Revista Española
de Educación Médica

UNIVERSIDAD
DE MURCIA

Learning Analytics in Medical and Health Professions Education: A Bibliometric Mapping of Research Structure and Thematic Patterns (2014–2026).

Analítica del aprendizaje en la educación médica y de las profesiones sanitarias: un mapeo bibliométrico de la estructura de investigación y los patrones temáticos (2014–2026).

Metin Kartal[1]*, Alper Bayazıt[2]

[1] Department of Educational Sciences, Dede Korkut Faculty of Education, Kafkas University, Kars, Türkiye, c.metinkartal@gmail.com, https://orcid.org/0000-0002-7201-6904.

[2] Department of Medical Education and Informatics, Faculty of Medicine, Ankara University, Ankara, Türkiye, abayazit@ankara.edu.tr, https://orcid.org/0000-0003-4369-587X.

* Correspondencia: c.metinkartal@gmail.com

Recibido: 8/29/26; Aceptado: 9/10/26; Publicado: 9/11/26

Abstract.
This study aimed to map the development, intellectual structure, and thematic patterns of learning analytics research in medical and health professions education. A bibliometric and science-mapping analysis was conducted using 177 publications indexed in the Web of Science Core Collection between 2014 and 2026. Data were analysed with the Bibliometrix package and its Biblioshiny interface. The analysis examined annual scientific production, source distribution, author productivity, country-level production, three-field relationships, keyword co-occurrence, and thematic structure. The results indicated rapid growth in publication output, particularly after 2021, with a marked increase from 2024 onwards. Publications were distributed across established medical education and health professions education journals, suggesting growing bibliometric visibility within the field. Authorship and Lotka's Law patterns indicated a collaborative but unevenly consolidated research community. The thematic structure showed visible associations among learning analytics, feedback, performance, competence, technology, artificial intelligence, big data, educational data mining, and machine learning. By contrast, programmatic assessment, workplace-based assessment, and entrustable professional activities appeared less central in the mapped literature. These findings should be interpreted as patterns in Web of Science-indexed publications rather than direct evidence of educational effectiveness or implementation quality. The study contributes by organising the field's publication structure and identifying areas where future research could connect learning analytics more explicitly with longitudinal, workplace-based, and curriculum-level assessment practices.

Keywords: learning analytics, medical education, health professions education, artificial intelligence, predictive analytics, bibliometric analysis, science mapping, competency-based medical education

Resumen.
Este estudio tuvo como objetivo mapear el desarrollo, la estructura intelectual y los patrones temáticos de la investigación sobre analítica del aprendizaje en la educación médica y de las profesiones sanitarias. Se realizó un análisis bibliométrico y de mapeo científico a partir de 177 publicaciones indexadas en la Web of Science Core Collection entre 2014 y 2026. Los datos se

analizaron mediante el paquete Bibliometrix y su interfaz Biblioshiny. El análisis examinó la producción científica anual, la distribución de fuentes, la productividad de los autores, la producción por países, las relaciones de tres campos, la coocurrencia de palabras clave y la estructura temática. Los resultados indicaron un rápido crecimiento de la producción científica, especialmente después de 2021, con un aumento marcado a partir de 2024. Las publicaciones se distribuyeron en revistas consolidadas de educación médica y educación de las profesiones sanitarias, lo que sugiere una creciente visibilidad bibliométrica dentro del campo. Los patrones de autoría y la Ley de Lotka indicaron una comunidad investigadora colaborativa, aunque desigualmente consolidada. La estructura temática mostró asociaciones visibles entre analítica del aprendizaje, retroalimentación, rendimiento, competencia, tecnología, inteligencia artificial, big data, minería de datos educativos y aprendizaje automático. En contraste, la evaluación programática, la evaluación basada en el lugar de trabajo y las actividades profesionales confiables aparecieron como menos centrales en la literatura mapeada. Estos hallazgos deben interpretarse como patrones en publicaciones indexadas en Web of Science, y no como evidencia directa de efectividad educativa o calidad de implementación. El estudio contribuye a organizar la estructura de publicación del campo e identifica áreas en las que futuras investigaciones podrían conectar de manera más explícita la analítica del aprendizaje con prácticas longitudinales, basadas en el lugar de trabajo y orientadas al currículo.

Palabras clave: analítica del aprendizaje, educación médica, educación de las profesiones sanitarias, inteligencia artificial, analítica predictiva, análisis bibliométrico, mapeo científico, educación médica basada en competencias

1. Introduction

Medical and health professions education is becoming increasingly data rich. Competency-based medical education, programmatic assessment, simulation-based training, e-portfolios, online learning platforms and AI-supported educational environments give rise to heterogeneous forms of learner data. These data may shed light on the topics of clinical reasoning, self-regulated learning, feedback uptake, collaborative practice and progression toward professional competence. But the availability of data alone does not create educational value. Within the framework of medical education, the value of learning analytics is not simply in data collection, prediction of performance or reporting dashboards, but rather in the way data are interpreted theoretically, translated pedagogically, governed ethically and used to support improvement at the level of the learner, course, program and institution.

Learning analytics is the collection and analysis of learner data to gain insights and improve learning environments (1). Early work was mostly focused on prediction, retention monitoring, early-warning systems and behavioural indicators from learning management systems (2,3). However, easily available traces such as logins, time-on-task, clickstream activity or participation counts are not necessarily meaningful representations of learning processes (4,5). Therefore, current scholarship stresses the importance of theory-informed and educationally legitimate analytics in which data are connected to higher-order constructs (6,7). This is particularly important in the education of medical and health professions, where evidence from OSCEs, simulations, e-portfolios, workplace-based assessments, supervisor narratives and entrustment decisions need to be interpreted in relation to clinical reasoning, feedback uptake, affective regulation, teamwork, professional identity formation and clinical trustworthiness (8-11). The closed-loop learning analytics take this logic one step further, connecting data collection, interpretation, feedback, intervention, and re-evaluation . Multimodal data such as speech, gesture, gaze, physiological signals, written reflections, and interaction traces may inform understanding of the embodied, social, temporal, and situated nature of clinical learning.

Developments in artificial intelligence and large language models have amplified the possibilities and responsibilities of learning analytics in medical education. Reflections, case analyses, feedback narratives, clinical reasoning notes and e-portfolio entries can be examined to look at how they are put together, the strength of the arguments, how the diagnoses are made and how the writers are progressing. But, as we know, AI-based interpretation must be handled with care. Statistical analysis of text sometimes can show things that are not true, especially when we use this information to decide how good someone is, how they are doing, what they need to learn, who is in charge of teaching them, or whether we can trust them (19-23).

The educational value of learning analytics in medical education relies on a human-centred, ethical and critical design. Analytic systems are not very accurate if they are not well integrated in teaching practice. They can also cause distrust in learners, increased surveillance, unclear interpretation and replacing professional judgement. Dashboards, alerts and feedback systems should be usable for educators, meaningful for learners, aligned with program decisions and transparent enough to promote dialogue rather than compliance (6,7). Being ethical also means being fair, giving learners control, and making sure the right people are held accountable. It also means acting when statistics indicate there are problems with learning or patient safety.

Despite growing research on learning analytics and related data-intensive approaches in medical and health professions education, the broader bibliometric structure of the field remains insufficiently mapped. Existing studies have examined specific applications such as prediction, feedback, simulation, assessment, artificial intelligence, and data-informed educational support. However, less is known about how this literature has developed over time, where it has been published, which countries and authors have contributed to it, how its intellectual and conceptual linkages are organised, and which themes are most visible within its indexed publication structure. A bibliometric and science-mapping analysis is therefore useful for clarifying these patterns and identifying potentially underrepresented areas.

This study was guided by the following research question: How has learning analytics research in medical and health professions education developed between 2014 and 2026 in terms of publication growth, source distribution, author productivity, geographical distribution, intellectual linkages, and thematic structure? To address this question, the study aimed to: (1) map the publication growth and general bibliometric profile of the field; (2) identify the main publication sources, author productivity patterns, and country-level distribution of research output; (3) examine the relational structure of the field through cited reference–author–keyword linkages and keyword co-occurrence patterns; (4) analyse the dominant thematic patterns within the indexed literature; and (5) identify potentially underrepresented areas and future research directions.

By providing an intellectual and thematic map of learning analytics research in medical and health professions education, this study clarifies how the topic has been represented in the indexed literature and which conceptual associations currently structure the field. Rather than evaluating the effectiveness of learning analytics interventions, it offers a bibliometric basis for identifying visible themes, less connected areas, and future lines of inquiry for more theory-informed, ethically governed, and educationally meaningful research.

2. Methods

2.1. Study Design

This study employed bibliometric analysis and science mapping to analyse the development, distribution, relational structure and thematic orientation of learning analytics research in medical and health professions education. Bibliometric analysis is a time-tested method to organise and evaluate scientific literature according to measurable indicators such as publications, authors, journals, countries, citations and keywords (24-26). In this study, bibliometric analysis was not used to evaluate the effectiveness of individual learning analytics interventions, but rather to map the field in terms of growth of publications, distribution of sources, productivity of authors, geographical distribution, intellectual linkages, relationships of keywords and dominant thematic patterns. The study data was constructed following a bibliometric design and included publications indexed in the Web of Science Core Collection related to learning analytics in medical and health professions education. The final dataset consisted of 177 eligible publications published between 2014 and 2026. The main bibliometric variables were year of publication, source title, authorship, country of affiliation, cited references, author keywords, Keywords Plus, citation indicators, and thematic clusters.

2.2. Data Source and Search Strategy

Bibliographic data were retrieved from the Web of Science Core Collection on June 22, 2026. The Web of Science database was chosen due to its structured bibliographic metadata and its widespread use in studies of bibliometrics and science mapping. We developed a search strategy to identify publications that intersect learning analytics with medical or health professions education. The search strategy comprised terms associated with learning analytics, educational data mining, predictive analytics, multimodal analytics, dashboard-based analytics, and AI-supported analytics combined with terms associated with medical education, health professions education, clinical education, simulation-based learning, clinical reasoning, competency-based medical education, and entrustable professional activities.

The following search string was used: ("learning analytics" OR "educational data mining" OR "academic analytics" OR "multimodal learning analytics" OR "predictive analytics" OR "learning dashboard*" OR "AI analytics") AND ("medical education" OR "health professions education" OR "clinical education" OR "simulation-based learning" OR "clinical reasoning" OR "competency-based medical education" OR "entrustable professional activities" OR "EPA").

The search was conducted using the Topic field, which covers titles, abstracts, author keywords, and Keywords Plus. The abbreviation “EPA” was retained because of its relevance to entrustable professional activities, but records retrieved through broad terms were manually checked for relevance to learning analytics and medical or health professions education. Records were exported from Web of Science in plain text format, including full records and cited references, for subsequent analysis in the Bibliometrix package and its Biblioshiny interface.

2.3. Eligibility Criteria and Screening Process

Publications were included if they addressed learning analytics, educational data mining, predictive analytics, multimodal learning analytics, dashboard-based analytics, or AI-supported educational analytics within the context of medical and health professions education. Studies focusing on medical education, clinical education, nursing education, residency training, simulation-based learning, clinical reasoning, competency-based medical education, entrustable professional activities, assessment, feedback, or curriculum improvement were considered eligible only when they had a clear connection to learning analytics or data-informed educational analysis.

No restrictions were applied in terms of language, document type, publication year or academic discipline so as to maximise coverage and capture the interdisciplinary nature of the field. The earliest record retrieved by the search strategy was from 2014 and this is not a year limit applied before the search. After the preliminary search, a manual screening of the records was performed. We removed duplicate records and excluded publications that did not meaningfully address learning analytics, educational analytics, or data-informed educational analysis in medical or health professions education. Publications that utilised analytics only in a clinical, biomedical, epidemiological or patient-care context without an educational focus were also excluded. This resulted in a final dataset of 177 publications from 2014 to 2026.

2.4. Data Cleaning and Preparation

The resulting dataset was imported into Bibliometrix/Biblioshiny in R for pre-processing and analysis. To enhance the consistency and interpretability of the bibliometric outputs, the data cleansing procedure was performed. The authors names, source titles, keywords and recurring terms were thoroughly examined for any possible inconsistencies. Where considered necessary, keyword harmonisation was applied with the aim of reducing terminological fragmentation. During the review process, closely related terms such as “learning analytics”, “educational data mining”, “predictive analytics”, “AI”, “artificial intelligence”, “medical education” and “medical-education” were reviewed. This approach was adopted to preserve the conceptual specificity whilst reducing redundancy. The cleaning process aimed to maintain the integrity of the original bibliographic metadata while improving the accuracy of keyword-based and thematic analyses. No full-text coding or quality appraisal of individual studies was conducted, as the aim of the study was to map the bibliographic and conceptual structure of the field rather than to synthesise intervention effectiveness.

2.5. Bibliometric and Science-Mapping Analysis

The cleaned dataset was analysed using the Biblioshiny interface of the Bibliometrix package in R (24). The analysis included descriptive bibliometric indicators, annual scientific production, most relevant publication sources, author productivity according to Lotka’s Law (29), country scientific production, three-field analysis, keyword co-occurrence analysis, and thematic mapping. Descriptive indicators were used to summarise the general characteristics of the dataset, including the number of publications, sources, authors, references, annual growth rate, international co-authorship, co-authors per document, and average citations per document. Annual scientific production was analysed to examine the development of the field over time. Source analysis identified the main publication venues. Lotka’s Law was applied to examine the distribution of author productivity. Country scientific production was used to map the geographical distribution of research output. Three-field analysis examined relationships among cited references, authors, and merged keywords. Keyword co-occurrence analysis identified the relational structure of the main concepts in the field and is provided as Supplementary Figure S1. Thematic mapping examined the centrality and density of dominant themes and is provided as Supplementary Figure S2.

2.6. Ethical Considerations

This study used publicly available bibliographic metadata and did not involve human participants, patient data, clinical records, biological materials, or experimental intervention. Therefore, ethical approval was not required. Nevertheless, the study followed principles of transparency and reproducibility by reporting the database, search strategy, screening criteria, data-cleaning procedures, and analytical techniques.

3. Results

The results are organised around the main bibliometric and science-mapping dimensions examined in the study. First, the general bibliometric profile and annual publication growth of the field are reported. Second, publication source distribution and author productivity patterns are examined. Third, the geographical and relational structure of the field is analysed through country-level production and three-field relationships among cited references, authors, and merged keywords. Finally, the thematic structure of the literature is summarised, with the keyword co-occurrence network and thematic map provided as Supplementary Figures S1 and S2. Together, these analyses show how learning analytics research in medical and health professions education has developed over time, where it has been published, how scholarly production is distributed, and which themes are most visible within the indexed literature.

3.1. General Bibliometric Profile

This analysis describes the general bibliometric pictures and characteristics of the dataset, including publication volume, source distribution, authorship structure, collaboration patterns, citation indicators, and keyword diversity.

Figure 1. Main bibliometric information of the dataset.

Figure 1 presents the general bibliometric characteristics and profiles of the dataset. The final dataset consisted of 177 publications emerged between 2014 and 2026 across 106 sources. This distribution indicates that researches on medical and health professions education are not solely focus on a small number of specialist releases, but is spread across a relatively broad publication landscape. Such spread is truly uniform with the interdisciplinary character of the topic, which connects medical education, health professions education, educational technology, assessment, data science, artificial intelligence, and learning analytics. The dataset under consideration comprised 767 authors, of whom only 13 produced single-authored documents. The average number of co-authors per document was 4.98 and the international co-authorship rate was 34.46%. The results indicate a predominantly collaborative publication profile. This may be indicative of the interdisciplinary expertise needed for standard research in the area of learning analytics in medical and health professions education. In this context, educational interpretation, clinical relevance, digital infrastructure and analytic techniques should be considered together. The annual growth rate of 21.15% and the average age of the documents of 3.84 years indicate that the literature mapped is recent and growing. The dataset also included 7,591 cited references with an average of 14.19 citations per document and 519 author keywords, indicating an emerging knowledge base with a significant conceptual diversity. The diversity of approaches reflects the multiple entry points through which learning analytics is viewed in the field including prediction, assessment, feedback, simulation, artificial intelligence, multimodal data, digital learning, and competency-based education. These indicators should be seen as a bibliometric profile of the indexed literature, and not as evidence of the effectiveness of the intervention or the quality of its implementation.

3.2. Temporal Trends in Publication Output

This part of analysis presents the annual distribution of publications and shows how scholarly production in the field has changed over time. Figure 2 depicts the annual distribution of publications on learning analytics in medical and health professions education between 2014 to 2026. The early period was characterised by limited publication output indicating that learning analytics entered medical and health professions education as a relatively marginal or exploratory research topic. From 2016 to 2020, publication activity became more trend, but still modest. The pattern emerged in the field have started to become visible before the most recent period of rapid growth. A more pronounced increase is seen after 2021, followed by a sharper rise from 2024 onwards. This acceleration may be explained by several overlapping developments: the expansion of digital and hybrid learning environments after the COVID-19 pandemic, increasing use of simulation and e-portfolios in medical education, growing attention to competency-based and programmatic assessment, and the rapid emergence of artificial intelligence and data-intensive educational tools. Driven perhaps by this convergence of digitalisation, assessment reform, and AI-supported educational innovation, the post-2024 increase may reflect a growing interest in the use of learner data to understand, monitor, and support medical education.

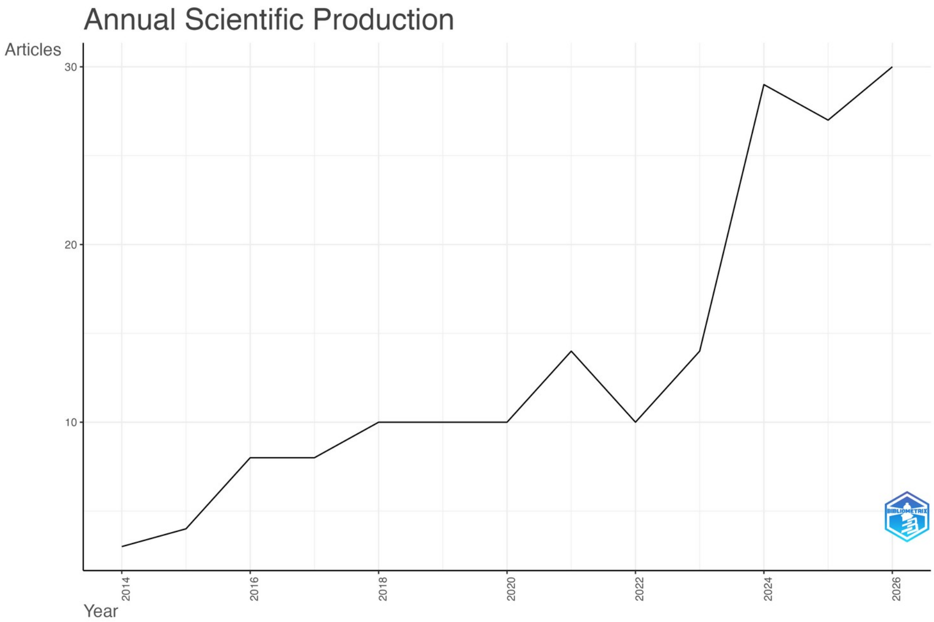

Figure 2. Annual scientific production.

However, the publication pattern should be interpreted cautiously. The increase in publication output indicates widely scholarly attention, but it does not by itself prove that learning analytics has become fully embedded in medical education practice or that the field has reached conceptual and epistemological saturation. Since the 2026 data were retrieved on 22 June 2026, the publication count for that year represents an incomplete publication year. Therefore, the apparent peak in 2026 should be read as evidence of strong early-year publication activity rather than as a definitive full-year comparison. All in all, the apparent peak in 2026 should be read as evidence of strong early-year publication activity rather than as a definitive full-year comparison. Overall, the temporal pattern suggests a rapidly expanding literature, while leaving open the question of whether this growth corresponds to deeper theoretical consolidation, methodological diversification, or practical implementation in medical education settings.

3.3. Source Distribution and Disciplinary Location

This analysis reports the main publication venues in which learning analytics research in medical and health professions education has appeared. Figure 3 presents the most relevant publication sources in the dataset. The highest number of publications appeared in *Medical Teacher* (n = 13), followed by *Academic Medicine* (n = 10), *BMC Medical Education* (n = 10), and *JMIR Medical Education* (n = 8). Other contributing sources included *Journal of Continuing Education in the Health Professions* (n = 5), *Teaching and Learning in Medicine* (n = 5), *Anatomical Sciences Education* (n =

4), *Medical Science Educator* (n = 4), *Frontiers in Education* (n = 3), and *Frontiers in Medicine* (n = 3). The distribution by source reveals that research on learning analytics in the domain of medical and health professions education is not confined to journals specialised in learning analytics, educational data mining or educational technology. A substantial proportion of the published research on this topic appears in well-established medical and health professions education journals. It could be claimed that the dialogue about learning analytics is happening in forums that are concerned with medical teaching, clinical training, assessment, feedback and health professions education. At the same time, the field is not concentrated in one dominant journal. Of the 177 publications, the journal with the highest output, Medical Teacher, accounted for only 13. The dispersion of the pattern indicates that medical education learning analytics has gained disciplinary visibility but is still scattered across different publication spaces including medical education, educational technology, assessment, simulation, artificial intelligence and other health professions education spaces. Hence, the distribution of sources points to a plural and still coalescing publication landscape rather than a tightly formed journal core.

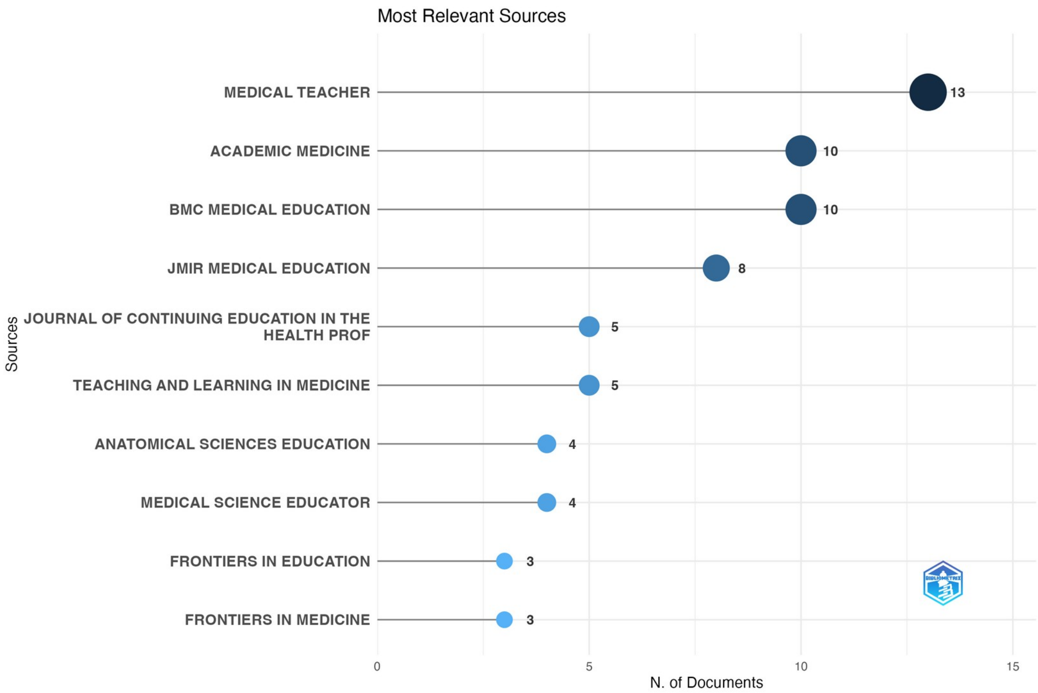

Figure 3. Most relevant sources.

This source pattern has two implications: (1) the presence of journals (from ascending to descending) indicates that learning analytics has acquired visibility in recognised medical education publication channels. (2) The relatively small number of publications per source indicates that the field remains distributed and still consolidating. The data, therefore, entail learning analytics seems to be increasing its bibliometric presence in medical and health professions education journals, but source distribution alone does not determine the extent of its embeddedness in curricula, assessment systems, or clinical training practices.

3.4. Authorship Productivity and Field Consolidation

This following analysis examines the distribution of author productivity and the extent to which publications are concentrated among recurrent contributors. Figure 4 presents the distribution of author productivity according to Lotka's Law (29) and shows whether publication activity is concentrated among recurrent contributors. The pattern was very skewed. Most authors contributed only one publication, and the proportion of authors declined sharply with increasing number of publications. The dataset showed a small number of recurrent contributors, with multiple publications. The fitted Lotka model showed a strong correspondence with the observed productivity distribution, as indicated by the high R^2 value (.9629) and the non-significant Kolmogorov–Smirnov test for the fitted model ($p = .938$). The estimated beta is 3.19 which is faster than the classical inverse-square pattern. This means, in practice, that the authorship productivity in the mapped literature is concentrated in a relatively small number of repeat contributors, while most authors have a single appearance. The author productivity pattern therefore reflects a broad but unevenly consolidated research community, with many occasional contributors and a smaller group of recurrent authors.

This pattern suggests a broad but uneven consolidation. The fact that there are so many contributors shows that learning analytics in medical and health professions education is something that researchers from all kinds of different backgrounds are interested in. But there aren't that many people writing regularly, which suggests that the field hasn't quite settled into a big and active group of authors yet. This could be because it's a topic that covers lots of different fields, like medical education, educational technology, assessment, simulation, data science, artificial intelligence and health professions education. But Lotka's Law (29) is more about how much work academics produce, not how good the ideas in the studies are, how well they're done, how relevant they are to teaching, or what kind of impact they have.

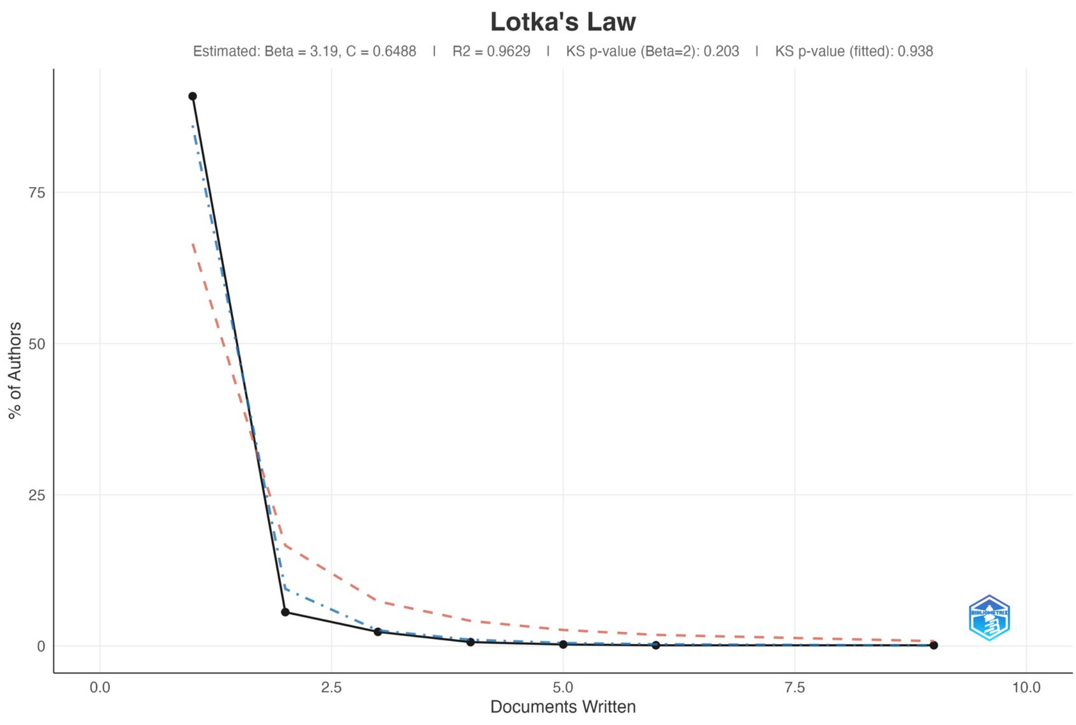

Figure 4. Author productivity distribution according to Lotka's Law

3.5 Geographical Distribution of Reaserch Output

This analysis presented here presents the country-level distribution of publications and identifies the geographical concentration of research output in the field. Figure 5 illustrates the geographical distribution of publications on learning analytics in medical and health professions education. The map illustrates that research production is globally distributed but unevenly concentrated. The largest contributor to the dataset is the United States, followed by Canada, the United Kingdom, several European countries, China, Australia and some countries in the Asia-

Pacific region. This shows that research on learning analytics in medical and health professions education is more visible in countries with established infrastructures for research in medical education, digital learning, educational technology and data science. North America, Western Europe, Australia and China are so prominent that this could simply reflect differences in research funding, the capacity of their institutions, digital education infrastructure, the extent to which they publish in English and whether they are included in journals that are indexed around the world. But we should remember that learning analytics and data-informed educational practices are not absent from Africa, South America, Central Asia and parts of the Middle East. But actually, it's showing lower representation in the analysed dataset indexed in Web of Science in this study. So, the country distribution is more like a map of how visible publications are than a complete picture of global learning analytics activities in medical and health professions education.

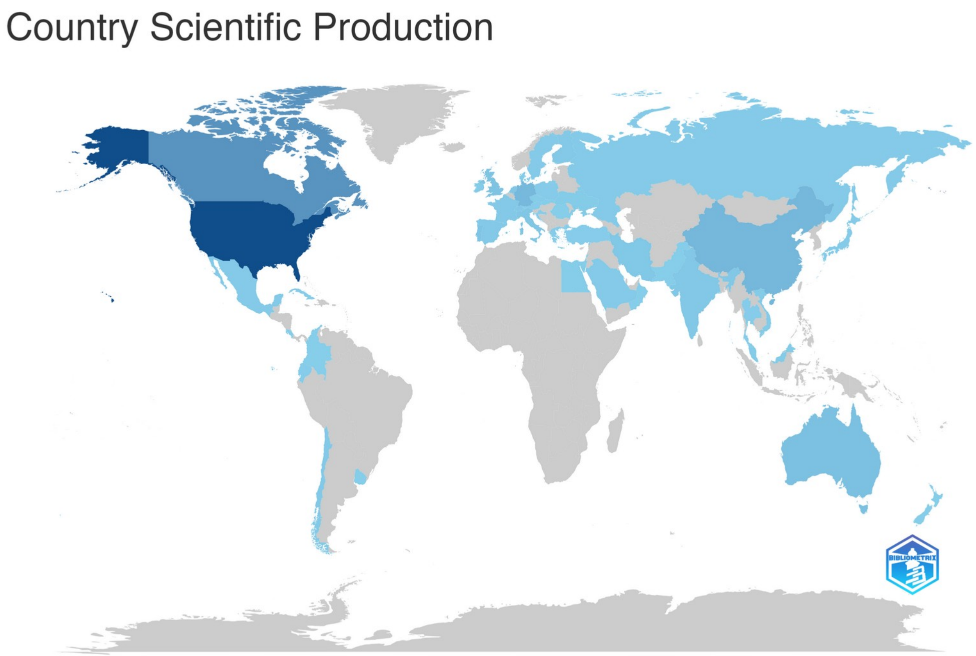

Figure 5. Country production in learning analytics research in medical education.

3.6. Intellectual and Conceptual Linkages

This analysis shows how cited references, authors, and merged keywords are connected within the field. Figure 6 visualises the relationships among cited references, authors, and merged keywords. The left field represents cited references, the middle field represents authors, and the right field represents the main merged keywords. The width of the links indicates the relative strength of the connections among these elements. As the three field plot shows, the literature is not organised around a single intellectual source, group of authors or thematic line. Rather, it leverages knowledge from disciplines such as learning analytics, educational data mining, medical education, assessment and feedback, simulation-based learning, competence development, technology-enhanced learning, big data and machine learning. There are a number of authors that appear to be key links to the references they cite and the keywords. Among them are Saqr M, Thoma B, Nouri J, Pusic M, Pusic MV, Lajoie SP, Hamstra SJ, Martinez-Maldonado R, Doleck T, Ferreira MA, Gašević D, Park YS. Their presence should be treated as an indicator of their bibliometric visibility in the analysed records and not a direct measure of their scholarly influence, methodological quality or educational impact.

The keyword field also makes clear what the literature thinks. The terms most frequently linked were “learning analytics”, “education”, “medical education”, “performance”, “collaborative learning”, “students”, “competence”, “assessment”, “technology”, “big data”, “educational data mining”, “analytics”, “feedback”, “simulation-based learning” and “machine learning”. These terms suggest that the field is about the combination of data on learners, educational performance, assessment, feedback, simulation, technology and competence related stuff. But these links only show how ideas are related within the Web of Science indexed literature. These don't demonstrate a cause and effect, or that anything's been put into practice, or that assessment practice, competence development or feedback has improved.

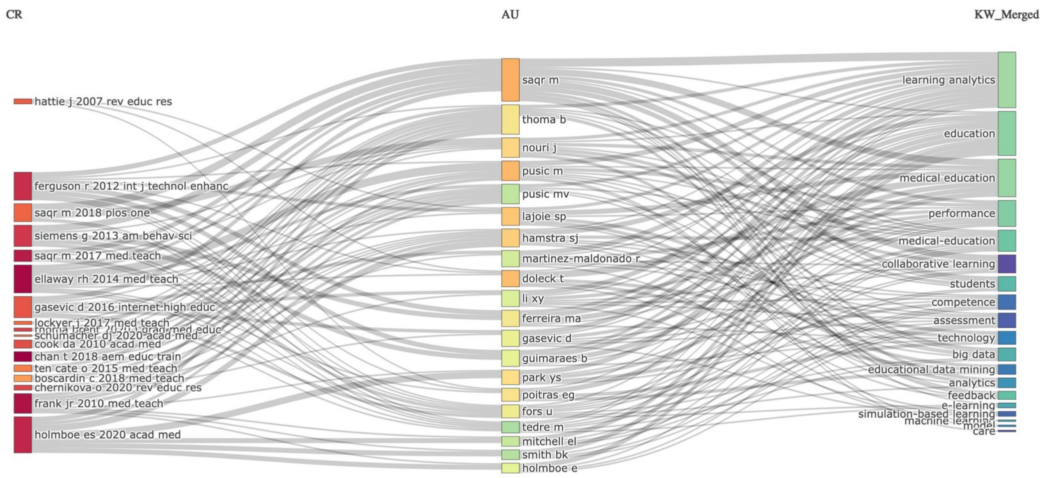

Figure 6. Three-field plot of cited references, authors, and merged keywords.

3.7. Thematic Structure

Looking at the most common keywords and used something called thematic mapping to see determines how medical and health professions education research is organised. The supplementary figures S1 and S2 show the keyword co-occurrence network and the thematic map. These analyses help to explain the connections between the main ideas in the dataset and how specific themes fit into the wider context of the field.

The keyword co-occurrence network analysis and thematic mapping show that “learning analytics” functions as a central term in the literature, as shown in Supplementary Figure S1. It is linked with “education”, “medical education”, “performance”, “students”, “assessment”, “feedback”, “competence”, “simulation-based learning”, “technology”, “big data”, “educational data mining”, and “machine learning”. This pattern suggests that the field is not organised around a single conceptual focus, but is shaped by overlapping concerns related to data-informed educational analysis, learner performance, assessment and feedback, simulation-based learning, competence research, and AI- or technology-supported educational environments.

The thematic map provides a second layer of interpretation by positioning themes according to centrality and density. Themes related to learning analytics, education, performance, competence, feedback, students, artificial intelligence, big data, and technology appear as relatively central elements of the indexed literature. By contrast, programmatic assessment, workplace-based assessment, entrustable professional activities, clinical skills, and time appear in more peripheral or weakly connected positions. This is relevant because these concepts are central to competency-based medical education and longitudinal assessment.

This finding should be interpreted cautiously. The peripheral location of programmatic assessment, workplace-based assessment, and entrustable professional activities does not demonstrate that these practices are absent from medical education or that learning analytics is not used in local assessment systems. Rather, it indicates that these concepts are less visible and less strongly connected within the Web of Science-indexed literature included in this analysis. Overall, the thematic structure points to a field that is most visibly associated with performance, feedback, assessment, competence, simulation, technology, artificial intelligence, big data, educational data mining, and machine learning, while its weaker connection with longitudinal and workplace-based assessment terminology identifies a potential area for future research rather than evidence of implementation failure.

4. Discussion

This study mapped the research structure and thematic patterns of learning analytics in medical and health professions education using 177 Web of Science-indexed publications published between 2014 and 2026. The results indicate a recent and rapidly expanding literature, a dispersed source structure, a collaborative but unevenly consolidated authorship pattern, and an internationally distributed but geographically uneven publication landscape. The conceptual analyses further suggest that the mapped literature is organised around learning analytics, medical education, performance, feedback, assessment, competence, simulation-based learning, artificial intelligence, big data, educational data mining, and machine learning. These findings should be interpreted as bibliometric and thematic patterns within the indexed literature, not as direct evidence of educational effectiveness, implementation quality, or changes in medical education practice.

4.1. Growth, Visibility, and Field Consolidation

The temporal pattern suggests that publication output remained relatively modest between 2014 and 2020, increased after 2021, and accelerated from 2024 onwards. In bibliometric terms, this pattern indicates a growing scholarly attention to learning analytics in medical and health professions education (24,25). This growth is associated with several overlapping developments including the expansion of digital and hybrid learning environments, simulation-based training, e-portfolios, AI-supported tools, and data-informed educational decision-making in health professions education (27,28). Driven perhaps by this convergence of digitalisation, assessment reform, and AI-related educational innovation, the post-2021 increase may reflect a widening research interest in data-informed approaches to medical education. However, publication growth should not be equated with conceptual and intellectual maturity or educational implementation. The field remains relatively recent, as reflected in the low average document age, and its publication activity is distributed across a wide range of sources. Although journals such as *Medical Teacher*, *Academic Medicine*, *BMC Medical Education*, and *JMIR Medical Education* appear as prominent venues, no single journal dominates the dataset.

The authorship pattern supports a similar interpretation. The dataset shows a highly collaborative publication structure, but the Lotka's Law distribution indicates that most authors contributed only one publication, while a smaller group appeared as recurrent contributors (29). This suggests a broad but unevenly consolidated research community. In practical terms, learning analytics in medical and health professions education appears to attract researchers from multiple domains, but it has not yet stabilised around a large and sustained core of highly productive authors. This is not a weakness in itself; rather, it is characteristic of a field still forming at the intersection of several disciplines

4.2. Thematic Associations and Disciplinary Positioning

The source distribution and three-field plot suggest that learning analytics is gaining bibliometric visibility embedded within medical and health professions education rather than remaining only a technical topic within educational data science. The appearance of established medical education and health professions education journals among the most relevant publication sources suggests that learning analytics is being discussed in venues regarding with to educational problems that matter directly to medical educators, such as feedback, assessment, clinical training, learner progression, simulation, and competence development. This is consistent with earlier arguments analytics in health professions education becomes meaningful when it is connected to educational purposes, professional contexts, and local decision-making needs (28).

The conceptual structure also points to a hybrid field. Terms such as "learning analytics", "medical education", "performance", "assessment", "feedback", "students", "competence", "simulation-based learning", "technology", "big data", "educational data mining", and "machine learning" show that the literature connects technical, pedagogical, and domain-specific concerns. This pattern suggests that the field is not organised around a single agenda. Instead, it appears to be shaped by multiple entry points: data-informed educational analysis, learner performance, assessment and feedback, simulation-based learning, AI-supported analytics, and competence-related research.

Nevertheless, these findings require careful interpretation. Keyword co-occurrence and thematic mapping show thematic proximity within publications; they do not demonstrate that these concepts are integrated in actual curricula, assessment systems, or clinical training practices. For example, the association between learning analytics, feedback, assessment, and competence indicates that these concepts are discussed together in the literature. It does not show whether learning analytics has improved feedback quality, strengthened assessment validity, or supported competence development in practice. This distinction is central to interpreting the present study. The dataset maps research about learning analytics; it does not analyse learner data or evaluate learning analytics interventions.

4.3. Methodological and Educational Implications

The conceptual structure of the mapped literature has methodological and educational implications, but these should be interpreted within the limits of bibliometric evidence. Terms such as "performance", "assessment", "feedback", "competence", "simulation-based learning", "artificial intelligence", "big data", and "machine learning" appear in close relation to learning analytics. This indicates that the field is frequently discussed in connection with educational monitoring, assessment, technology-supported learning, and data-informed decision-making. However, these associations do not demonstrate that learning analytics has improved feedback quality, assessment validity, competence development, or educational outcomes in practice.

One implication is the need for more theory-informed and transparent learning analytics research. Prior scholarship has emphasised that learning analytics should be grounded in learning theory and should not reduce learning to easily measurable behavioural traces (4-7). This is particularly important in medical and health professions education, where learning is situated, longitudinal, and closely connected to clinical judgement. Future studies should therefore report more explicitly how analytic indicators are selected, linked to educational constructs, validated, and translated into feedback, supervision, remediation, or curriculum-level decisions.

AI-supported analytics also requires cautious methodological development. The mapped literature shows visible associations among artificial intelligence, big data, technology, machine learning, and learning analytics, consistent with broader developments in medical education (19-23). This growing attention is also reflected in recent medical education studies addressing AI use in undergraduate clinical training, ChatGPT-supported pedagogy, and medical students' perceptions of AI tools (36-38). Nevertheless, AI-related visibility in bibliometric maps should not be read as evidence of effectiveness. Predictive or automated systems can support educational decision-making only when their outputs are explainable, valid, fair, pedagogically actionable, and ethically governed.

A further implication concerns the weaker bibliometric connection between learning analytics and longitudinal assessment architectures. Programmatic assessment, workplace-based assessment, and entrustable professional activities are central to competency-based medical education and clinical progression decisions (10,30-33), yet they appear less strongly connected within the mapped thematic structure. This pattern should be understood as a possible research gap, not as evidence that these practices are absent from medical education or unused in local assessment systems. Future research could examine how learning analytics is connected to programmatic assessment, workplace-based assessment, entrustment decisions, clinical supervision, competency committees, and curriculum-level decision-making. Recent work on learning analytics in clinical competency committees illustrates the potential of this direction while showing that its connection with competency-based decision-making remains developing (34).

4.4. Limitations

This study has several limitations. First, the analysis was based only on records retrieved from the Web of Science Core Collection. Although Web of Science provides structured bibliographic metadata and is widely used in bibliometric research, it does not include all relevant publications in medical education, health professions education, learning analytics, educational technology, computer science, or data science (35). Studies indexed only in Scopus, PubMed, ERIC, IEEE Xplore, Google Scholar, or other databases may therefore have been excluded. The findings should consequently be interpreted as a map of Web of Science-indexed literature, not as a complete representation of the global field.

Second, bibliometric and science-mapping analyses depend on the quality and consistency of database metadata. Variations in author names, affiliations, source titles, keywords, and indexing practices may affect the results. Although keyword cleaning and harmonisation were conducted, some terminological fragmentation may remain, especially for closely related terms such as "learning analytics", "educational data mining", "AI", "artificial intelligence", "medical education", and "medical-education". Citation and productivity indicators should also be interpreted cautiously because they are influenced by publication age, journal visibility, disciplinary norms, and database coverage (24,25).

Third, the visual outputs used in this study describe bibliometric and conceptual relationships, but they do not assess the quality, effectiveness, or practical impact of individual studies. The three-field plot, keyword co-occurrence network, and thematic map help identify intellectual linkages and thematic associations, but they cannot determine whether learning analytics improves learning outcomes, feedback processes, assessment validity, competence development, supervision, or curriculum decision-making. Finally, because the dataset included publications up to 2026 and was retrieved on 22 June 2026, the 2026 publication year was incomplete. Annual production and recent thematic patterns for 2026 should therefore be interpreted as provisional.

5. Conclusions

- The bibliometric evidence directly indicates that the field of learning analytics has expanded rapidly, particularly after 2021, and that the literature is distributed across a broad range of publication sources. The results also show a collaborative but unevenly consolidated authorship structure, an internationally distributed but geographically uneven publication landscape, and a conceptual structure associated with learning analytics, medical education, performance, feedback, assessment, competence, simulation-based learning, artificial intelligence, big data, educational data mining, and machine learning.
- The interpretation of these findings should remain within the limits of bibliometric evidence. The prominence of terms related to feedback, assessment, competence, simulation, artificial intelligence, and data-informed decision-making suggests that these themes are visible within the indexed literature. However, this does not demonstrate that learning analytics has already improved feedback quality, assessment validity, competence development, or educational outcomes in medical education practice. Similarly, the weaker bibliometric connection with programmatic assessment, workplace-based assessment, and entrustable professional activities should be interpreted as a thematic gap in the indexed literature rather than as direct evidence of absence or weak implementation in educational settings.
- The take-home message is that learning analytics in medical and health professions education is a developing research field that includes multiple stakeholders, but more research is required to understand how it can support education. Learning analytics should be explored in future studies within medical education. We need more research, particularly on the links between learning analytics, feedback, assessment validity, learner agency, clinical supervision, competency-based decision-making, entrustment, and curriculum-level educational improvement.

Funding: There was no funding.

Declaration of conflict of interest: The authors declare that they have no conflict of interest.

Author contributions: M.K. contributed to conceptualisation, study design, data analysis, interpretation, writing, and revision. A.B. contributed to conceptualisation, methodological review, interpretation, critical revision, and final approval. Both authors read and approved the final version of the manuscript.

6. References.

1. Long P, Siemens G. Penetrating the fog: Analytics in learning and education. *EDUCAUSE Rev*. **2011**, 46(5), 31-40. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education
2. Ferguson R. Learning analytics: Drivers, developments and challenges. *Int J Technol Enhanc Learn*. **2012**, 4(5/6), 304-317. https://doi.org/10.1504/IJTEL.2012.051816
3. Siemens G. Learning analytics: The emergence of a discipline. *Am Behav Sci*. **2013**, 57(10), 1380-1400. https://doi.org/10.1177/0002764213498851
4. Gašević D, Dawson S, Siemens G. Let's not forget: Learning analytics are about learning. *TechTrends*. **2015**, 59(1), 64-71. https://doi.org/10.1007/s11528-014-0822-x
5. Wise AF, Shaffer DW. Why theory matters more than ever in the age of big data. *J Learn Anal*. **2015**, 2(2), 5-13. https://doi.org/10.18608/jla.2015.22.2
6. Knight S, Buckingham Shum S. Theory and learning analytics. In Lang C, Siemens G, Wise AF, Gašević D, editors. *Handbook of Learning Analytics*. Society for Learning Analytics Research; **2017**. p. 17-22. https://doi.org/10.18608/hla17.027

7. Lodge JM, Knight S, Kitto K. Theory and learning analytics, a historical perspective. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 3-18. https://doi.org/10.1007/978-3-031-60571-0_1
8. Frank JR, Snell L, Englander R, Holmboe ES. Implementing competency-based medical education: Moving forward. *Med Teach*. **2017**, 39(6), 568-573. https://doi.org/10.1080/0142159X.2017.1315069
9. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR, ICBME Collaborators. A call to action: The controversy of and rationale for competency-based medical education. *Med Teach*. **2017**, 39(6), 574-581. https://doi.org/10.1080/0142159X.2017.1315067
10. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. **2011**, 33(6), 478-485. https://doi.org/10.3109/0142159X.2011.565828
11. Wise AF. Collaborative learning theory and analytics. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 105-124. https://doi.org/10.1007/978-3-031-60571-0_7
12. Clow D. The learning analytics cycle: Closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Association for Computing Machinery; **2012**. p. 134-138. https://doi.org/10.1145/2330601.2330636
13. Sailer M, Ninaus M, Huber S, Bauer E, Greiff S. The end is the beginning is the end: The closed-loop learning analytics framework. *Comput Human Behav*. **2024**, 158, 108305. https://doi.org/10.1016/j.chb.2024.108305
14. Dever DA, Cloude EB, Wiedbusch MD, Azevedo R. Emotion theory and learning analytics: A theoretical framework for capturing emotion regulation using process data. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024.** p. 125-137. https://doi.org/10.1007/978-3-031-60571-0_8
15. Giannakos M. Theory and intermediate-level knowledge in multimodal learning analytics. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 87-104. https://doi.org/10.1007/978-3-031-60571-0_6
16. Nguyen A, Järvelä S. Learning analytics framework for analysing regulation in collaborative learning (FARCL). In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 71-85. https://doi.org/10.1007/978-3-031-60571-0_5
17. Poquet O, Chen B. Integrating theories of learning and social networks in learning analytics. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 139-151. https://doi.org/10.1007/978-3-031-60571-0_9
18. Raković M, Bannert M, Molenaar I, Winne PH, Gašević D. In conversation: Bannert, Molenaar, & Winne—Multiple perspectives on researching and supporting self-regulated learning via analytics. In Bartimote K, Howard SK, Gašević D, editors. *Theory Informing and Arising from Learning Analytics*. Springer; **2024**. p. 57-69. https://doi.org/10.1007/978-3-031-60571-0_4
19. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Med Educ*. **2019**, 5(1), e13930. https://doi.org/10.2196/13930
20. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, Hanson J, Haas M, Spadafore M, Grafton-Clarke C, Gasiea RY, Michie C, Corral J, Kwan B, Dolmans D, Thammasitboon S. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach*. **2024**, 46(4), 446-470. https://doi.org/10.1080/0142159X.2024.2314198
21. Masters K. Artificial intelligence in medical education. *Med Teach*. **2019**, 41(9), 976-980. https://doi.org/10.1080/0142159X.2019.1595557
22. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. *Med Teach*. **2023**, 45(6), 574-584. https://doi.org/10.1080/0142159X.2023.2186203
23. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. **2019**, 5(2), e16048. https://doi.org/10.2196/16048

24. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr*. **2017**, 11(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007
25. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *J Bus Res*. **2021**, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070
26. Lazarides MK, Lazaridou I-Z, Papanas N. Bibliometric analysis: Bridging informatics with science. *Int J Low Extrem Wounds*. **2025**, 24(3), 515-517. https://doi.org/10.1177/15347346231153538
27. Bojic I, Mammadova M, Ang CS, Teo WL, Diordieva C, Pienkowska A, Gašević D, Car J. Empowering health care education through learning analytics: In-depth scoping review. *J Med Internet Res*. **2023**, 25, e41671. https://doi.org/10.2196/41671
28. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. *Med Teach*. **2014**, 36(3), 216-222. https://doi.org/10.3109/0142159X.2014.874553
29. Lotka AJ. The frequency distribution of scientific productivity. *J Wash Acad Sci*. **1926**, 16(12), 317-323.
30. van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach*. **2012**, 34(3), 205-214. https://doi.org/10.3109/0142159X.2012.652239
31. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. **2007**, 29(9-10), 855-871. https://doi.org/10.1080/01421590701775453
32. ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. **2005**, 39(12), 1176-1177. https://doi.org/10.1111/j.1365-2929.2005.02341.x
33. ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? *Acad Med*. **2007**, 82(6), 542-547. https://doi.org/10.1097/ACM.0b013e31805559c7
34. Carney PA, Sebok-Syer SS, Pusic MV, Gillespie CC, Westervelt M, Goldhamer MEJ. Using learning analytics in clinical competency committees: Increasing the impact of competency-based medical education. *Med Educ Online*. **2023**, 28(1), 2178913. https://doi.org/10.1080/10872981.2023.2178913
35. Birkle C, Pendlebury DA, Schnell J, Adams J. Web of Science as a data source for research on scientific and scholarly activity. *Quant Sci Stud*. **2020**, 1(1), 363-376. https://doi.org/10.1162/qss_a_00018
36. Bonilla Mejia JJ, Cortés Fuenzalida TD, Polanco Aliaga DH, Martínez Carrillo C, Herrera Alcaíno ÁA. Artificial intelligence in the practical clinical training of undergraduate medical students: a scoping review of applications, outcomes, and gaps. *Spanish Journal of Medical Education*. 2026, 7(3). https://doi.org/10.6018/edumed.710071
37. Jiménez Salcedo SC, Granda Cruz CA. Use of ChatGPT as a pedagogical tool in medical education: systematic review of the literature (2020-2025). *Spanish Journal of Medical Education*. **2026**, 7(4). https://doi.org/10.6018/edumed.716681
38. Cortés-Fraile L, Suescun-Elizalde D, García-Estañ J. Use and perception of artificial intelligence among Spanish medical students. *Spanish Journal of Medical Education*. **2026**, 7(2). https://doi.org/10.6018/edumed.704241