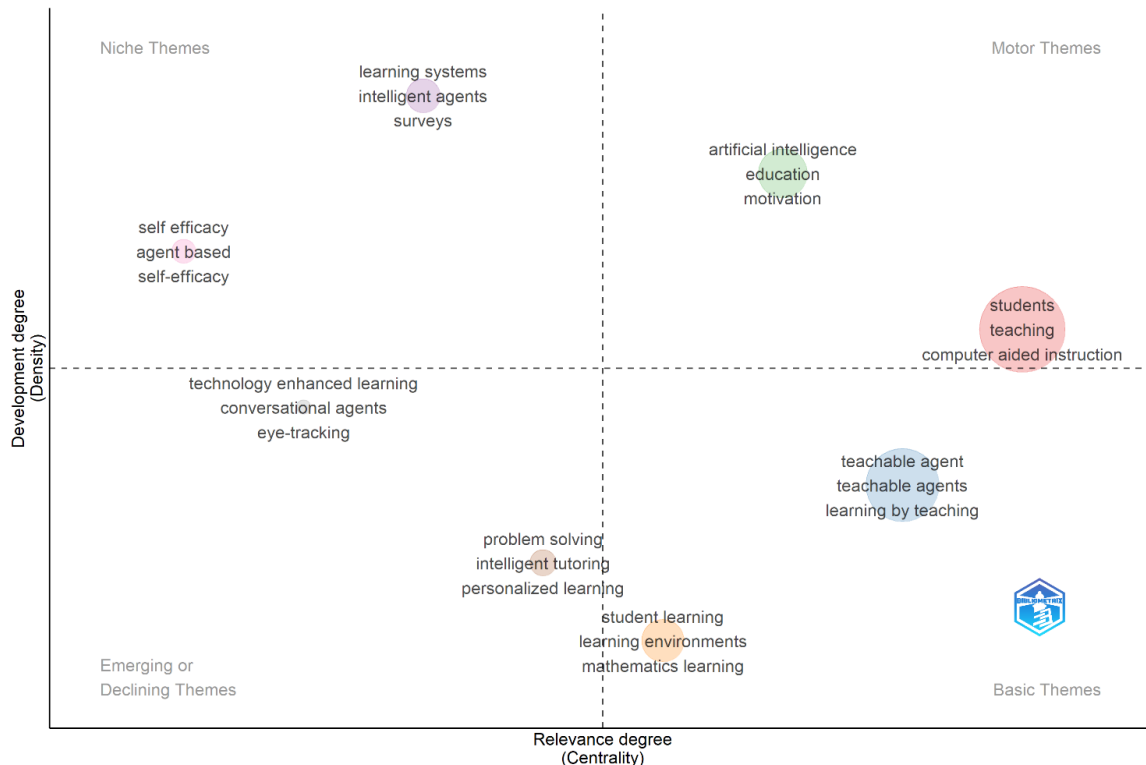
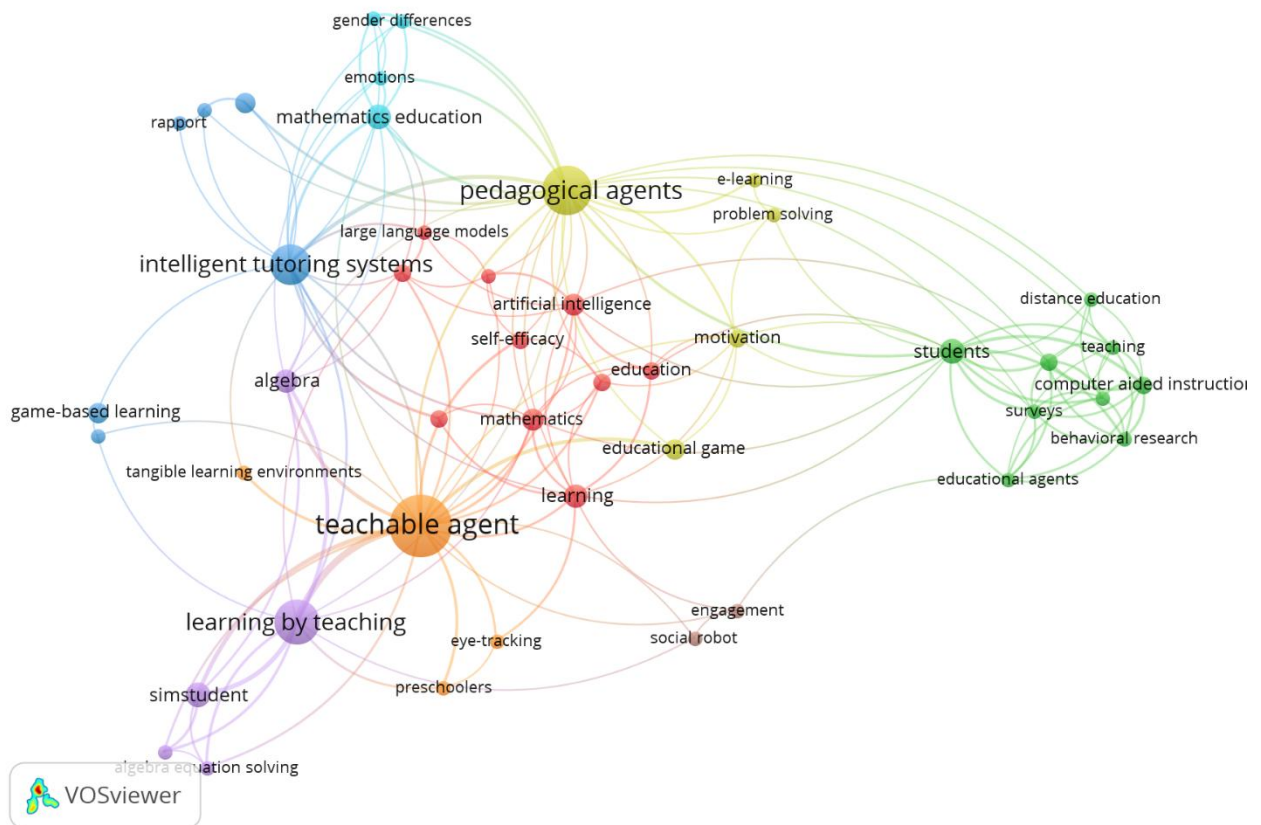


Figure 15. Thematic Map

Moving from a citation-based structure to a thematic structure, Figure 15 maps the positions of keyword clusters across two conceptual dimensions. The thematic map (Figure 15) locates keyword clusters by centrality (relevance to the field) and density (internal development), and its configuration is more revealing than the cluster inventory itself. The motor-theme quadrants mature and central is occupied not by mathematics but by generic instructional-technology and AI vocabulary ("students–teaching–computer-aided instruction" and "artificial intelligence–education–motivation"), indicating that what drives the field discursively is the agent-and-instruction apparatus rather than mathematical content. The field's signature paradigm, "teachable agent–learning by teaching", sits in the basic-themes quadrant central but low in density a foundational commitment not yet matured into a dense, self-sustaining research front, while isolated niche clusters ("learning systems–intelligent agents–surveys"; "self-efficacy–agent-based") remain peripheral. Most diagnostic is the position of mathematics itself: the cluster "student learning–learning environments–mathematics learning" falls on the basic/declining border at mid-centrality and low density, confirming at the thematic level what the source and keyword analyses imply mathematics learning is the field's application setting, not its conceptual engine. The newest, low-centrality material ("technology-enhanced learning–conversational agents–eye-tracking") sits at the emerging/niche margin, consistent with the still-incipient status of conversational-agent and LLM work. (Map constructed via the centrality–density co-word method of Cobo et al. (2011))

Table 1. Clusters Formed from the Results of Keyword Analysis Using VOSviewer



The thematic mapping using the Biblioshiny approach in Figure 15 needs to be validated and refined through network visualisation using different algorithms, as shown in Figure 16 and detailed in Table 1. Figure 16 displays a co-occurrence network visualisation of author keywords from the ‘pedagogical agents in mathematics education’ corpus, constructed using VOSviewer (van Eck & Waltman, 2010) with a minimum co-occurrence threshold of three, equivalent to approximately 2.3% of the total 133 articles. Node size reflects the frequency of keyword occurrence, line thickness illustrates the strength of co-occurrence between keyword pairs, and colour indicates membership of clusters generated by a modularity-based clustering algorithm. Eight thematic clusters were identified and are detailed in Table 1.

Cluster 1 (red) represents the convergence of AI-based pedagogical technology with personalised mathematics learning. The co-location of animated pedagogical agent (first-generation), artificial intelligence, large language models, and personalised learning in a single cluster suggests a subset of studies that deliberately bridge the classical pedagogical-agent paradigm and contemporary LLM technology (full keyword membership in Table 1).

Cluster 2 (green) groups the technical vocabulary of conventional learning systems computer-aided instruction, intelligent tutoring systems, distance education, and related terms (Table 1). Its relatively isolated position on the right-hand side of the network indicates that it forms the terminological foundation of the field's classic studies rather than an active contemporary front.

Cluster 3 (light blue) centres on game-based and interactive learning environments (Table 1). The analytically important feature is where 'mathematics learning' sits: in this game-based/interactive cluster rather than in a larger, more central one implying that mathematics is engaged more often through game-based and interactive designs than through the teachable-agent paradigm or generative-AI technologies.

Cluster 4 (yellow) gathers e-learning, motivation, and problem-solving terms around 'pedagogical agents' (Table 1). That 'pedagogical agents' is the largest node here rivalling 'teachable agent' indicates the term remains a central anchor of the field even though its cluster is no longer the largest by volume.

Cluster 5 (purple) represents the teachable-agent paradigm applied to specific mathematical content algebra and algebraic equation solving via SimStudent (Table 1). Its composition maps directly onto the Carnegie Mellon group (Matsuda and Koedinger) identified in Figure 8, tying this thematic cluster to an identifiable research lineage.

Cluster 6 (cyan/teal) is a small socio-emotional grouping emotions, gender differences, motivation and affect (Table 1). Its significance is that it situates 'mathematics education' in an affective frame distinct from the technical framing dominant elsewhere in the network, marking one of the few places where mathematics is tied to learner emotion rather than system architecture.

Cluster 7 (orange) combines teachable agents with eye-tracking, preschoolers, and tangible learning environments (Table 1) a configuration that reflects the Lund University group (Gulz, Haake, Sjöden) and its focus on teachable agents for young children.

Cluster 8 (brown/peach) is the smallest, pairing only 'engagement' and 'social robots' (Table 1). It marks a marginal but distinct direction: social robots as a physical alternative to screen-based pedagogical agents.

Three cross-cluster observations can be noted from Figure 16. Firstly, there are two largest nodes that act as conceptual hubs: 'pedagogical agents' (Cluster 4) and 'teachable agents' (Cluster 7), each of which attracts dense connections from almost all clusters. Second, the keywords 'mathematics' and 'math education' are distributed across three different clusters (Cluster 1, Cluster 3, and Cluster 6), implying a fragmentation of the mathematics learning context into several relatively distinct sub-approaches. Thirdly, the emergence of large language models in Cluster 1 confirms that LLM terminology already has sufficient co-occurrence to appear on the map, supporting previous observations regarding a paradigmatic transition during the 2024–2025 period (Figure 4).

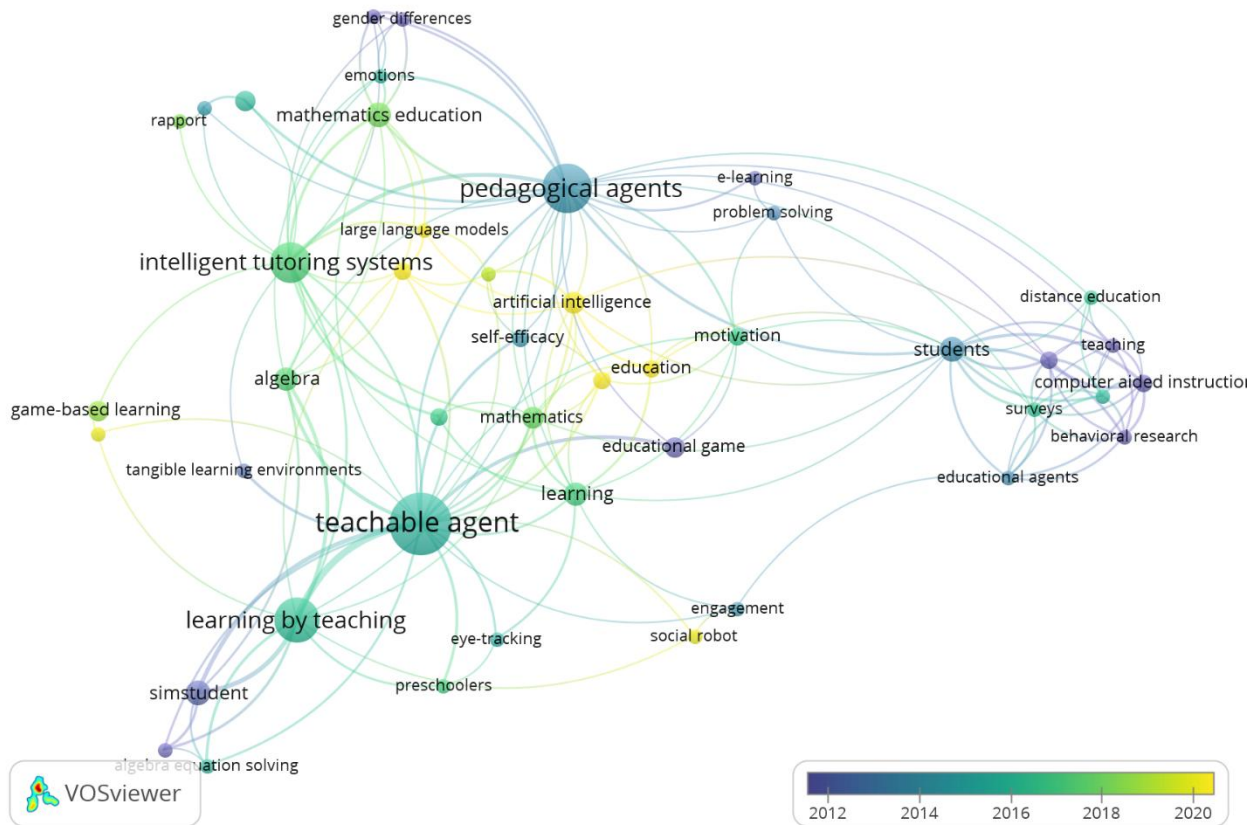


Figure 17. Overlay Visualisation

The network visualisation in Figure 16 depicts a static conceptual structure; the temporal dimension of this structure can be traced via the publication year overlay. Figure 17 presents the same network as Figure 16 but with a colour scheme based on the average year of publication of each author keyword's occurrence, ranging from dark purple (≈ 2012) to bright yellow (≈ 2020). Three temporal patterns can be identified from this map.

First, the bright yellow keywords (more recent, ≈ 2019 – 2020) include artificial intelligence, education, large language models, social robots, and game-based learning (in part). The concentration of yellow keywords around artificial intelligence and large language models indicates that discussions regarding AI and LLMs in the context of pedagogical agents are a relatively new theme within the corpus, consistent with the post-2022 surge in publications shown in Figure 4.

Secondly, the keywords highlighted in yellow-green (≈ 2017 – 2019) include teachable agents, pedagogical agents, intelligent tutoring systems, mathematics education, and algebra. This pattern indicates that the teachable agent and intelligent tutoring paradigms reached a peak in co-occurrence during the mid-to-late 2010s, prior to the emergence of the LLM wave.

Thirdly, the dark purple to blue keywords (≈ 2012 – 2015) include teaching, students, computer-aided instruction, gender differences, SimStudent, behavioural research, and educational games. Their status as legacy terms indicates that these core pedagogical terms and the contributions of SimStudent have historically been well-established in the corpus since the early stages of the observation period. Specifically or simstudent, its earlier temporal position (≈ 2012 – 2014) is consistent with the publication date of the seminal SimStudent paper by Matsuda et al. (2013), which was cited in the paragraph accompanying Figure 7.

DISCUSSION

Based on the results of the bibliometric mapping outlined in the previous section, five dimensions of discussion are presented here to synthesise the findings, situate the field within the broader scientific landscape, and draw theoretical and practical implications for future research.

The Dynamics of Growth and Maturity in the Field

The trajectory of scientific output in the field of pedagogical agents in mathematics education exhibits the characteristics of an emerging research frontier that is developing at a rate above the average for science in general. The annual growth rate of 11.25% (CAGR $\approx$ 11.3% when measured end-to-end) substantially exceeds the growth in global scientific publications reported by [NSB \(2019\)](#) of 4% for 2007–2017 and the estimate by [Bornmann and Mutz \(2015\)](#) of 8–9% for the modern phase of science up to 2012. This discrepancy is not merely an artefact of the small corpus size: the increase in the average annual citation count from $\approx$ 0.4 in the first decade to $\approx$ 2.0 in the last decade (Figure 5) indicates that volumetric growth is accompanied by a substantial increase in citation intensity. The combination of both high publication growth accompanied by an increase in age-normalised impact per article is a hallmark of a field in an active expansion phase, not stagnation.

However, the corpus's internationalisation profile suggests that this field has not yet fully moved beyond the 'nationally oriented' character commonly found in educational research. The international co-authorship rate of 19.55% is indeed comparable to the level of internationalisation in educational research reported by [Sezgin et al. \(2022\)](#), but remains below the global cross-disciplinary average of 25.7% in 2021 ([Aksnes & Sivertsen, 2023](#)). [Aman and Botte \(2017\)](#) explain this pattern as a consequence of the tradition of educational research that has historically been oriented towards the national context, where learning interventions are designed and tested within the domestic curriculum ecosystem. For the field of pedagogical agents in mathematics education, the practical consequence is that cross-context replicability and the generalisation of learning outcomes remain open empirical questions.

Institutional Concentration and Knowledge Production

Mapping of academic actors reveals a structure of knowledge production that is highly concentrated in three institutional clusters: Lund University (Sweden) with the Gulz–Haake–Sjödén–Pareto network, which developed a game-based teachable agent ([Pareto et al., 2012](#)); Carnegie Mellon University, with the Matsuda–Koedinger network, which developed SimStudent for algebra learning through 'learning by teaching' ([Matsuda et al., 2013](#)); and the University of Massachusetts Amherst, with the Arroyo–Woolf–Burleson–Muldner network, which developed Wayang Outpost as an Intelligent Tutoring System with affective components ([Arroyo et al., 2014](#)). The concentration of productivity within these three clusters aligns with a pattern frequently observed in niche fields, where a small proportion of research groups generate the majority of output a pattern consistent with Lotka's productivity law.

A comparison between productivity rankings (Figure 7) and local citation rankings (Figure 8) reveals an important dimension of influence. The Wayang Outpost group particularly Arroyo and Woolf dominates the local citation rankings (15 citations each) despite their absolute productivity ranking only fifth to eighth. Conversely, Matsuda, who ranks second in productivity, drops to ninth in local citations. This pattern indicates that intellectual impact is not proportional to the volume of output: the Wayang Outpost group has produced works that serve as the dominant internal references for this research community, whilst the contributions of SimStudent despite being productive are more widely absorbed by the learning sciences and artificial intelligence in education communities in general than by studies specifically focused on the PA $\times$ mathematics learning intersection. This asymmetry reinforces the argument by [Castro-Alonso et al. \(2021\)](#) that the field of pedagogical agents comprises internal sub-communities that utilise different citation bases despite sharing the same theoretical paradigm.

These structural claims, however, warrant three reflexive caveats. First, the apparent dominance of three clusters and of the United States and Sweden (Figures 9–10) is partly an artefact of the sampling frame rather than a complete census: Scopus over-represents English-language and STEM-leaning outlets and under-represents non-English and social-science venues ([Aksnes & Sivertsen, 2023](#); [Mongeon & Paul-Hus, 2016](#)), so the geography reported here is more accurately a map of visibility within Scopus's English-indexed window than of all underlying research activity. Second, the local- and global-citation counts on which the influence analysis rests index attention and visibility rather than validity or instructional quality, and in a small corpus they are amplified by cumulative-advantage dynamics in which already-visible works accrue further citations; the Wayang Outpost group's local-citation dominance therefore reflects internal canon formation as much as intrinsic merit ([Aksnes et al., 2019](#)). Third, bibliometric mapping characterises the structure and attention economy of this literature, not the effect sizes or methodological rigour of the individual studies within it; claims about what actually improves mathematics learning with pedagogical agents must rest on the primary meta-analytic evidence ([Castro-Alonso et al., 2021](#); [Schroeder et al., 2013](#)) rather than on citation

structure. Read with these caveats, the concentration documented here describes how attention and output are distributed in the indexed field, not a definitive ranking of quality or a global accounting of activity.

Continuity in Paradigm Shift: from Teachable Agents to Large Language Models

The analysis of the intellectual base in Figure 14 confirms that this field is rooted in the ‘learning by teaching’ paradigm, which is underpinned by three conceptual pillars: the cognitive principle of the benefits of teaching (Bargh & Schul, 1980), its implementation in educational software based on teachable agents (Biswas et al., 2005), and empirical documentation of the protégé effect as a motivational mechanism (Chase et al., 2009). The fact that Bargh and Schul’s 1980 educational psychology paper remains a local reference with 13 citations in the 2025 corpus demonstrates remarkable paradigmatic continuity: the fundamental concept that students learn better when they become teachers has remained the theoretical backbone of this field for nearly half a century.

This continuity, however, is now encountering a second wave of paradigmatic transformation triggered by the emergence of Large Language Models. Recent empirical evidence particularly from Pardos and Bhandari (2024), who compared the learning effects of ChatGPT-generated hints with those of human tutor-authored hints in algebra and statistics shows that LLM-based pedagogical agents can generate learning gains comparable to those of human tutors for specific algebraic topics. The sharp surge in scientific output in 2024–2025 identified in Figure 4 and the emergence of large language models in Cluster 1 of the VOSviewer visualisation (Figure 16) as entries converging with first-generation animated pedagogical agents indicate that a transition is underway not a discontinuity, but an accumulation: researchers are not abandoning the ‘learning by teaching’ and ‘teachable agent’ paradigms, but rather integrating LLM capabilities into an established framework. This aligns with the retrospective reflections of Johnson and Lester (2016), who state that pedagogical agents have entered their third decade as a productive and continuously evolving research paradigm.

The Anomalous Position of Mathematics Learning in the Field

One of the most counter-intuitive findings of this mapping is the marginal position of mathematics learning within the thematic structure of a field explicitly named ‘pedagogical agents in mathematics education’. In the thematic map shown in Figure 15, the cluster containing ‘mathematics learning’ is situated on the boundary between Basic Themes and Emerging/Declining Themes rather than at the centre of the Motor Themes quadrant. In the VOSviewer network visualisation in Figure 16, the terms “mathematics”, “maths education”, and “mathematics learning” are scattered across three different clusters (Clusters 1, 3, and 6), without forming a single consolidated cluster. Meanwhile, the dominant Motor Themes actually contain general teaching terminology (“students–teaching–computer-aided instruction”) and AI (“artificial intelligence–education–motivation”).

The most likely interpretation of this pattern is that the field under study is in fact stronger as a study of pedagogical technology than as a study of mathematics learning per se. Mathematics education functions as an application context and testing ground, rather than as an epistemic core driving the research agenda. This aligns with the findings of Castro-Alonso et al. (2021) in a meta-analysis of the effectiveness of multimedia pedagogical agents, which showed that the proportion of studies in the mathematics domain is relatively small compared to the science or language domains. The methodological consequence is that mathematics education researchers wishing to enter this field are actually entering an open space for contribution: they can situate questions characteristic of mathematics education such as conceptual understanding, representational fluency, mathematical discourse, and equity within the pedagogical agent paradigm, rather than merely treating mathematics as a content delivery vehicle.

Implications

A synthesis of the above bibliometric findings yields four noteworthy implications. Firstly, for researchers new to the field particularly those from the Global South, including Indonesia, which does not appear in Figures 10–11 this research area presents a rare combination: a field with high growth momentum (CAGR 11.3%), a stable intellectual foundation, and conceptual space that remains open at the intersection of generative AI and mathematics education. Second, for educational technology developers, the finding that teachable agents and the ‘learning by teaching’ paradigm remain dominant suggests that the integration of LLMs into educational products should not abandon proven conceptual architectures, but rather reinforce them. Thirdly, for research policymakers, the extreme institutional concentration (three universities dominate the entire field) indicates a need to broaden the contributor base through international partnerships and cross-border funding. Fourthly, for the mathematics education community, this mapping serves as a call to more actively position the distinctive agenda of mathematics education within the dialogue with the learning sciences and AI in education communities.

Limitations and Future Research Agenda

Several limitations bound these findings. The corpus is Scopus-only, so work indexed solely in Web of Science, ERIC, IEEE Xplore, the ACM Digital Library, or regional databases may be absent; because Scopus over-represents English-language and STEM outlets (Aksnes & Sivertsen, 2023; Mongeon & Paul-Hus, 2016), research from non-Anglophone contexts including Indonesia is likely under-represented (Aman & Botte, 2017), and, given the field's reliance on proceedings (AIED, ITS, CHI; Figure 6), some intellectually significant conference papers may be undersampled.

With only 133 documents, annual metrics particularly citation intensity (Figure 5) and local-citation rankings (Figures 8 and 13) are sensitive to small-number noise; the swing from ≈ 2.9 (2014) to ≈ 0.6 (2015) in Figure 5 is largely such noise rather than a real shift. The 2024–2025 output is further deflated by Scopus indexing lag (typically 6–12 months), so the ≈ 16 articles recorded for 2025 will likely rise on re-extraction, while the 2024 citation-intensity peak (≈ 3.3) is inflated by a short citation window (Bormmann & Mutz, 2015).

Finally, residual metadata artefacts duplicate Lund University entries ('Lund University Cognitive Science' vs. 'Lunds Universitet'; Figure 9), the ampersand-encoding of 'Texas A&M', an orphan 'School of Computer Science', split Lecture Notes in Computer Science entries (Figure 6), and the fragmentation of mathematics-related keywords across clusters (Figure 16) indicate that affiliation and source disambiguation and a keyword thesaurus should be applied more rigorously in future work.

Five directions regarding future research agenda follow. First, multi-database extraction (Scopus + Web of Science + IEEE), ideally within a hybrid bibliometric–systematic-review design, would reduce coverage bias and pair structural mapping with qualitative appraisal of core works. Second, systematic empirical study of LLM-based pedagogical agents is needed across specific mathematical sub-domains (geometry, calculus, statistics) and learner profiles, grounded in mathematics-education theory (e.g., APOS, RME) and extending the algebra-focused evidence of Pardos and Bhandari (2024). Third, cross-regional collaboration with Southeast Asia, Africa, and Latin America would test whether the 'teachable agent' and 'protégé effect' (Chase et al., 2009) generalise beyond Anglophone settings or are culturally contingent. Fourth, because the field's foundational studies (Bargh & Schul, 1980; Biswas et al., 2005) are predominantly short-term, multi-year longitudinal designs are needed to assess effects on mathematical identity and STEM persistence. Fifth, co-citation and thematic-evolution analyses (Cobo et al., 2011) would refine the intellectual map developed here.

CONCLUSION

This study maps the landscape of the scientific literature on pedagogical agents in mathematics education based on 133 Scopus-indexed documents from the period 1999–2025, with the aim of addressing five research questions relating to production dynamics, academic actors, geographical distribution, intellectual base, and thematic structure.

Across the five research questions, a coherent portrait emerges. The field is in active expansion (CAGR $\approx 11.3\%$, with citation intensity rising roughly fivefold across decades) yet structurally concentrated three institutions (Lund, Carnegie Mellon, UMass Amherst) and a CS-proceedings publication culture dominate, while the United States and Sweden account for most citations within a still nationally oriented collaboration profile (19.55% international). Its intellectual base is strikingly stable the 'learning by teaching', teachable-agent, and protégé-effect lineage now absorbing rather than being displaced by LLM-based agents. The defining paradox is that mathematics learning, though it names the field, sits at its thematic margin, functioning as an application context rather than an epistemic core.

This study makes two contributions. Substantively, it provides the first bibliometric map of the pedagogical-agents $\times$ mathematics-education intersection and identifies its central dynamic continuity-in-shift from teachable agents to LLM-based agents. Methodologically, it shows how triangulating Biblioshiny and VOSviewer yields a more robust structural reading than either tool alone. These conclusions are bounded by a single-database, English-leaning corpus of modest size; we therefore recommend (i) multi-database replication, (ii) domain-specific empirical work on LLM-based agents in mathematics, and (iii) longitudinal evaluation. For mathematics-education researchers especially in the Global South—the field offers a rare opening: to move mathematics learning from application context to the epistemic core of the next generation of pedagogical-agent research.

REFERENCE

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1), 1–17. <https://doi.org/10.1177/2158244019829575>
- Aksnes, D. W., & Sivertsen, G. (2023). Global trends in international research collaboration, 1980–2021. *Journal of Data and Information Science*, 8(2), 26–42. <https://doi.org/10.2478/jdis-2023-0015>

-
- Aman, V., & Botte, A. (2017). A bibliometric view on the internationalization of European educational research. *European Educational Research Journal*, 16(6), 843–868. <https://doi.org/10.1177/1474904117729903>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective gendered learning companions. In *Artificial intelligence in education* (pp. 41–48). IOS Press.
- Arroyo, I., Woolf, B. P., Cooper, D. G., Burleson, W., & Muldner, K. (2011). The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In *2011 IEEE 11th international conference on advanced learning technologies* (pp. 506–510). IEEE. <https://doi.org/10.1109/ICALT.2011.157>
- Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology*, 105(4), 957–969. <https://doi.org/10.1037/a0032748>
- Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4), 387–426. <https://doi.org/10.1007/s40593-014-0023-y>
- Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72(5), 593–604. <https://doi.org/10.1037/0022-0663.72.5.593>
- Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(2), 95–115. [https://doi.org/10.3233/IRG-2005-15\(2\)02](https://doi.org/10.3233/IRG-2005-15(2)02)
- Biswas, G., Leelawong, K., Schwartz, D., & Vye, N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3–4), 363–392. <https://doi.org/10.1080/08839510590910200>
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222. <https://doi.org/10.1002/asi.23329>
- Bringula, R. P., Fosgate Jr, I. C. O., Garcia, N. P. R., & Yorobe, J. L. M. (2018). Effects of pedagogical agents on students' mathematics performance: A comparison between two versions. *Journal of Educational Computing Research*, 56(5), 701–722. <https://doi.org/10.1177/0735633117722494>
- Büyükkıdık, S. (2022). A bibliometric analysis: a tutorial for the bibliometrix package in R using IRT literature. *Journal of Measurement and Evaluation in Education and Psychology*, 13(3), 164–193. <https://doi.org/10.21031/epod.1069307>
- Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review*, 33(3), 989–1015. <https://doi.org/10.1007/s10648-020-09587-1>
- Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4), 334–352. <https://doi.org/10.1007/s10956-009-9180-4>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Daulay, L. A., Nasution, A. K. P., Asnawi, A., Zahari, C. L., & Wardani, H. (2025). Bibliometric study on research trends in artificial intelligence and mathematics education. *Alifmatika: Jurnal Pendidikan dan Pembelajaran Matematika*, 7(1), 124–147. <https://doi.org/10.35316/alifmatika.2025.v7i1.124-147>
- 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>
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1), 35–51. [https://doi.org/10.1016/S1389-0417\(99\)00005-4](https://doi.org/10.1016/S1389-0417(99)00005-4)
- Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a teachable agent with a social conversation module: Effects on student experiences and learning. In *Proceedings of International Conference on Artificial Intelligence in Education* (pp. 106–114). Springer. https://doi.org/10.1007/978-3-642-21869-9_16
- Johnson, W. L., & Rickel, J. (1997). Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin*, 8(1–4), 16–21. <https://doi.org/10.1145/272874.272877>
-

-
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26(1), 25–36. <https://doi.org/10.1007/s40593-015-0065-9>
- Johnson, W. L., & Lester, J. C. (2018). Pedagogical agents: back to the future. *AI Magazine*, 39(2), 33–44. <https://doi.org/10.1609/aimag.v39i2.2793>
- Kim, Y., & Wei, Q. (2011). The impact of learner attributes and learner choice in an agent-based environment. *Computers & Education*, 56(2), 505–514. <https://doi.org/10.1016/j.compedu.2010.09.016>
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. <https://doi.org/10.1007/s10648-007-9049-0>
- Krämer, N. C., Karacora, B., Lucas, G., Dehghani, M., Rütther, G., & Gratch, J. (2016). Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Computers & Education*, 99, 1–13. <https://doi.org/10.1016/j.compedu.2016.04.002>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 359–366). Association for Computing Machinery. <https://doi.org/10.1145/258549.258797>
- Li, M. (2025). Integrating artificial intelligence in primary mathematics education: Investigating internal and external influences on teacher adoption. *International Journal of Science and Mathematics Education*, 23(5), 1283–1308. <https://doi.org/10.1007/s10763-024-10515-w>
- Liu, L., Yan, S., Deng, X., & Yu, Z. (2024). Bibliometric analyses of the impact of pedagogical agents on learning. *International Journal of Mobile and Blended Learning*, 16(1), 1–19. <https://doi.org/10.4018/IJMBL.346976>
- Lyu, B., Li, C., Li, H., Oh, H., Song, Y., Zhu, W., & Xing, W. (2025a). Exploring the role of teachable AI agents' personality traits in shaping student interaction and learning in mathematics education. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 887–894). <https://doi.org/10.1145/3706468.3706532>
- Lyu, B., Li, C., Li, H., Oh, H., Song, Y., Zhu, W., & Xing, W. (2025b). The role of teachable agents' personality traits on student-AI interactions and math learning. *Computers & Education*, 234, Article 105314. <https://doi.org/10.1016/j.compedu.2025.105314>
- Maass, K., & Engeln, K. (2019). Professional development on connections to the world of work in mathematics and science education. *ZDM*, 51(6), 967–978. <https://doi.org/10.1007/s11858-019-01047-7>
- Martha, A. S. D., & Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, 16(1), Article n1. <https://doi.org/10.9743/jeo.2019.16.1.8>
- Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Cohen, W. W., Stylianides, G. J., & Koedinger, K. R. (2013). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology*, 105(4), 1152–1163. <https://doi.org/10.1037/a0031955>
- Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., Cohen, W. W., & Koedinger, K. R. (2011). Learning by teaching SimStudent: An initial classroom baseline study comparing with Cognitive Tutor. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 213–221). Springer. https://doi.org/10.1007/978-3-642-21869-9_29
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- National Science Board. (2019). *Publications output: U.S. trends and international comparisons* (Science and Engineering Indicators 2020, NSB-2020-6). National Science Foundation. <https://nces.nsf.gov/pubs/nsb20206/>
- Nguyen, D. T., & Pham, Q. V. (2025). The evolving landscape of AI integration in mathematics education: A systematic review of trends (2015–2025). *Eurasia Journal of Mathematics, Science and Technology Education*, 21(10), Article em2714. <https://doi.org/10.29333/ejmste/17078>
- Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, 19(5), e0304013. <https://doi.org/10.1371/journal.pone.0304013>
-

-
- Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education*, 24(3), 251-283. <https://doi.org/10.1007/s40593-014-0018-8>
- Pareto, L., Arvemo, T., Dahl, Y., Haake, M., & Gulz, A. (2011). A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 247-255). Springer. https://doi.org/10.1007/978-3-642-21869-9_33
- Pareto, L., Haake, M., Lindström, P., Sjöden, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development*, 60(5), 723–751. <https://doi.org/10.1007/s11423-012-9246-5>
- Samuel, T., Chen, H. L., & Yohannes, A. (2025). Robot-assisted mathematics education: A systematic literature review. *Journal of Computer Assisted Learning*, 41(4), Article e70095. <https://doi.org/10.1111/jcal.70095>
- Schlimbach, R., Rinn, H., Markgraf, D., & Robra-Bissantz, S. (2022). A literature review on pedagogical conversational agent adaptation. In *Proceedings of the Pacific Asia Conference on Information Systems* (Vol. 1, pp. 1-14). Association for Information Systems. https://www.researchgate.net/publication/361228223_A_Literature_Review_on_Pedagogical_Conversational_Agent_Adaptation
- Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research*, 49(1), 1–39. <https://doi.org/10.2190/EC.49.1.a>
- Sezgin, A., Orbay, K., & Orbay, M. (2022). Educational research review from diverse perspectives: A bibliometric analysis of Web of Science (2011–2020). *SAGE Open*, 12(4). <https://doi.org/10.1177/21582440221141628>
- Sjöden, B., Tärning, B., Pareto, L., & Gulz, A. (2011). Transferring teaching to testing—an unexplored aspect of teachable agents. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 337-344). Springer. https://doi.org/10.1007/978-3-642-21869-9_44
- Subroto, P. W., Malik, M., Raditya, A., & Saputra, N. N. (2024). A bibliometric analysis on artificial intelligence in mathematics education. *Journal of Research and Advances in Mathematics Education*, 9(1), 1-15. <https://doi.org/10.23917/jramathedu.v9i1.2429>
- Tao, Y., Zhang, G., Zhang, D., Wang, F., Zhou, Y., & Xu, T. (2022). Exploring persona characteristics in learning: A review study of pedagogical agents. *Procedia computer science*, 201, 87-94. <https://doi.org/10.1016/j.procs.2022.03.014>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Lewin, S., Godfrey, C. M., Macdonald, M. T., Langlois, E. V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., & Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- 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>
- Wei, C. W., Hung, I., Lee, L., & Chen, N. S. (2011). A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *Turkish Online Journal of Educational Technology-TOJET*, 10(2), 11-23.
- Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R., & Christopherson, R. M. (2010). The effect of motivational learning companions on low achieving students and students with disabilities. In *International conference on intelligent tutoring systems* (pp. 327-337). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13388-6_37
- Wulandari, T., Hasanah, R. U., & Fadillah, D. T. (2024). Kemampuan pembuktian matematis mahasiswa: Systematic literature review. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 1(3), 350-359.
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- Yıldız, S. G., & Körpeoğlu, S. G. (2025). Trends and insights of AI in mathematics education: A bibliometric analysis. *International Electronic Journal of Mathematics Education*, 20(3), Article em0837. <https://doi.org/10.29333/iejme/16401>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. <https://doi.org/10.1177/1094428114562629>
-

ACKNOWLEDGEMENT

The authors would like to express their gratitude and appreciation to the Indonesian Endowment Fund for Education (LPDP) under the Ministry of Finance of the Republic of Indonesia for sponsoring and providing assistance for the realisation of this work.

AUTHOR CONTRIBUTION STATEMENT

The contributions made by the authors to the preparation of this document are as follows: (1) Fadhil Zil Ikram contributed to the conceptualisation, methodology, data curation, formal analysis, visualisation and writing of the manuscript; and (2) Kusnandi contributed to the conceptualisation, writing of the article and supervision.

AI DISCLOSURE STATEMENT

During the preparation of this work, the author(s) used DeepL and Claude Pro to enhance the the readability and language of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Fadhil Zil Ikram

Department of Mathematics Education Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia,

Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari, Bandung, Jawa Barat, Indonesia

Email: dhilikram@upi.edu

***Kusnandi (Corresponding Author)**

Department of Mathematics Education Faculty of Mathematics and Science Education
Universitas Pendidikan Indonesia,

Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari, Bandung, Jawa Barat, Indonesia

Email: kusnandi@upi.edu
