e-ISSN: 2710-2122, p-ISSN: 2710-2114
2025: Vol 7(2), pp 1-22. https://doi.org/10.38140/ijer-2025.vol7.2.06
Interdisciplinary Journal of Education Research

Learning Analytics in Computer Programming Education: A Bibliometric Scoping Review

Kelibone Eva Mamabolo[1]
Siyabonga Mhlongo[2]
Moeketsi Mosia[3]
Hossana Twinomurinzi[4]

AFFILIATIONS

[1]College of Business & Economics, University of Johannesburg, Johannesburg, South Africa.
[1]Department of Computer Science and Information Technology, Sol Plaatje University, Kimberley, South Africa.
[2&4]College of Business & Economics, University of Johannesburg, Johannesburg, South Africa.
[3]Department of Mathematics, Natural Sciences and Technology Education, University of the Free State, Bloemfontein, South Africa.

CORRESPONDENCE

Email: eva.mamabolo@spu.ac.za*

EDITORIAL DATES

Received: 11 March 2025
Revised: 15 June 2025
Accepted: 01 July 2025
Published: 17 August 2025

DOI: 10.38140/ijer-2025.vol7.2.06

Abstract: There are often high failure rates and student attrition in programming education due to challenges with syntax, debugging, and abstract concepts. Traditional teaching approaches have struggled to meet the diverse learning needs of students. This paper presents a scoping review incorporating bibliometric analysis that examines Learning Analytics (LA) research in programming education within Computer Science, Engineering, and Mathematics. The study identifies thematic trends, research gaps, and instructional implications. A bibliometric scoping review was conducted on documents published from 2014 to 2023, retrieved from Scopus and Web of Science. After screening, 1,208 documents were analysed. The review reveals a growing focus on data mining, predictive modelling, and student-centred learning. Most research outputs emerge from Europe and North America, while Africa shows a growing contribution. However, programming-specific applications such as debugging and formative feedback remain underexplored. The study highlights the limited integration of learning theories in LA applications. It also suggests that aligning LA with frameworks like cognitive load theory can foster personalised learning, enhance engagement, and support skill acquisition. These findings provide evidence-based insights to guide instructional innovation, research collaboration, and the development of adaptive programming education systems.

Keywords: Learning analytics, programming education, engineering, mathematics, data mining, bibliometric analysis.

1. Introduction

The challenges of learning programming are widely acknowledged, particularly for students with minimal academic backgrounds who struggle in large, instructor-paced courses. Those lacking self-regulated learning skills often exhibit inattentive behaviour, resulting in low performance and dropout (Utamachant et al., 2023). This difficulty is especially evident for students transitioning to a new academic curriculum, with high failure rates in introductory programming courses being a major concern (Utamachant et al., 2023). Scholars in computer science education have extensively investigated these issues, employing early interventions to reduce student attrition (Obaido et al., 2023).

Programming requires persistence, strategic thinking, and strong problem-solving skills, extending beyond syntax mastery. Students face challenges with syntax, debugging, and skill development, all of which are fundamental to achieving programming proficiency (Moon et al., 2020). Debugging, a critical skill, remains one of the most challenging aspects of programming and is essential for problem-solving and code efficiency (Venigalla & Chimalakonda, 2020). Programming exercises improve computational thinking and language capabilities, which are vital for solving real-world

How to cite this article:
Mamabolo, K. E., Mhlongo, S., Mosia, M., & Twinomurinzi, H. (2025). Learning analytics in computer programming education: A bibliometric scoping review. *Interdisciplinary Journal of Education Research, 7*(2), a06. https://doi.org/10.38140/ijer-2025.vol7.2.06

problems (Moon et al., 2020; Utamachant et al., 2023). However, tackling complex problems necessitates metacognitive skills that extend beyond syntax and flow (Bosse & Gerosa, 2017).

Traditional teaching methods, which rely on static resources like textbooks and presentations, are increasingly inadequate for programming education (Zhang et al., 2019). Pedagogically, programming languages should be simplified to prevent negative impacts on students' motivation (Qian & Lehman, 2017). Despite efforts to simplify programming instruction, traditional methods continue to challenge educators and students, as evidenced by consistently high failure rates across courses (Asai et al., 2019). The discussion above highlights programming's cognitive demands, requiring extensive knowledge absorption. While flipped classrooms and active learning tools have been adopted, they account for only a fraction of educational resources and often overlook diverse learning needs (Giannakos et al., 2016). Learning Analytics (LA) offers promising pedagogical applications through theory-based research; however, its use in programming education remains limited, despite its potential to address cognitive challenges and optimise resource utilisation. This underutilisation stems from an over-reliance on data-driven approaches and a lack of integration with learning theories to better interpret student behaviour and performance (Wang et al., 2022). While some studies employ LA for real-time feedback and course design (Yu et al., 2023; Eloy et al., 2022), most focus on self-regulated learning and social constructivism, leaving gaps in programming education applications. Additionally, LA's role in debugging and problem-solving remains underexplored.

Against this background, there is a critical need to integrate learning theories into LA research to better understand how students acquire programming skills. This paper addresses this gap through a bibliometric review of LA research in programming education within Computer Science, Engineering, and Mathematics. It maps existing scholarship, identifies research gaps, and answers the two guiding questions below.

- What are the emerging trends and research gaps in LA within programming education?
- How can these insights inform effective teaching strategies and educational interventions?

The review analyses publication patterns, citation metrics, and thematic areas, highlighting key contributors, institutional affiliations, and collaborative networks.

2. Materials and Methods

This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) framework. It adopts a two-step approach: a scoping review and a bibliometric analysis, as outlined in Figure 1. PRISMA-ScR guides transparent and systematic synthesis through structured stages of identification, screening, and inclusion (Tricco et al., 2018; Agrawal et al., 2024).

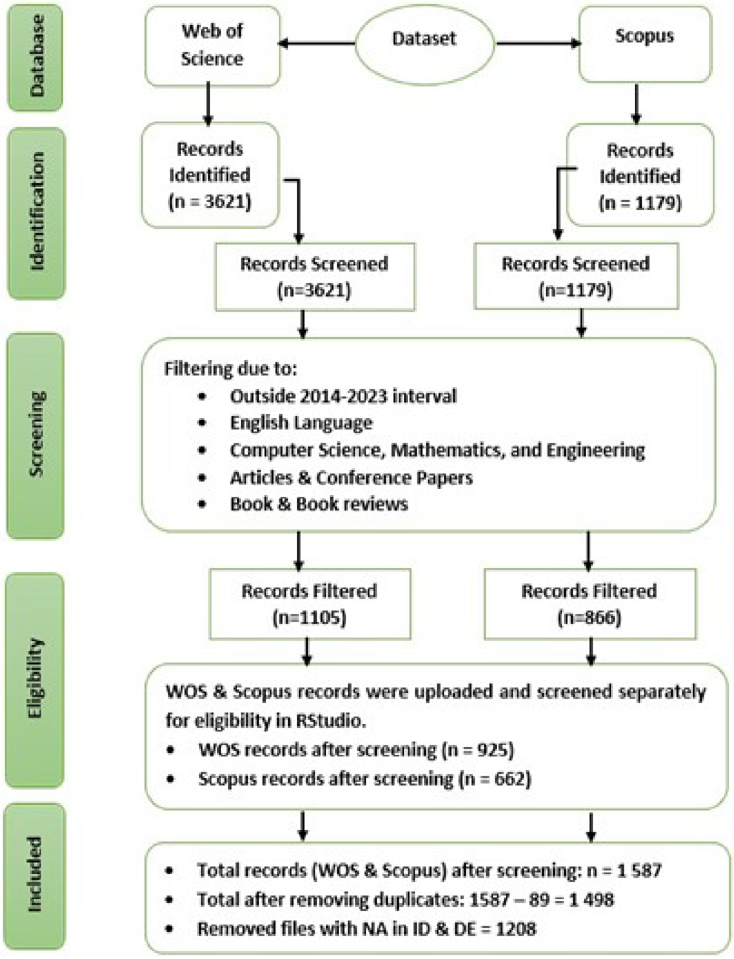

Figure 1: *Process flowchart*

Step 1: "The Arksey and O'Malley framework guided the scoping review. This method helps identify research gaps while ensuring rigour and quality (Pirri et al., 2020)."

Step 2: "Bibliometric analysis provided a comprehensive view of Learning Analytics (LA) research in Computer Science, Engineering, and Mathematics. This method quantifies scientific literature to assess research expansion, publication trends, and citation impact (Muhuri et al., 2019; Okumuş Dağdeler, 2023)."

2.1 Development of research objectives

This study maps LA research in Computer Science, Engineering, and Mathematics related to programming education. It aims to identify research gaps and analyse publications from 2014 to 2023. Using bibliometric analysis, the study explores research trends and identifies influential scholars, institutions, and key themes.

2.1.1 Framework stage: Identification of databases, relevant studies, and dataset pre-processing

The study employed Scopus and Web of Science (WOS) as primary data sources. The accurate merging of these databases is essential to mitigate errors that could adversely affect bibliometric analysis (Echchakoui, 2020). These databases were selected due to their comprehensive indexing of peer-reviewed literature across STEM disciplines, including Computer Science and Engineering. Both databases are widely acknowledged for their compatibility with bibliometric tools and their extensive coverage of interdisciplinary publications. The integration process adhered to the four-step PRISMA method (Echchakoui, 2020; Ogundeji & Okolie, 2022).

PRISMA encompasses four selection stages: identification, screening, eligibility, and inclusion (Pirri et al., 2020). The search string utilised to retrieve data from WOS and Scopus was: ("Educational Data Mining"), ("Learning Analytics"), or ("Online Learning Environment") AND (Teaching OR Learning) AND (Programming OR Coding OR "Computer Science" OR "CS"). Following the identification stage, 3,621 publications from WOS and 1,179 from Scopus were retrieved. Subsequent filtering reduced the dataset to 1,105 entries from WOS and 866 from Scopus (Figure 1). These entries were further screened in RStudio based on specified inclusion and exclusion criteria (Table 1).

***Table 1:** Inclusion and exclusion criteria*

Select papers based on inclusion and exclusion criteria	
Inclusion Criteria (IC)	**Exclusion Criteria (EC)**
(IC1) Papers on LA in Computer Science, Engineering, and Mathematics	(EC1) Lecture Notes (90 records removed)
(IC2) Papers written in English	(EC2) Articles on ethics, policy, or industry-related LA
	(EC3) LA reviews, posters, discussions, and workshop
	(EC4) Papers unrelated to CS, Engineering, or Mathematics (e.g., marketing, finance, medical analysi

This study used the Bibliometrix R package for bibliometric mapping and Biblioshiny to measure the impact of LA. These tools analysed keyword trends, research themes, and collaboration networks in Computer Science, Engineering, and Mathematics.

2.1.2 Dataset synthesis

Table 2 summarises LA research in Computer Science, Engineering, and Mathematics from 2014 to 2023. Research activity grew from 2015 onwards, with 1,208 documents from 410 sources, a 4.28% annual growth rate, and an average of 11.02 citations per paper. A total of 3,277 authors contributed, with 80% of the papers being single-authored, 3.82% co-authored, and 16.56% involving international collaboration. The dataset includes 745 journal articles, 451 conference papers, and a few books, reviews, and proceedings papers. Additionally, 3,095 Keywords Plus and 3,088 author-defined keywords highlight the focus of the research.

***Table 2:** Primary information on data synthesis and dataset summary*

Description	**Results**
Timespan	2014:2023
Sources (Journals, Books, etc.)	410
Documents	1208
Annual Growth Rate %	4,28
Document Average Age	3,75
Average citations per doc	11,02
References	39625
DOCUMENT CONTENTS	
Keywords Plus (ID)	3095
Author's Keywords (DE)	3088
AUTHORS	
Authors	3277
Authors of single-authored docs	76
AUTHORS COLLABORATION	

Single-authored docs	80
Co-Authors per Doc	3,82
International co-authorships %	16,56
DOCUMENT TYPES	
article	745
article: proceedings paper	3
book	1
book chapter	4
conference paper	451
review	4

3. Results and Discussions

3.1 Growth and trends of learning analytics

This section analyses the annual scientific production of LA articles in computer science, engineering, and mathematics from 2014 to 2023. Figure 2 illustrates the yearly publication count, a key measure of research output. Publications increased steadily from 2014 to 2019, with notable growth between 2016 (79 articles) and 2018 (117 articles). This period aligns with broader institutional uptake and research diversification in LA, particularly within higher education (Viberg et al., 2018). A sharper rise occurred between 2019 and 2020, followed by a slight decline in 2021, which may reflect evolving thematic priorities or project-based publication cycles within the field.

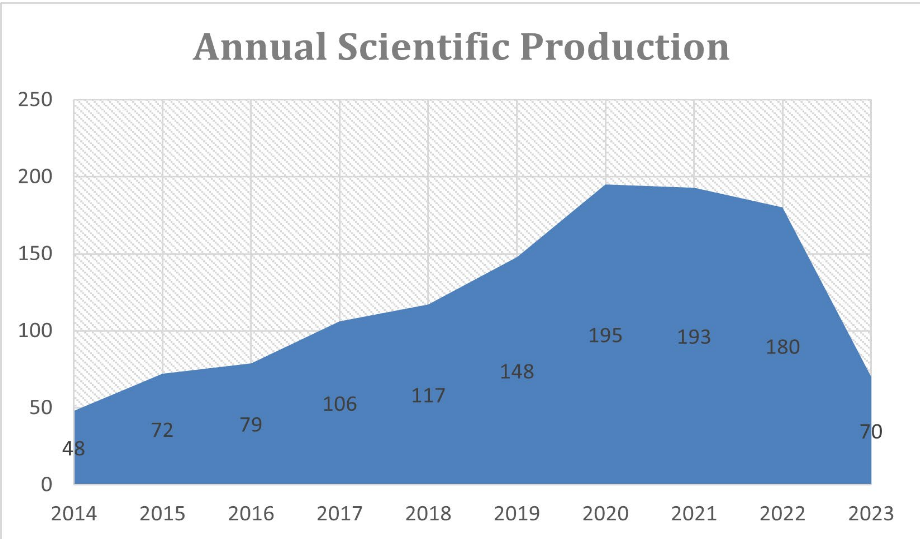

Figure 2: *Annual Scientific Production*

Figure 3 shows the annual total citations from 2014 to 2023. MeanTCperYear, which measures total publications and average citations per year, provides insights into publication impact (Junjia et al., 2023). Citations declined in 2018 and dropped significantly during the 2018–2020 period. While research volume increased, the gap between output and citation impact widened. Scientific publications typically gain influence over time, peaking after a delay (Ioannidis et al., 2022). However, pandemic-related disruptions (Rodrigues et al., 2020) affected research dissemination while simultaneously accelerating Learning Analytics (LA) due to the rise of emergency remote learning. This transition led to an increased reliance on educational technologies and digital trace data, driving renewed interest and urgency in LA. This trend extended to LA, where emergency remote teaching environments generated large volumes of learner data, prompting a surge in LA studies focused on digital engagement and educational data mining. In contrast, studies focusing on

COVID-19 dominated citations, with 98 of the 100 most-cited papers centred on the pandemic, reflecting a shift in research priorities (Rodrigues et al., 2020).

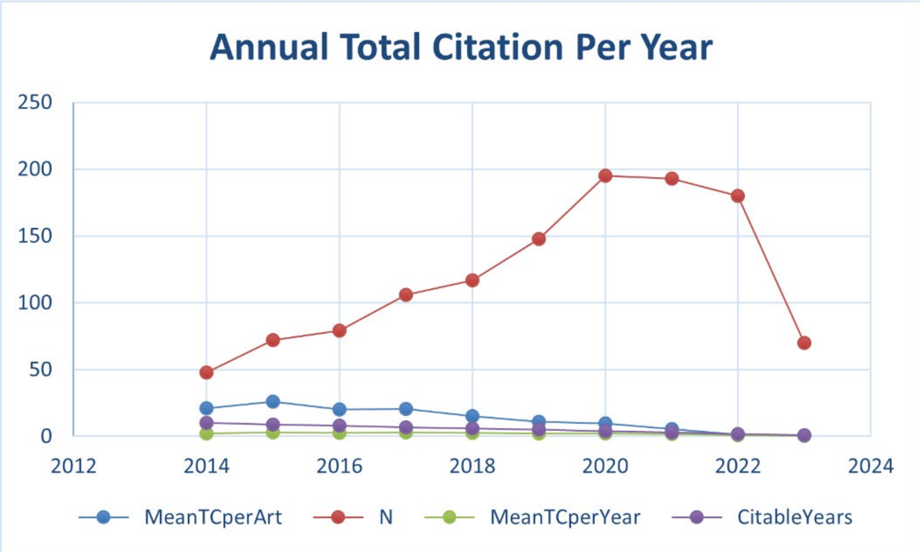

Figure 3: Annual total citation per year

3.2 Relevant sources and documents of learning analytics

Figure 4 highlights the top 20 publishing platforms for LA research in computer science, engineering, and mathematics. The ACM International Conference Proceeding Series leads with 121 articles, followed by Computers & Education (60 papers), IEEE Access (59), and IEEE Transactions on Learning Technologies (47). Other key contributors include Applied Sciences Basel (43), the International Journal of Artificial Intelligence in Education (31), and Computer Applications in Engineering Education (28). LA research spans computer science, education, engineering, and AI, fostering collaboration and innovation. ACM conferences and IEEE journals are vital for research visibility and impact. Staying engaged with these sources and actively participating in conferences and journals enhances community collaboration and knowledge sharing.

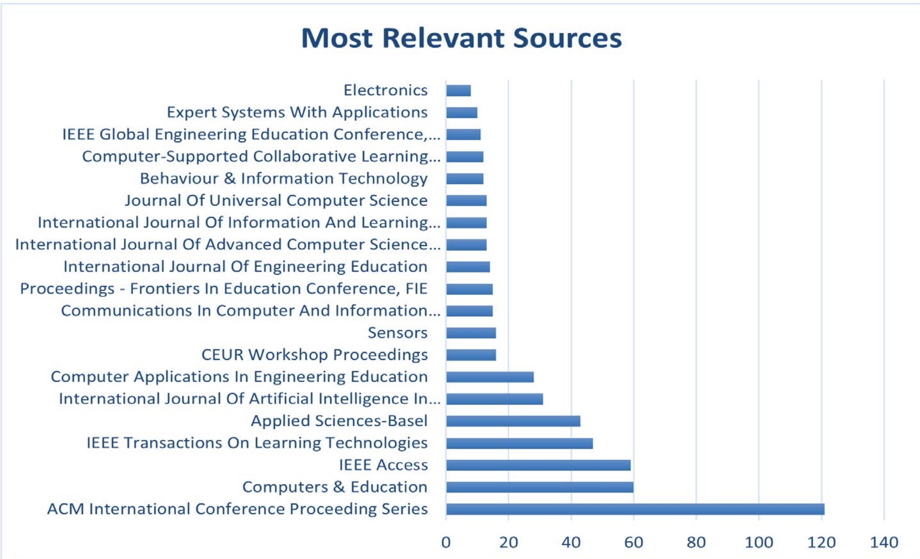

Figure 4: Most relevant sources

Figure 5 highlights the regional impact of key publications. Lecture Notes in Computer Science leads with 551 local citations, reflecting a strong regional influence. IEEE Transactions on Learning Technologies follows with 534 citations, emphasising its role in educational technology research. Technology Education & Society (312 citations) demonstrates regional interest in the sociological aspects of technology in education. Meanwhile, IEEE Access (289 citations) attracts multidisciplinary specialists, further enhancing its local impact.

Expert Systems with Applications (274 citations) showcases the practical use of expert systems in real-world learning analytics (LA) solutions. The Internet and Higher Education (262 citations) highlights the role of internet technology in higher education and digital learning. J Learning Anal (252 citations) reflects strong regional interest in LA methodologies and applications. Other local papers (195–114 citations) explore various LA topics, demonstrating a diverse research focus. Specialized conferences, like LAK16, contribute to local research by focusing on specific themes that enrich LA inquiry. The diversity of local sources reflects active scholarly participation across platforms. Additionally, the prominence of journals like Expert Systems with Applications underscores a regional preference for practical LA solutions, emphasising actionable discoveries.

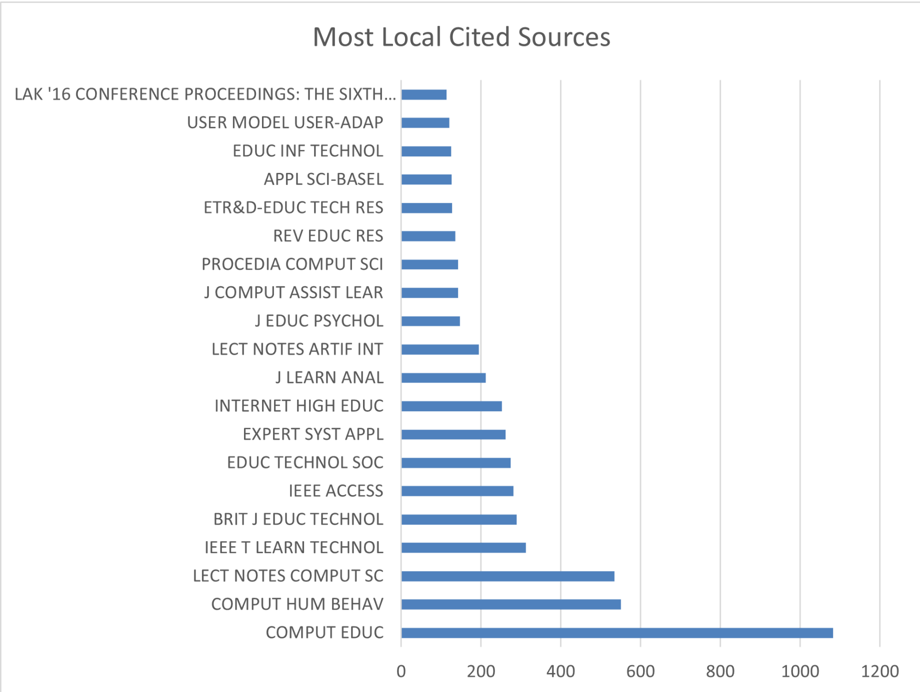

Figure 5: *Most local cited sources*

Table 3 evaluates the citations of key documents in Learning Analytics (LA) across computer science, engineering, and mathematics. Citations serve as markers of a document's significance and intellectual influence (Grant et al., 2000; Waheed et al., 2018; Agbo et al., 2021). The top 20 references, ranked by local citations, highlight the role of educational data mining and LA in programming education. Notably, Ihantola et al. (2015) received the most local (38) and global (210) citations. This foundational work offers comprehensive insights and case studies.

The second most referenced study, Carter et al. (2015), received 19 local and 82 global citations. This study focuses on prediction models generated from programming behaviour. Ahadi et al. (2015) garnered 14 local and 123 global citations, underscoring the value of machine learning in assisting students. Grover and Korhonen (2017) and Grover et al. (2017) both received 9 local citations, contributing significantly to LA in computer education. Other key documents address diverse topics, such as learning curve analysis to identify students needing assistance and predictive models. These documents demonstrate the importance of data-driven techniques in enhancing student success in programming education.

Table 3: *Top 20 references by local citations*

#	Document Title	Authors & Year Published	Publication Source	Local Citations	Global Citation
1	Educational Data Mining and LA in Programming: Literature Review and Case Studies	(Ihantola et al., 2015)	ITICSE-WGR '15: Proceedings of the 2015 ITICSE on Working Group	38	210
2	The Normalized Programming State Model: Predicting Student Performance in Computing Courses Based on Programming Behavior	(Carter et al., 2015)	ICER '15: Proceedings of the eleventh annual International Conference on International Computing Education Research	19	82
3	Exploring Machine Learning Methods to Automatically Identify Students in Need of Assistance	(Ahadi et al., 2015)	ICER '15: Proceedings of the eleventh annual International Conference on International Computing Education Research	14	123
4	Unlocking the Potential of LA in Computing Education	(Grover & Korhonen, 2017)	ACM Transactions on Computing Education (TOCE)	9	5
5	A Framework for Using Hypothesis-Driven Approaches to Support Data-Driven LA in Measuring Computational Thinking in Block-Based Programming Environments	(Grover et al., 2017)	ACM Transactions on Computing Education (TOCE)	9	63
6	How novices tackle their first lines of code in an IDE: analysis of programming session traces	(Vihavainen et al., 2014)	Koli Calling '14: Proceedings of the 14th Koli Calling International Conference on Computing Education Research	9	32

7	Learning Curve Analysis for Programming: Which Concepts do Students Struggle With?	(Rivers et al., 2016)	ICER '16: Proceedings of the 2016 ACM Conference on International Computing Education Research	8	45
8	Evaluating Neural Networks as a Method for Identifying Students in Need of Assistance	(Castro-Wunsch et al., 2017)	SIGCSE '17: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education	7	47
9	Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses	(Costa et al., 2017)	Computers in Human Behaviour	6	253
10	Automatic Inference of Programming Performance and Experience from Typing Patterns	(Leinonen et al., 2016)	Proceedings of the 47th ACM Technical Symposium on Computing Science Education, 2016	6	45
11	Integrating LA in an Educational MMORPG for Computer Programming	(Malliarakis et al., 2014)	2014 IEEE 14th International Conference on Advanced Learning Technologies	6	27
12	Discriminating Programming Strategies in Scratch: Making the Difference between Novice and Experienced Programmers	(Kesselbacher & Bollin, 2019)	WiPSCE '19: Proceedings of the 14th Workshop in Primary and Secondary Computing Education	5	1
13	Personalizing Computer Science Education by Leveraging Multimodal LA	(Azcona et al., 2018)	2018 IEEE Frontiers in Education Conference (FIE)	5	4

14	Detecting students-at-risk in computer programming classes with LA from students' digital footprints	(Azcona et al., 2019)	User Modeling and User-Adapted Interaction, 2019	5	40
15	Analysis Method of Student Achievement Level Utilizing Web-Based Programming Education Support Tool Pgtacer	(Murata & Kakeshita, 2016)	2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)	5	6
16	Identification of programmers from typing patterns	(Longi et al., 2015)	Proceedings of the 15th Koli Calling conference on computing education research, 2015	5	26
17	Amoeba: Designing for collaboration in computer science classrooms through live LA	(Berland, Davis and Smith, 2015)	International Journal of Computer-Supported Collaborative Learning, 2015	5	51
18	LA to improve coding abilities: a fuzzy-based process mining approach	(Ardimento et al., 2019)	2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019	4	10
19	Blending measures of programming and social behaviour into predictive models of student achievement in early computing courses	(Carter, Hundhausen, Adesope, 2017)	ACM Transactions on Computing Education (TOCE), 2017	4	32
20	Data mining of students' behaviours in programming exercises	(Kato, Kambayashi, Kodama, 2016)	Smart Education and e-Learning 2016	4	6

A word cloud analysis of the top 20 articles highlights LA research's strong focus on programming education. Key themes include students, programming, data, learning, courses, and analytics, all emphasising student engagement. Technical terms such as programming, data, process, and technique indicate detailed investigations, while learning analytics, data mining, and predictive models reflect data-driven insights aimed at improving learning outcomes. Words like support, educational, and improve showcase efforts to enhance teaching tools and frameworks. Meanwhile, terms such as performance, identify, and novice signal a focus on early-stage programming challenges. Overall, Figure 6 illustrates the broad scope of LA research in programming education, covering student engagement, data-driven methods, and educational support to improve learning outcomes.

Figure 6: *Most local cited sources*

3.3 Scientific publication by region/countries

This study assessed LA's global impact in computer science, engineering, and mathematics by analysing publication and citation rates to measure research influence. Evaluating regional contributions helps map global knowledge flows and reveals disparities in research capacity, collaboration, and visibility within LA (Ifenthaler & Yau, 2020; Viberg et al., 2018). Slovenia leads with an average of 59 citations, reflecting a strong research impact, while the Netherlands (33 citations) and Serbia (28 citations) also make notable contributions. Spain, despite a lower average citation rate of 12.40, enriches LA's knowledge through a high publication volume and diverse research efforts.

Table 4: *Top 20 most cited countries*

#	Country	Total Citation (Tc)	Average Article Citations
1	Spain	1633	12,40
2	United States of America (USA)	1511	15,30
3	China	874	8,70
4	Canada	501	25,00
5	United Kingdom	481	20,00
6	Australia	407	11,30
7	Netherlands	402	33,50
8	Brazil	371	24,70
9	Saudi Arabia	239	8,90
10	Serbia	231	28,90
11	Turkey	191	23,90
12	India	171	5,00
13	Belgium	140	20,00
14	Korea	129	9,20
15	Greece	123	7,20
16	Italy	119	8,50
17	Chile	116	8,30
18	Finland	107	6,70
19	Ireland	107	21,40
20	France	93	13,30

Countries such as Saudi Arabia, Canada, Iran, Brazil, Turkey, Portugal, Ireland, the UK, and Belgium demonstrate diverse expertise and global impact in LA research, fostering opportunities for international collaboration. This study evaluates total citations (research volume) and average citations (paper quality) for a balanced assessment. Table 4 shows Spain leading in Europe with 1,633 publications, followed by the UK, Australia, the Netherlands, and Serbia. In North and South America, the USA dominates with 1,511 publications, while Canada and Brazil reinforce academic engagement through robust research infrastructures. China and Saudi Arabia have emerged as key contributors in Asia, reflecting a growing academic influence. Although Africa lags behind Europe, Asia, and the Americas, interest is rising in Egypt, Nigeria, South Africa, Tunisia, Morocco, Botswana, and Ghana, signalling potential for a stronger research presence. However, LA implementation remains concentrated in the Global North, leading to a heavy reliance on imported educational models that may widen existing inequalities.

Reform programmes often fail to meet their goals and may exacerbate imbalances. A critical review of past North-South collaborations is essential to promote local ownership and engagement (Prinsloo & Kaliisa, 2022). If challenges remain unaddressed, LA deployment in Africa may fall short, echoing past disappointments in educational technology (Prinsloo & Kaliisa, 2022). Despite its relative novelty in Africa, the increasing number of LA publications signals a rising interest and potential. Strengthening global collaboration, especially between Europe, North America, Asia, and Africa, could enhance research impact and drive innovation in LA.

3.4 Prolific scholars, institutions, and collaborations network

3.4.1 Prolific scholars of learning analytics

Analysing author output trends shows increased research activity after 2015, reflecting growing engagement. Identifying prolific scholars in a scoping review reveals intellectual leadership, scholarly influence, and collaboration patterns that shape the trajectory of LA research (Ergul Sonmez, 2024). Munoz-Merino P (18 papers, 16 citations from three works) and Gasevic D (17 papers, 24 citations per year in 2021) made significant contributions. Hsiao I (17 papers, 10.6 citations per year in 2017), Kloos C (16 citations per year in 2020), and Ruiperez-Valiente J (13 citations per year in 2017) also had a notable impact. Despite starting in 2019, Chen G (25.5 citations per year in 2020 from five publications) showed immediate influence. Gasevic D and Munoz-Merino P each averaged 11.25 citations per year, reinforcing their relevance. Gasevic D's 2021 study (seven citations) further solidified his impact. The collaborations between Chen G and Gasevic D suggest strong research synergies. The varied impact trajectories, from rapid influence to steady growth, highlight a diverse research landscape, with scholars driving collaboration, debate, and meaningful contributions.

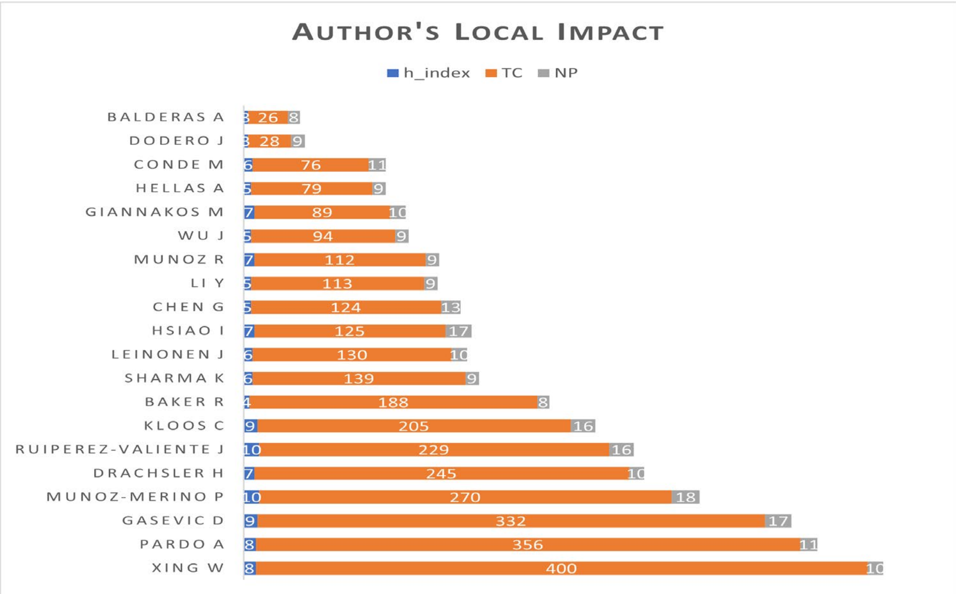

Figure 7: *Author's production overtime*

Figure 7 presents the authors' production and impact, focusing on h-index and total citations. Xing W (400 citations, h-index 8), Pardo A (356 citations, h-index 8), and Gasevic D (332 citations, h-index 9) lead in influence. Munoz-Merino P (270 citations, 18 papers) and Ruiperez-Valiente J (229 citations, 16 papers) hold the highest h-index (10), indicating strong recognition. In contrast, Dodero J (28 citations, h-index 3) and Balderas A (26 citations, h-index 3) have lower impact. Authors with h-indices between 8 and 10 demonstrated significant influence, with at least eight citations per top publication, while those with h-indices between 3 and 7 had a moderate impact, with at least three citations per top work. Researchers with over 300 citations established notable reputations, while those with 20 to 30 citations maintained productivity within niche areas. Xing W, Pardo A, and Gasevic D balanced high output and strong influence, consistently contributing to LA research. Munoz-Merino P and Ruiperez-Valiente J, despite fewer publications, maintained a high impact in specialised areas. Figure 7 highlights the varying impact of authors, with some gaining recognition through highly cited papers, while others thrived on prolific publishing.

Figure 8 applies Lotka's Law, analysing author productivity and research concentration. The study reveals widely distributed contributions, with 79.2% of papers (2,594 out of 3,277) single-authored, while 12.5% involved two authors and 4.5% had three authors. This suggests that LA research is not dominated by a few experts but rather a diverse pool of contributors, reinforcing its collaborative nature. In Figure 8, the solid line represents the observed distribution of author productivity in the dataset, while the dotted line depicts the expected theoretical distribution based on Lotka's Law. The close alignment supports the typical bibliometric pattern, where a few authors produce most publications, a finding also observed in other bibliometric studies applying Lotka's Law (Kushairi & Ahmi, 2021).

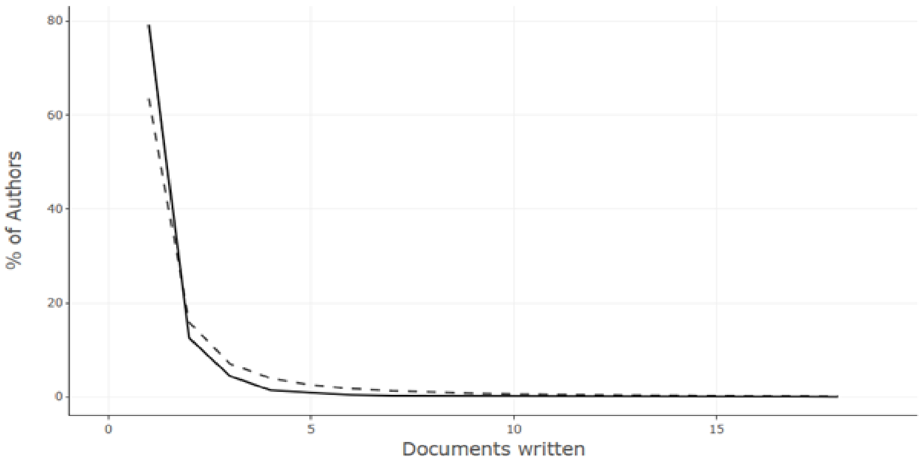

Figure 8: *Frequency distribution of publications through lotka's law*

3.4.2 Institutions, co-authorship, and collaboration network

Table 5 presents the top 20 unique institutions contributing to LA research, based on aggregated publication output and author affiliations. Institutions such as Monash University, King Abdulaziz University, Universidad Carlos III de Madrid, Central China Normal University, and the University of South Australia demonstrate significant engagement in LA scholarship, often through cross-institutional collaboration and diverse research contributions.

Table 5: *Most relevant institutions*

#	Affiliation	Articles	Countries
1	Monash University	30	Australia
2	King Abdulaziz University	27	Saudi Arabia

3	Central China Normal University	19	China
4	University of South Australia	17	Australia
5	Universidad Carlos III De Madrid	17	Spain
6	San Diego State University	14	USA
7	UOC (Universitat Oberta De Catalunya)	13	Spain
8	California State University System	12	USA
9	Complutense University of Madrid	12	Spain
10	University of Florida	11	USA
11	Zhejiang University	11	China
12	National Central University	10	China
13	University of Murcia	10	Spain
14	Egyptian Knowledge Bank (EKB)	10	Egypt
15	Norwegian University of Science and Technology	10	Europe
16	State University System of Florida	9	USA
17	The University of Hong Kong	9	China
18	Ming Chuan University	9	China
19	Massachusetts Institute of Technology	9	USA

Figure 9 illustrates institutional collaborations in research and academic activities. Monash University, the University of Edinburgh, and King Abdulaziz University are central to global research clusters, fostering knowledge exchange and joint initiatives. These partnerships drive new projects, joint publications, and academic events across multiple disciplines. The cooperation between universities in Australia, Saudi Arabia, Spain, and the UK highlights the global impact of their research contributions. Such collaborations enhance research quality, integrate diverse perspectives, and address multidisciplinary challenges. International cooperation also fosters cultural exchange, creating an inclusive academic environment. Sustained partnerships lead to stronger research alliances, expanded collaboration, and greater innovation, reinforcing LA’s global research structure.

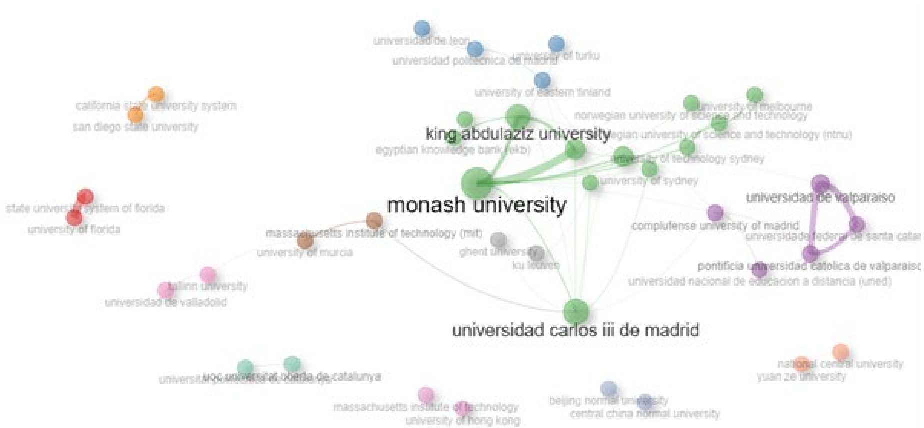

Figure 9: *Institutions’ collaboration network*

Figure 10 compares single-country (SCP) and multiple-country (MCP) publications by authors' affiliations. SCPs were more prevalent, aligning with Lotka's Law, which states that a small number of authors produce most papers. Spain led with 106 SCPs compared to 26 MCPs, followed by China (78 SCPs, 22 MCPs) and the USA (85 SCPs, 14 MCPs out of 99 studies). While these countries engage in extensive individual research, African nations exhibit limited collaboration, which may hinder productivity. However, this also presents opportunities for growth through increased global and regional partnerships in LA research.

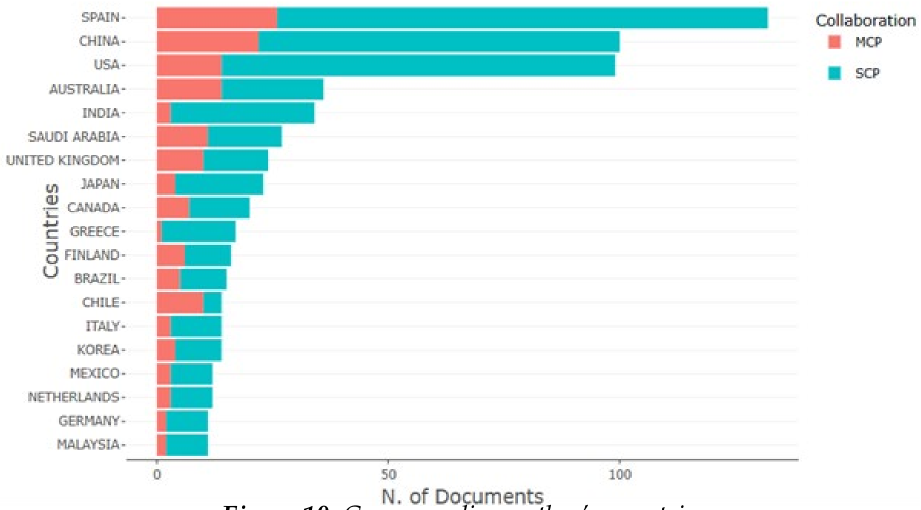

Figure 10: *Corresponding author's countries*

This study used Biblioshiny to analyse co-authorship and social collaboration, examining cooperation among individuals, institutions, and countries. Figure 11 illustrates the author collaboration network, where larger nodes represent stronger networks (Agbo et al., 2021). The top 20 prolific authors, including Munoz-Merino P, Ruiperez-Valiente J, Kloos C, Gasevic D, and Pardo A, had the greatest influence, often collaborating within the same field (brown cluster).

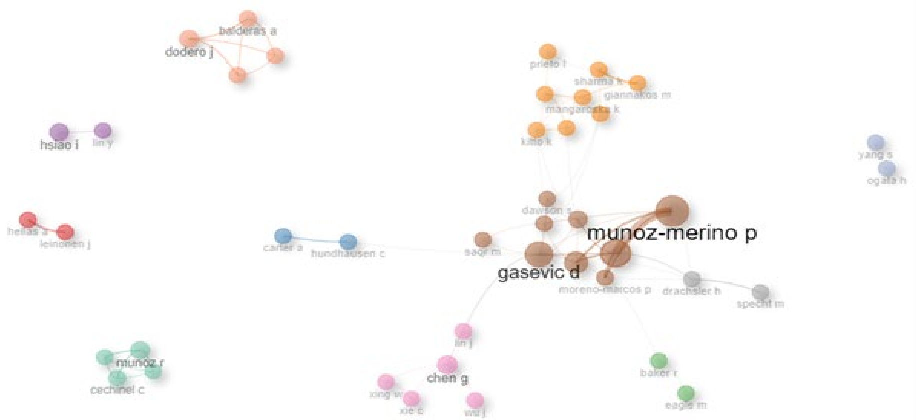

Figure 11: *Author's collaboration network*

Intensive collaborations within clusters foster expertise, drive innovation, and advance specialised knowledge. These networks facilitate knowledge exchange, promoting research breakthroughs. Understanding collaboration patterns provides insights into LA knowledge generation and helps institutions and policymakers optimise investment and research strategies.

3.5 Thematic focus of learning analytics

This section explores the major themes and areas of study in the field of LA research. It discusses the most common words that emerged from the word cloud, the tree map, trending topics, and co-occurrence networks. Furthermore, it examines whether there is a shift in LA research among field researchers.

3.5.1 Keywords analysis, co-occurrence network, and trend topics

Analysing authors' keywords helps to identify trending research topics and interests (Song et al., 2019; Sarpong et al., 2023). This study employed keyword analysis to track annual increases and keyword prevalence in LA research. Figure 12 highlights the most frequently used terms, with "learning analytics" appearing 527 times out of 3,088 keywords, ranking first among the top 10. Keywords Plus (Figure 13) is a semi-automated technique that scans reference and article titles, providing a broader perspective on knowledge structure (Della Corte et al., 2019).

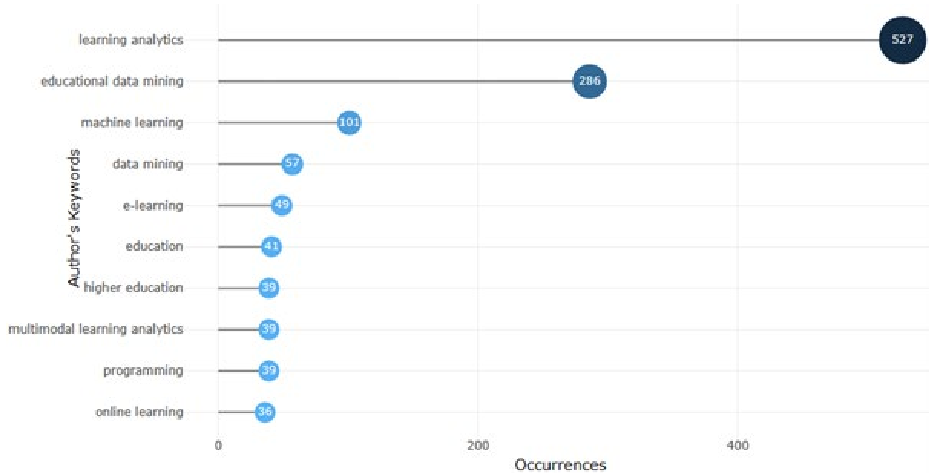

Figure 12: *Authors' most frequent words*

Figure 14's word cloud highlights key research themes through term frequency and relevance. Larger terms such as "students" (351), "learning analytics" (229), and "e-learning" (152) indicate a strong focus on student analysis, data-driven learning, and online education. "Computer programming" (138) emphasises programming skills in education, while "performance" (101) and "teaching" (98) reflect interest in student outcomes and instructional methods. Moderately frequent terms like "education computing" (119) and "learning systems" suggest an interest in integrating education and technology. The prominence of "learning analytics" and "students" underscores a focus on analysing student data for learning patterns and performance metrics. Frequent mentions of "e-learning," "computer programming," and "education computing" reinforce an emphasis on technology-enhanced learning and online education. The word cloud indicates a dataset centred on students, learning analytics, e-learning, and programming education.

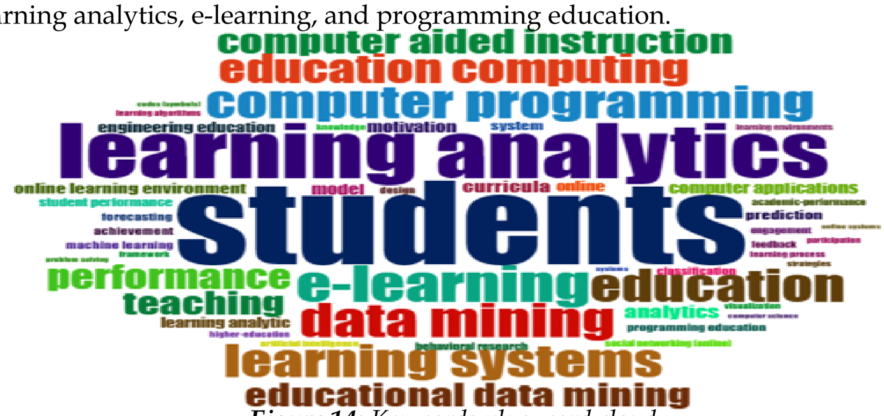

Figure 14: *Keywords plus word cloud*

The co-occurrence network analysis examines keyword relationships to provide deeper insights into the field. Figure 15 visualises these connections, where thicker lines denote stronger keyword associations, while absent links indicate no connection (Tlili et al., 2022; Sarpong et al., 2023). The analysis identified "students" as the most interconnected node, strongly linked to "learning analytics," "educational data mining," "education," "data mining," "learning systems," "computer-aided instruction," "e-learning," and "educational computing." These connections emphasise a primary focus on student-related issues in education, analytics, technology, and data mining. Weaker associations with "computer programming," "teaching," and "computer-aided instruction" suggest their relevance but lesser prominence. The dominance of "students" (351) in LA research highlights a student-centric focus, while "teachers" (98) appear less frequently, indicating a secondary emphasis. This suggests a research focus on student data analysis, learning behaviours, and educational technology for enhanced learning. Potential research areas include personalised learning, educational data analysis for student outcomes, learning management systems, and technology integration in education.

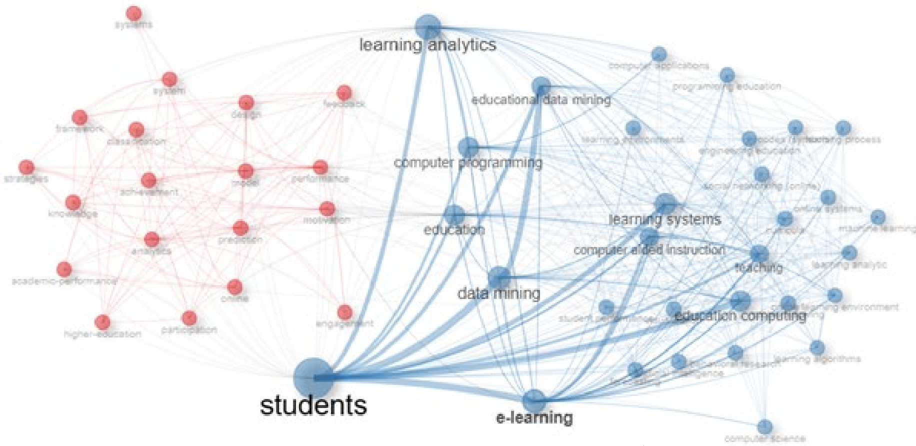

Figure 15: *Co-occurrence network*

Trending topics in published articles reveal shifts in popular research themes over time. Figure 16 illustrates how their importance has changed annually, with node sizes representing usage frequency in journal articles. Authors' keywords, aligned with journal content, provide insights into the evolving LA research landscape (Song et al., 2019; Agbo et al., 2021). Figure 16 offers a hierarchical view of key themes mentioned annually, reflecting changing interests and objectives. Tracking trends highlights shifts in research focus. In 2015, "assessment" and "flow visualisation" emphasised student performance monitoring and data analysis. By 2017, "teaching," "education," and "computer programming" gained attention, indicating increased interest in pedagogy and technology-enhanced learning. In 2018, "LA" (229 times), "data mining," and "e-learning" became prominent, underscoring a growing focus on data-driven education and online learning.

During COVID-19, discussions on students surged due to the pandemic's impact on education. Simultaneously, "learning systems" and "education computing" gained traction, reflecting a heightened interest in remote learning and technology integration. In 2021 and 2022, emphasis on "performance" increased alongside sustained interest in LA, highlighting its continued relevance. The frequent mention of "performance" signals strong interest in assessing student outcomes and instructional effectiveness. These trends reflect a dynamic research landscape shaped by technology, remote learning, and data-driven education, showcasing continuous efforts to enhance learning experiences in a rapidly evolving environment.

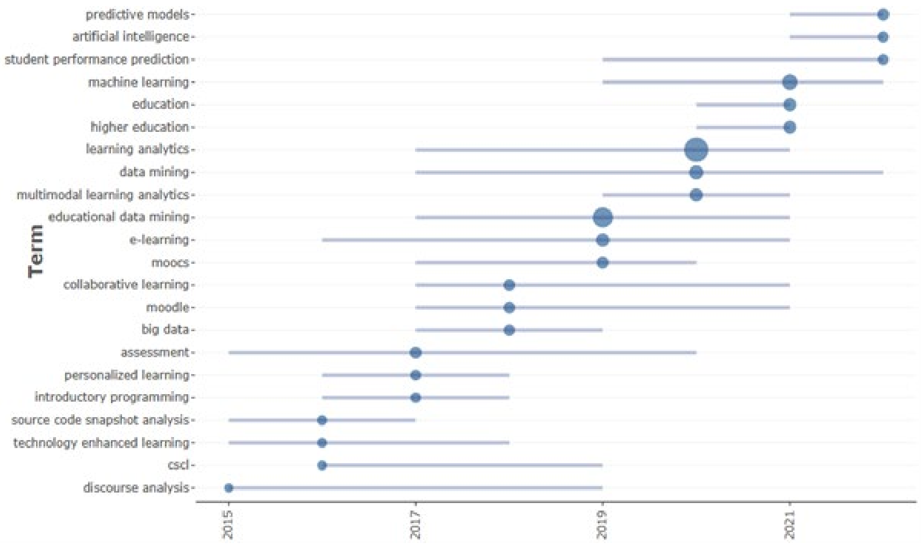

***Figure 16:** Trending topics*

The current study generated a thematic map of the LA sector, as shown in Figure 17. This map offers valuable insights into the evolving landscape of LA. The number of associations between nodes indicates their centrality and relevance within the thematic network. Additionally, the coherence among nodes reflects the density of research topics, suggesting their potential for growth and long-term sustainability.

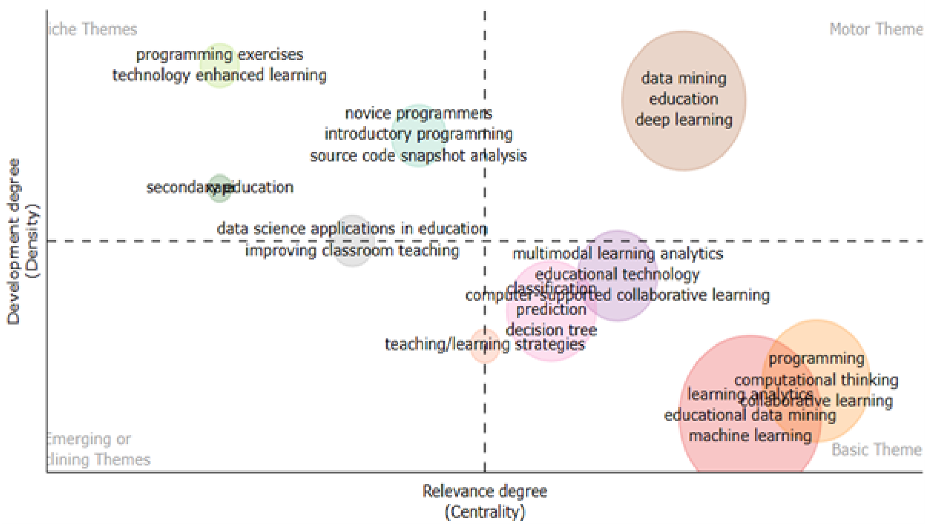

***Figure 17:** Thermal map*

Figure 17 divides themes into four quadrants (Q1–Q4), each representing different research categories. Q1 encompasses overarching ideas, Q2 focuses on specific issues, Q3 highlights emerging or fading themes, and Q4 contains key themes essential to LA research. Q4 themes, such as programming, computational thinking, collaborative learning, LA, educational data mining, machine learning, and educational technology, are crucial for advancing LA research. Although teaching and learning strategies in Q3 are emerging or declining, their connection to Q4 themes indicates their continued role in shaping LA's direction and future development.

Q2 themes, including the introduction to programming, source code snapshot analysis, novice programmers, programming exercises, technology-enhanced learning, secondary education, and data science applications, exhibit strong internal links but have yet to make a significant impact on LA. The study highlights their latent potential, suggesting that deeper integration with LA could drive multidisciplinary research and innovation. Additionally, data mining, education, and deep learning are identified as major concerns likely to shape the direction of LA research.

The theme map illustrates LA as a dynamic, multidisciplinary field. Q4's key themes provide a foundational research framework, while emerging issues in teaching and learning offer opportunities for future growth. The intersections between programming, data mining, education, and deep learning highlight collaboration opportunities and emphasise the need for scholars to bridge these fields. Collaboration, particularly in areas like data mining, education, and deep learning, is crucial for advancing LA research. The map underscores the importance of continuous adaptation and integration of diverse themes, fostering innovation and progress. Over time, the significance of themes evolves, reflecting LA's ability to adapt to new ideas, technologies, and pedagogical approaches.

4. Conclusion

This study explores the applications of Learning Analytics (LA) in programming education by addressing two core research questions. Through a comprehensive bibliometric and thematic analysis of 1,208 publications from 2014 to 2023, key insights were identified regarding research trends, challenges, and emerging priorities in the field.

To address RQ1, the analysis reveals that LA research in programming education is dominated by themes such as educational data mining, performance prediction, and student engagement. Although countries like Spain, the USA, and China lead in publication volume, collaboration patterns and regional impact vary, with limited integration of LA into core programming concepts like debugging and exercises. These studies highlight underexplored areas and regional disparities in research. In response to RQ2, the study identifies effective strategies informed by LA, including early detection of at-risk students, personalised feedback systems, and the use of programming artefacts to monitor learning. However, gaps remain in how LA is leveraged to support higher-order skills and collaborative learning. The findings suggest that aligning LA with pedagogical frameworks such as cognitive load theory and self-regulated learning can enhance its practical value in programming instruction.

In summary, this study reinforces the critical role of LA in improving teaching strategies and learner outcomes in programming education. By mapping existing efforts and uncovering research gaps, it lays the foundation for more targeted, theory-informed, and data-driven educational practices. Future research should focus on deepening LA's integration with pedagogical design, particularly in underrepresented areas such as formative feedback, debugging, and collaborative problem-solving.

4.1 Further research agenda

Future research in Learning Analytics (LA) within programming education should consider several interconnected themes. One area deserving attention is the use of data generated from programming exercises and source code snapshots, which could provide valuable insight into students' coding progression, strategy development, and instructional responsiveness. Additionally, evolving and declining teaching strategies, particularly those aimed at improving classroom engagement and effectiveness, remain underexamined in the LA literature. Understanding how pedagogical approaches intersect with LA adoption, implementation, and outcomes can enrich insights into the dynamic nature of programming education.

Another area of growing importance is the integration of LA with core programming topics and cognitive theories. This integration can foster adaptive, theory-informed feedback, enhance engagement, and provide more tailored learning support. Frameworks such as behaviourism, cognitive load theory, constructivism, connectivism, and humanism offer valuable perspectives for exploring how students learn, interact, and reflect within LA-supported environments. Advancing such interdisciplinary intersections among computer science, cognitive science, educational psychology, and data science will pave the way for more holistic and impactful LA applications in programming education.

Continued exploration of the above themes will not only expand the theoretical and empirical foundations of LA but will also contribute to more effective, inclusive, and data-informed approaches to programming education. While this study offers valuable insights, certain limitations should be acknowledged. Although data were sourced from reputable databases (Scopus and WOS), studies published in non-indexed venues or non-English languages may have been excluded. Furthermore, bibliometric indicators emphasise publication frequency and citation patterns, which may not fully capture the methodological rigour or pedagogical impact of the reviewed studies.

5. Declarations

Author Contributions: Conceptualisation (K.E.M. & H.T.); literature review (K.E.M., H.T., S.M. & M.M.); methodology (K.E.M., S.M. & H.T.); software (K.E.M. & H.T.); validation (K.E.M., H.T., S.M. & M.M.); formal analysis (K.E.M.); investigation (K.E.M.); data curation (K.E.M.) drafting and preparation (K.E.M.); review and editing (K.E.M., H.T., S.M., & M.M.); supervision (S.M., H.T., M.M.); project administration (K.E.M., S.M., H.T. & M.M.); funding acquisition (N/A). All authors have read and approved the published version of the article.

Funding: The study did not receive any external funding.

Acknowledgements: We sincerely thank Sol Plaatje University for supporting this research, the University of Johannesburg for providing publication support, and my supervisors for their invaluable guidance, insights, and encouragement.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: This review is based entirely on publicly available data and information sourced from peer-reviewed articles, reports, and other academic publications cited in the manuscript. No new primary data were generated or analysed during this study. Readers may refer to the cited sources for detailed information.

References

Agbo, F. J., Oyelere, S. S., Suhonen, J., & Tukiainen, M. (2021). Scientific production and thematic breakthroughs in smart learning environments: A bibliometric analysis. *Smart Learning Environments*, *8*(1), 1.13. https://doi.org/10.1186/s40561-020-00145-4

Agrawal, S., Oza, P., Kakkar, R., Tanwar, S., Jetani, V., Undhad, J., & Singh, A. (2024). Analysis and recommendation system based on PRISMA checklist to write systematic review. *Assessing Writing*, *61*, 100866. https://doi.org/10.1016/j.asw.2024.100866

Asai, S., Phuong, D. T. D., Harada, F., & Shimakawa, H. (2019). Predicting cognitive load in acquisition of programming abilities. *International Journal of Electrical and Computer Engineering*, *9*(4), 3262–3271. https://doi.org/10.11591/ijece.v9i4.pp3262-3271

Bosse, Y., & Gerosa, M. A. (2017). Why is programming so difficult to learn? *ACM SIGSOFT Software Engineering Notes*, *41*(6), 1–6. https://doi.org/10.1145/3011286.3011301

Della Corte, V., Del Gaudio, G., Sepe, F., & Sciarelli, F. (2019). Sustainable tourism in the open innovation realm: A bibliometric analysis. *Sustainability (Switzerland)*, *11*(21), 1-15. https://doi.org/10.3390/su11216114

Echchakoui, S. (2020). Why and how to merge Scopus and Web of Science during bibliometric analysis: The case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, *8*(3), 165–184. https://doi.org/10.1057/s41270-020-00081-9

Eloy, A., Achutti, C. F., Fernandez, C., & Lopes, R. D. D. (2022). A data-driven approach to assess computational thinking concepts based on learners' artifacts. *Informatics in Education*, *21*(1), 33–54. https://doi.org/10.15388/infedu.2022.02

Ergul Sonmez, E. (2024). A comprehensive bibliometric analysis of learning analytics in education research. *International Journal on Studies in Education*, *6*(3), 446–462. https://doi.org/10.46328/ijonse.234

Giannakos, M. N., Krogstie, J., & Aalberg, T. (2016). Video-based learning ecosystem to support active learning: Application to an introductory computer science course. *Smart Learning Environments*, *3*(1), 1-13. https://doi.org/10.1186/s40561-016-0036-0

Grant, J., Cottrell, R., Fawcett, G., & Cluzeau, F. (2000). Evaluating "payback" on biomedical research from papers cited in clinical guidelines: Applied bibliometric study. *British Medical Journal*, *320*(7242), 1107–1111. https://doi.org/10.1136/bmj.320.7242.1107

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Ioannidis, J. P., Bendavid, E., Salholz-Hillel, M., Boyack, K. W., & Baas, J. (2022). Massive covidization of research citations and the citation elite. *Proceedings of the National Academy of Sciences*, *119*(28), e2204074119. https://doi.org/10.1073/pnas.2204074119

Junjia, Y., Alias, A. H., Haron, N. A., & Abu Bakar, N. (2023). A bibliometric review on safety risk assessment of construction based on CiteSpace Software and WoS Database. *Sustainability*, *15*(15), 11803. https://doi.org/10.3390/su151511803

Kushairi, N., & Ahmi, A. (2021). Flipped classroom in the second decade of the millennia: A bibliometrics analysis with Lotka's law. *Education and Information Technologies*, *26*(4), 4401–4431. https://doi.org/10.1007/s10639-021-10457-8

Moon, J., Do, J., Lee, D., & Choi, G. W. (2020). A conceptual framework for teaching computational thinking in personalised OERs. *Smart Learning Environments*, *7*(1), 1-14. https://doi.org/10.1186/s40561-019-0108-z

Muhuri, P. K., Shukla, A. K., & Abraham, A. (2019). Industry 4.0: A bibliometric analysis and detailed overview. *Engineering Applications of Artificial Intelligence*, *78*, 218–235. https://doi.org/10.1016/j.engappai.2018.11.007

Obaido, G., Agbo, F. J., Alvarado, C., & Oyelere, S. S. (2023). Analysis of attrition studies within the computer sciences. *IEEE Access*, *11*, 53736–53748. https://doi.org/10.1109/ACCESS.2023.3280075

Ogundeji, A. A., & Okolie, C. C. (2022). Perception and adaptation strategies of smallholder farmers to drought risk: A Scientometric Analysis. *Agriculture*, *12*(8), 1129. https://doi.org/10.3390/agriculture12081129

Pirri, S., Lorenzoni, V., & Turchetti, G. (2020). Scoping review and bibliometric analysis of big data applications for medication adherence: An explorative methodological study to enhance consistency in literature. *BMC Health Services Research*, *20*(1), 1-16. https://doi.org/10.1186/s12913-020-05544-4

Prinsloo, P., & Kaliisa, R. (2022). Learning analytics on the African continent: An emerging research focus and practice. *Journal of Learning Analytics*, *9*(2), 218–235. https://doi.org/10.18608/jla.2022.7539

Qian, Y., & Lehman, J. (2017). Students' misconceptions and other difficulties in introductory programming: A literature review. In *ACM Transactions on Computing Education* (Vol. 18, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3077618

Rodrigues, M., Franco, M., & Silva, R. (2020). COVID-19 and disruption in management and education academics: Bibliometric mapping and analysis. *Sustainability (Switzerland)*, *12*(18), 1-16. https://doi.org/10.3390/SU12187362

Sarpong, A. A., Arabiat, D., Gent, L., & Towell-Barnard, A. (2023). A bibliometric analysis of missed nursing care research: Current themes and way forward. *Nursing Forum*, *2023*, 1–17. https://doi.org/10.1155/2023/8334252

Song, Y., Chen, X., Hao, T., Liu, Z., & Lan, Z. (2019). Exploring two decades of research on classroom dialogue by using bibliometric analysis. *Computers and Education*, *137*, 12–31. https://doi.org/10.1016/j.compedu.2019.04.002

Tlili, A., Huang, R., Shehata, B., Liu, D., Zhao, J., Metwally, A. H. S., Wang, H., Denden, M., Bozkurt, A., Lee, L. H., Beyoglu, D., Altinay, F., Sharma, R. C., Altinay, Z., Li, Z., Liu, J., Ahmad, F., Hu, Y., Salha, S., & Burgos, D. (2022). Is metaverse in education a blessing or a curse: A combined content and bibliometric analysis. In *Smart Learning Environments* (Vol. 9, Issue 1). Springer. https://doi.org/10.1186/s40561-022-00205-x

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., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. In *Annals of Internal Medicine* (Vol. 169, Issue 7, pp. 467–473). American College of Physicians. https://doi.org/10.7326/M18-0850

Utamachant, P., Anutariya, C., & Pongnumkul, S. (2023). i-Ntervene: Applying an evidence-based learning analytics intervention to support computer programming instruction. *Smart Learning Environments*, *10*(1), 1-20. https://doi.org/10.1186/s40561-023-00257-7

Venigalla, A. S. M., & Chimalakonda, S. (2020). G4D - A treasure hunt game for novice programmers to learn debugging. *Smart Learning Environments*, *7*(1), 1–21. https://doi.org/10.1186/s40561-020-00129-4

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. In *Computers in Human Behaviour* (Vol. 89, pp. 98–110). Elsevier Ltd. https://doi.org/10.1016/j.chb.2018.07.027

Waheed, H., Hassan, S. U., Aljohani, N. R., & Wasif, M. (2018). A bibliometric perspective of learning analytics research landscape. *Behaviour and Information Technology*, *37*(10–11), 941–957. https://doi.org/10.1080/0144929X.2018.1467967

Wang, Q., Mousavi, A., & Lu, C. (2022). A scoping review of empirical studies on theory-driven learning analytics. *Distance Education*, *43*(1), 6–29. https://doi.org/10.1080/01587919.2021.2020621

Yu, J. H., Chang, X. Z., Liu, W., & Huan, Z. (2023). An online integrated programming platform to acquire students' behaviour data for immediate feedback on teaching. *Computer Applications in Engineering Education*, *31*(3), 520–536. https://doi.org/10.1002/cae.22596

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning-based recommender system: A survey and new perspectives. *ACM computing surveys*, *52*(1), 1-38. https://doi.org/10.1145/3285029