

REVIEW

Artificial intelligence in educational assessment
in the age of generative AI: A bibliometric review

Premalatha Perumal^{1*}, Asokan Vasudevan¹, Soon Eu Hui¹, Ganesan Palanisamy², and Aruna Dev Roy³

¹Faculty of Business and Communications, INTI International University, Negeri Sembilan, Malaysia

²Kalasalingam Academy of Research and Education, Kalasalingam Academy of Research and Education, India

³Royal School of Commerce, Royal Global University, Guwahati, Assam, India

Abstract

The rapid diffusion of artificial intelligence (AI), particularly generative AI tools, has significantly reshaped educational assessment practices, creating new opportunities and challenges for feedback, academic integrity, and curriculum design. Despite the growing volume of scholarship, research in this area remains fragmented, making it difficult to discern dominant trends, key contributors, and emerging themes. This study employs a bibliometric review to map the intellectual structure and evolution of research on AI in educational assessment between 2015 and 2025. Using open-access journal articles indexed in the Dimensions.ai database and aligned with Sustainable Development Goal 4 (Quality Education), a curated dataset of 89 studies was analysed. Bibliometric techniques, including co-authorship, citation, co-citation, and keyword co-occurrence analyses, were applied using VOSviewer and descriptive statistics. The findings reveal a sharp growth in publications following the emergence of generative AI, with influential clusters focusing on assessment redesign, feedback, academic integrity, and higher education applications. While established institutions and authors dominate the field, collaboration networks remain fragmented, and contributions from non-Western contexts are comparatively limited. The study highlights the need for stronger international collaboration and context-sensitive research to support equitable and responsible integration of AI in educational assessment.

Keywords: Artificial intelligence; Generative AI; Educational assessment; Bibliometric analysis; Higher education; Academic integrity; Sustainable Development Goal 4

***Corresponding author:**

Premalatha Perumal
(premalatha.perumal@newinti.edu.my)

Citation: Perumal, P., Vasudevan, A., Hui, S. E., Palanisamy, G. & Roy, A. D. (2026). Artificial intelligence in educational assessment in the age of generative AI: A bibliometric review. *Int J Systematic Innovation*, 10(3): 026060012.
[https://doi.org/10.6977/IJoSI.202606_10\(3\).0001](https://doi.org/10.6977/IJoSI.202606_10(3).0001)

Received: February 8, 2026

Revised: March 26, 2026

Accepted: April 15, 2026

Published online: June 29, 2026

Copyright: © 2026 Author(s).

This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Artificial intelligence (AI) has rapidly made its way into classrooms. Over the past decade, its role in assessment has grown, making it impossible to overlook. AI can grade work, give feedback and create tasks that adjust to student progress. At the same time, concerns over plagiarism, cheating and the authenticity of student work are equally pressing (Chiu, 2024; Lye & Lim, 2024). These developments highlight both the promise and the risks of AI in education. Though the research on this topic has grown rapidly, the field remains scattered. Many studies overlap, and it is unclear which ideas or

authors have the greatest influence. Bibliometric analysis can address this issue. It maps research patterns, affiliations between authors and institutions and highlights the main themes (Donthu *et al.*, 2021). This analysis offers a clearer picture of how the field's knowledge base has developed. AI in assessment also connects to global education policy, in particular Sustainable Development Goal (SDG) 4, which emphasises inclusive and equitable quality education for all (UNESCO, 2023). Moreover, much of the current debate comes from Western, well-resourced systems. This raises questions about how far the findings apply to other contexts.

The purpose of this study is to use bibliometric analysis to map publication trends, identify key authors and institutions, and examine thematic developments in the use of AI in educational assessment. By analysing citation patterns and thematic clusters, the study provides critical insights into how scholarly knowledge in the field is being constructed and advanced. The articles between 2015 and 2025 were selected deliberately to capture the full lifecycle of contemporary AI integration in educational assessment. Machine-learning applications began to enter mainstream educational discourse around 2015, marking a transition from experimental analytics toward more applied AI systems. Furthermore, the endpoint (2025) captures the inflexion point associated with the rise of large language models between 2022 and 2023. Therefore, the 10-year window reflects a coherent technological and pedagogical cycle rather than an arbitrary convenience-based boundary. The scope covers education, curriculum and pedagogy, education systems and policy, and sociology and psychology. Three research questions guide the study: (i) What publication trends appear in this period? (ii) Who are the most active authors and institutions? (iii) Which countries contribute most?

Additionally, this study contributes in three ways. First, it clarifies the structure of research on AI in assessment. Second, it links these patterns to debates on integrity, feedback, and curriculum. Third, it places the discussion within the global policy frame of SDG4. Together, these points highlight what is known and what still needs attention.

2. Methodology

This study used a bibliometric approach to examine how AI has been framed in educational assessment. The bibliometric method was chosen because it provides a systematic approach to map research output, authorship, collaboration, and thematic direction. Although bibliometric analysis does not assess the depth or quality of individual contributions, it can reveal patterns of

knowledge production. Additionally, it can highlight areas of growth, influence, and fragmentation in AI in assessment (Donthu *et al.*, 2021).

2.1. Data collection

Dimensions (Digital Science, United Kingdom; <https://www.dimensions.ai/>) was used to construct the dataset for this study. It was selected for its broad coverage of peer-reviewed research and its ability to filter results by SDGs and disciplinary domains. To align with SDG 4 (Quality Education), the search was confined to four research fields: Education; Curriculum and Pedagogy; Education Systems and Policy; and Sociology and Psychology. SDG 4 (Quality Education) seeks to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (UNESCO, 2023). Its relevance to AI-driven assessment lies in its emphasis on equity, inclusion, accessibility and fairness in learning systems. As AI technologies increasingly mediate grading, feedback and academic progression, questions arise about algorithmic bias, digital divides and cultural inclusivity. Framing the dataset within SDG4, therefore, ensures alignment with global policy imperatives and situates AI assessment research within broader sustainability and equity discourses rather than purely technological innovation narratives. In this context, the search terms were defined as: “artificial intelligence” AND “assessment” AND “education.”

The timeframe of articles was limited to 2015–2025 to capture both early developments and the surge of research following the emergence of generative AI. To maintain accessibility and integrity, only open-access journal articles were included. The present analysis excluded book chapters, conference papers, and editorials. The initial search returned 4,218 records. Through filters for open access and article type, the number was reduced to 177. Abstract and keyword screening further narrowed the set to 105. Finally, a manual review and removal of non-English or irrelevant studies resulted in a curated dataset of 89 articles. The process is demonstrated in the PRISMA flow diagram (Figure A1).

2.2. Data analysis

Multiple tools were employed to ensure rigour and reproducibility. VOSviewer (version 1.6.19, Centre for Science and Technology Studies, Leiden University, Netherlands) was selected for its strength in visualising bibliometric networks through distance-based mapping and clustering techniques. Unlike general statistical tools, VOSviewer applies the visualisation of similarities mapping technique, which enables the identification of intellectual clusters, thematic proximity and collaboration density with high interpretative clarity (Van Eck & Waltman, 2018).

Its suitability for co-authorship, co-citation and keyword co-occurrence analyses makes it particularly appropriate for mapping emerging interdisciplinary domains such as AI in educational assessment.

Microsoft Excel (Microsoft Corporation, United States of America) supported the calculation of publication trends and institutional frequencies. Dimensions.ai was further used to extract citation counts and verify author affiliations. Combining these tools allowed for both quantitative analysis (e.g., citation and keyword networks) and qualitative interpretation (e.g., thematic shifts and institutional collaboration patterns).

3. Findings

3.1. Publication trends

Between 2015 and 2021, research on AI in educational assessment remained minimal and confined to small-scale or domain-specific studies. A noticeable increase began in 2019, followed by a sharp rise in 2023 as generative AI tools, such as ChatGPT, entered educational debates. In 2024, publications accelerated further, reflecting both practical applications and growing ethical concerns. By 2025, output peaked, accounting for more than half of all studies in the dataset. As shown in Figure 1, this timeline illustrates a clear transition from limited studies in the

early years to widespread, systemic and ethically focused research in the most recent period.

3.2. Key authors: Most prolific contributors

The analysis identified eight authors with two or more publications in AI in educational assessment research from 2015 to 2025. Kohnke, Lucas, and Khlaif, Zuheir N., emerged as the most prolific contributors, each with three publications. They are followed by Zou, Di; Ahmed, Vian; Alkouk, Wejdan Awadallah; Bahroun, Zied; Ou, Amy Wanyu; and Chiu, Thomas K. F., each contributing two publications (see Table 1).

These authors represent the leading voices in the field, consistently contributing to research on AI-driven grading frameworks, teacher perceptions of generative AI and related assessment technologies. Their repeated contributions indicate sustained engagement with key issues in AI-based educational assessment, making them central to the development of scholarly discourse in this domain.

3.3. Key institutions: Most prolific contributors

The institutional analysis highlights a clear clustering of research productivity (see Table 2). Universities such as RMIT University and the Education University of Hong Kong emerge as central players, producing a steady stream

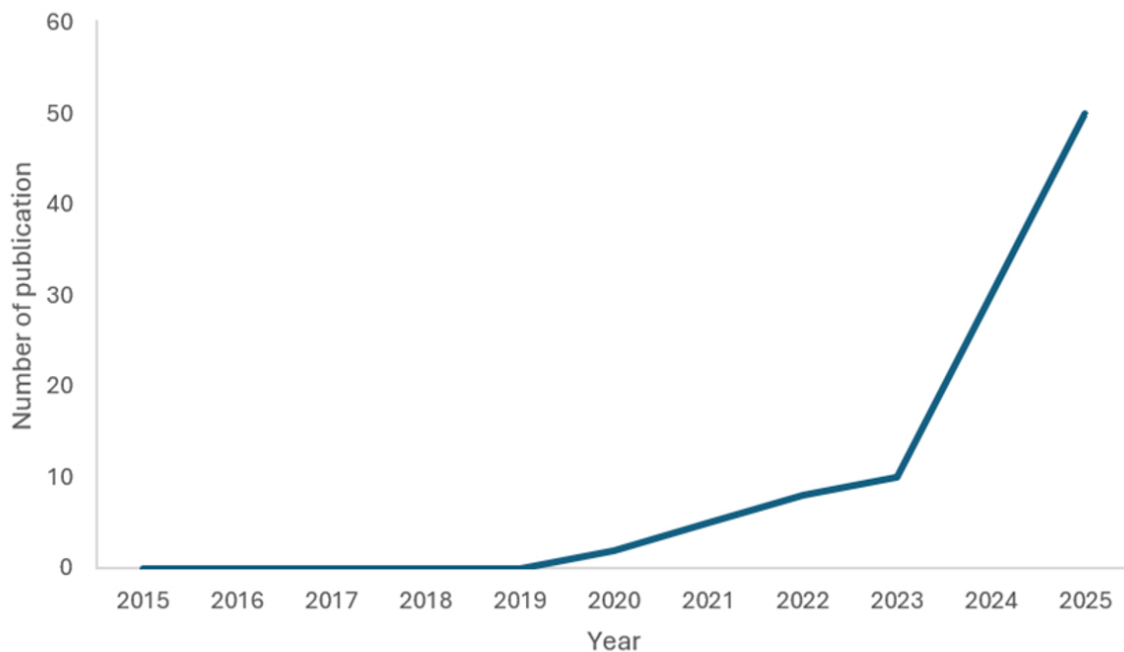


Figure 1. Publication trends in artificial intelligence and educational assessment research (2015–2025). The dataset analysis shows growth in publications over time.

of publications while maintaining dense collaboration networks. Similarly, An-Najah National University emerges as a regional leader in the Palestinian Territories, strengthening its profile by consistently contributing to the field despite resource and geopolitical constraints; this finding highlights how institutions outside the traditional academic hubs are emerging as visible contributors to the field. Stockholm University, although positioned slightly apart, plays a distinctive role as a bridge-builder, linking clusters and facilitating the circulation of knowledge across regions. In contrast, Dhofar University, though geographically distant from the main hub, forges strong international partnerships and positions itself as an emerging contributor to the global conversation. These patterns suggest that while established Western and Asian institutions continue to dominate output, newer regional contributors are forging significant roles through collaboration, ensuring that the discourse on AI in assessment is not confined wholly to traditional academic centres.

Table 1. Most prolific authors in research on artificial intelligence in educational assessment (2015–2025)

Author name	Documents	Link strength
Kohnke, Lucas	3	3
Khlaif, Zuheir N.	3	2
Zou, Di	2	3
Ahmed, Vian	2	2
Alkouk, Wejdan Awadallah	2	2
Bahroun, Zied	2	2
Ou, Amy Wanyu	2	2
Chiu, Thomas K. F.	2	0

Note: Link strength refers to the total link strength in the co-authorship network generated in VOSviewer.

The VOSviewer visualisation reinforces this picture by illustrating the collaborative dynamics among institutions (see Figure 2). Here, the red cluster highlights the strong interconnections between RMIT University and the Education University of Hong Kong, while the green cluster shows Dhofar University’s bridging role, and the blue cluster depicts Stockholm University as a connector across networks. The map makes it clear that institutional influence is about publication volume and the effectiveness of collaboration networks.

3.4. Keyword co-occurrence

Building on the institutional and authorship patterns highlighted earlier, it is equally important to examine the thematic directions that shape this body of research. The keyword co-occurrence analysis (Figure 3) reveals the conceptual clusters and interconnections that underpin the literature. This offers insights into the dominant themes and the way they evolve across different contexts.

The first cluster (red) centres on “students,” “teachers,” “assessment,” “ChatGPT,” and “AI tools.” This grouping reflects the practical dimension of the literature, with a strong emphasis on how generative AI tools are reshaping assessment practices, classroom dynamics and teacher-student interactions.

The second cluster (green) is anchored around “higher education,” “teaching,” “applications,” and “reviews.” This indicates a focus on institutional-level engagement, where AI integration is examined not only in classroom contexts but also in broader systemic adoption across higher education.

The third cluster (blue) connects “education,” “generative AI,” and “educators.” Unlike the other clusters, this theme appears to represent more conceptual and theoretical perspectives, where discussions extend beyond immediate application to situate AI within disciplinary discourse and educational philosophy.

Across the map, “education” and “assessment” emerge as central nodes, functioning as bridges that connect the different thematic clusters. Similarly, “ChatGPT” occupies a linking position between the practical concerns of assessment and the wider pedagogical debates in higher education. These interconnections suggest that the literature not only addresses immediate challenges in teaching and learning but also situates them within ongoing debates about the future direction of education in the era of AI.

3.5. Citation and co-citation analysis

The citation mapping (see Figure 4) highlights a small but significant cluster of highly influential works that anchor contemporary debates on the role of AI in educational assessment. The largest nodes, such as those representing Lim *et al.* (2023), Chiu (2023), and Khlaif *et al.* (2024, 2025), correspond to publications with the highest citation counts, indicating their centrality within the field. Among these, “Generative AI and the Future of Education: Ragnarök or Reformation?” by Lim *et al.* (2023) stands out with substantial influence, framing both the transformative potential and risks of generative AI in higher education. Similarly, Chiu’s (2023) “The Impact of

Table 2. Most prolific institutions in research on artificial intelligence and educational assessment (2015–2025)

Institution	Country/Region	Publications	Notes (specialisation/cluster)
Stockholm University	Sweden	5	Digital learning, ed-tech
An-Najah National University	Palestinian Territories	3	Regional leadership
Chinese University of Hong Kong	Hong Kong	3	Higher education focus
Education University of Hong Kong	Hong Kong	3	Pedagogy and assessment
RMIT University	Australia	3	Cross-continental collaborations
University of Wollongong	Australia	2	Applied educational tech
Birzeit University	Palestinian Territories	2	Local/regional contribution
Arizona State University	United States of America	2	Artificial intelligence and assessment integrity

Note: The full list of institutions is provided in [Table A1](#).

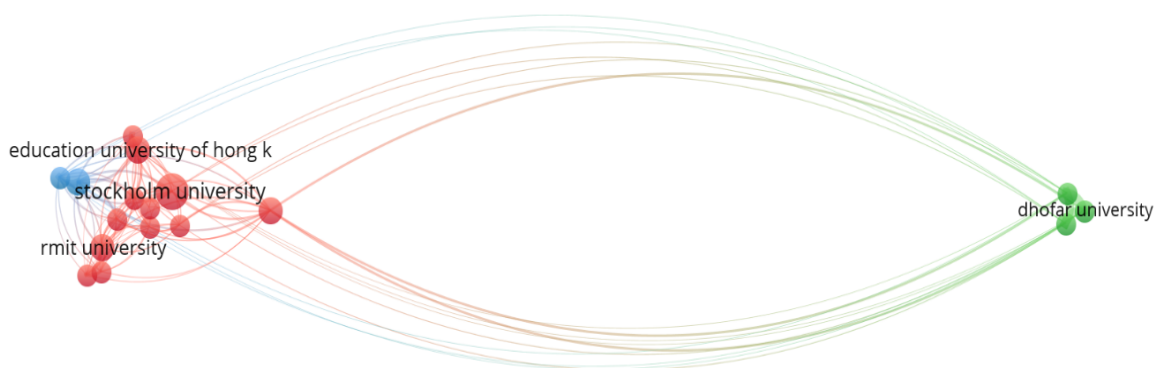


Figure 2. Institutional collaboration network of artificial intelligence and educational assessment research (2015–2025)

Note: Node size represents the number of publications; connecting lines indicate inter-institutional collaborations.

Generative AI on Practices, Policies and Research Direction in Education: A Case of ChatGPT and Midjourney” serves as a foundational work in mapping opportunities, risks, and implications for policy and practice. Another pivotal publication, “Redesigning Assessments for AI-Enhanced Learning: A Framework for Educators in the Generative AI Era” by Khlaif *et al.* (2025), underscores the systemic and long-term implications of AI integration into educational systems.

Other influential contributions include Ou *et al.*'s (2024) exploration of AI-powered language tools in academic communication and Bahroun *et al.*'s (2023) bibliometric review of generative AI in education, both of which appear as strong, well-connected nodes within the network. At the mid-tier level, studies such as Lye and Lim's (2024) principles for redesigning assessments with generative AI represent more specialised narratives. In the map, these

appear as smaller nodes clustered around the dominant authors, exemplifying the gradual diversification of research interests into specific pedagogical, disciplinary and ethical dimensions. Overall, the citation patterns suggest that a handful of landmark publications dominate scholarly attention and set the conceptual frame of the field; however, an expanding layer of mid-tier studies is diversifying the field.

The co-citation analysis further reveals the intellectual underpinnings of this research. Unlike co-authorship networks, which highlight collaboration patterns, co-citation networks show which scholars are most frequently cited together. This indicates a shared intellectual influence and conceptual alignment pattern. As shown in [Figure 5](#), a prominent cluster includes David Boud, Phillip Dawson, Joanna Tai, Rola Ajjawi, and Margaret Bearman. Their work on assessment, feedback literacy and

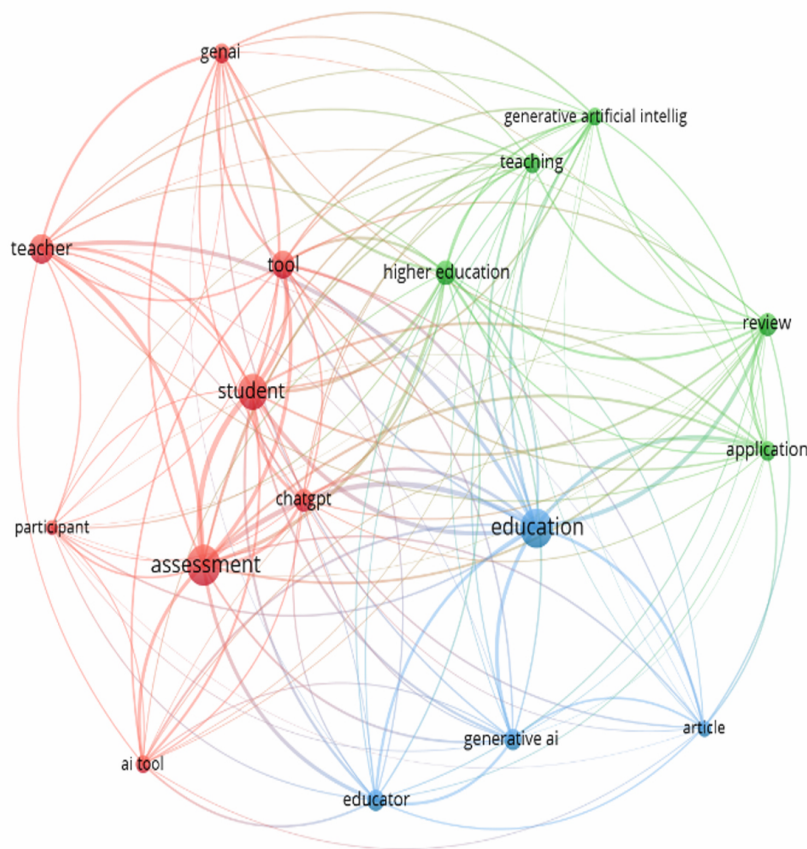


Figure 3. Keyword co-occurrence network in artificial intelligence and educational assessment research (2015–2025)
 Note: Node size reflects keyword frequency; connecting lines indicate co-occurrence relationships; line thickness shows strength of association.

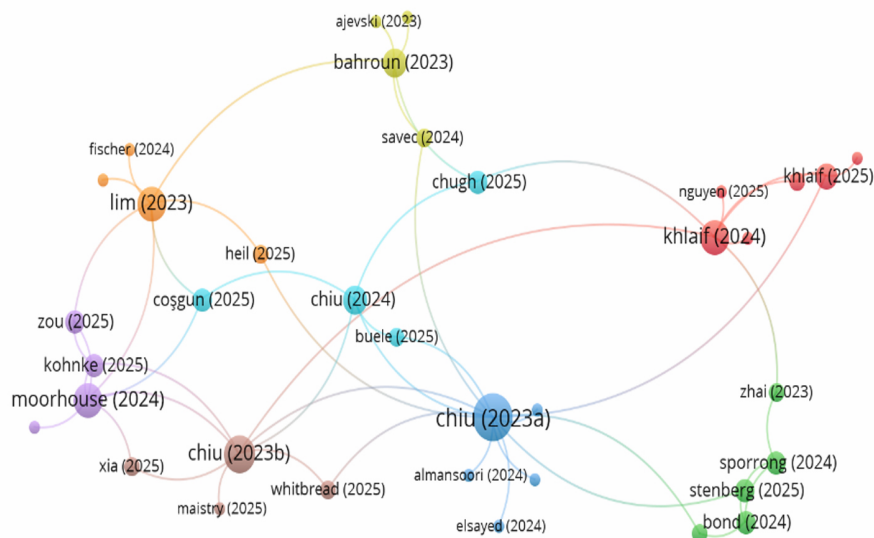


Figure 4. Most cited documents network in artificial intelligence and educational assessment research (2015–2025)
 Note: Node size reflects citation frequency; thicker links indicate stronger citation connections.

curriculum design continues to shape how generative AI is understood and applied in educational contexts. Although not the most prolific contributors to AI-specific studies, their frameworks remain central to the way the field conceptualises assessment and feedback.

The map illustrates that these scholars form a tightly connected cluster, suggesting that debates on AI in assessment build upon long-standing pedagogical traditions. Boud and Dawson's contributions on assessment futures and integrity resonate strongly with AI-related concerns, while Tai and Ajjawi's focus on feedback literacy intersects with the potential of AI to provide adaptive, personalised feedback. Bearman further bridges curriculum design and educational innovation, reinforcing links between foundational assessment theory and contemporary shifts. Together, this co-citation cluster demonstrates that AI in assessment does not emerge in isolation but is grounded in robust intellectual traditions that continue to guide new explorations of its pedagogical role.

Taken together, the co-citation map of foundational assessment scholars and the citation network of recent AI-focused studies illustrate a clear trajectory: long-standing theories of assessment and evaluative judgment provide the conceptual grounding, while newer research on generative AI is shaping the current and future agenda in educational assessment.

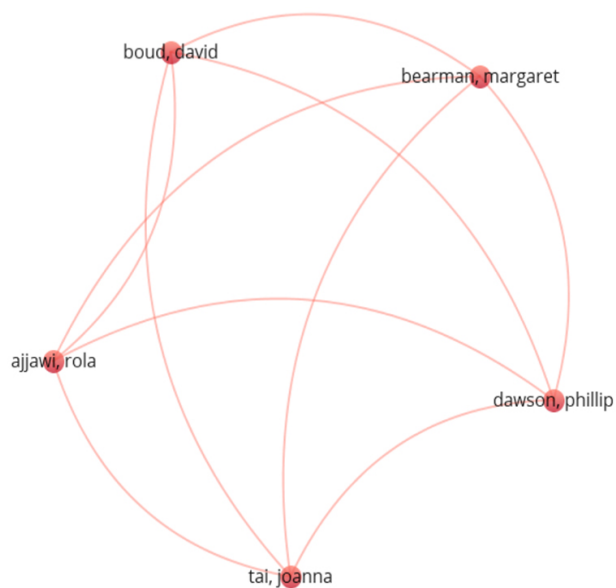


Figure 5. Co-citation network of influential authors in artificial intelligence and educational assessment research (2015–2025)
 Note: Node size reflects frequency of citation; connecting lines indicate co-citation relationships; line thickness shows strength of intellectual association.

3.6. Co-authorship networks: Visualising collaboration between authors

The co-authorship analysis highlights how leading scholars in AI and educational assessment collaborate. Unlike citation and co-citation patterns, which capture intellectual influence, co-authorship networks reveal the social and collaborative structures of knowledge production.

As shown in Figure 6, collaboration is relatively limited, with only a small cluster of active co-authors emerging. Within this cluster, Kohnke, Lucas; Zou, Di; and Ou, Amy Wanyu are closely linked, reflecting repeated joint work and a productive research partnership that has made a meaningful contribution to the field. By contrast, other prolific authors such as Khlaif, Zuheir N.; Ahmed, Vian; Alkhouk, Wejdan Awadallah; Bahroun, Zied; and Chiu, Thomas K. F. appear disconnected in the visualisation, reflecting independent publication patterns or collaboration with partners not captured in this dataset.

Overall, the network suggests a fragmented scholarly landscape. Rather than a broad, interconnected community, the field is characterised by isolated efforts and small teams. While this independence allows for diverse perspectives, it also risks slowing the diffusion of innovative practices and restricting the exchange of expertise. Strengthening collaboration across authors and institutions will therefore be essential to develop a more cohesive and impactful body of knowledge.

4. Interpretive synthesis of bibliometric findings

The study shows that research on AI in educational assessment from 2015 to 2025 has evolved through distinct phases, reflecting not only technological advancements but also shifting pedagogical practices, ethical considerations, and policy reforms that have shaped its implementation. The four chronological phases were derived from observable shifts in publication frequency, keyword density and citation centrality, indicating meaningful transitions in the field's development. Specifically, the period 2015–2019 represents a low-volume exploratory stage; 2020–2022 reflects a formative refinement stage marked by stabilised terminology; 2023–2024 corresponds to an expansion stage linked to institutional AI adoption; and 2025 constitutes a generative AI disruption phase characterised by exponential publication growth and the emergence of intensive ethical discourse. These divisions represent empirically identifiable inflexion points rather than arbitrary chronological segmentation.

(a) Early work (2015–2019)

Bibliometric mapping shows that publications in this

period were scarce and concentrated in specialised fields such as medical education and learning analytics. For example, Chan and Zary (2019) demonstrated technical feasibility but did not extend into pedagogical discussions. At this stage, AI was framed as an add-on to existing practices rather than a transformative force of assessment reform.

(b) Expansion (2020–2022)

From 2020 onwards, bibliometric clusters revealed a broader scope, with frequent co-occurrence of terms such as applications, teaching and reviews (Figure 3). This shift reflects a growing institutional and systemic engagement with AI in assessment, which was moving beyond pilot projects toward grading systems and predictive analytics. The literature supported this trend. For example, Ifenthaler and Yau (2020) examined AI's role in educational decision-making, while UNESCO's AI and Education report (Miao *et al.*, 2021) emphasised fairness, inclusivity and policy alignment. Collectively, these works suggest that studies were not confined to classroom practice but had begun to influence curriculum design and governance at this stage. However, while enthusiasm for innovation was high, the adoption was uneven due to concerns over equity, transparency and access.

(c) Generative artificial intelligence surge (2023–2024): Ethical and pedagogical turning point

The launch of ChatGPT served as a major turning point, driving an increase in research volume and triggering wider theoretical discussion. Citation analysis points to works such as Chiu (2024) and Lim *et al.* (2023) as central to discussions on personalised feedback, authenticity and academic integrity. At the same time, UNESCO (2023) warned that adoption was outpacing governance, raising concerns over plagiarism, inequity and cultural bias. This phase represents a shift from questions of whether AI can be used in assessment to how it should be implemented responsibly.

(d) Consolidation and diversification (2025): Ethical integration and specialisation

By 2025, research on AI in educational assessment reached peak productivity and thematic diversification. This research field has branched into AI literacy, AI-resistant assessments, fairness in grading and cross-cultural perspectives (Alkouk & Khlaif, 2024; Kohnke *et al.*, 2025). Bibliometric co-citation mapping (Figure 5) reveals that, despite these new directions, debates remain grounded in established assessment theory. Foundational scholars such as Boud, Dawson, Tai, Ajjawi, and Bearman continue to shape how feedback, curriculum and evaluative judgement are conceptualised in the AI era (Bearman *et al.*, 2024).

Nevertheless, co-authorship networks (Figure 6) reveal limited collaboration, suggesting a fragmented field where innovative practices risk being siloed.

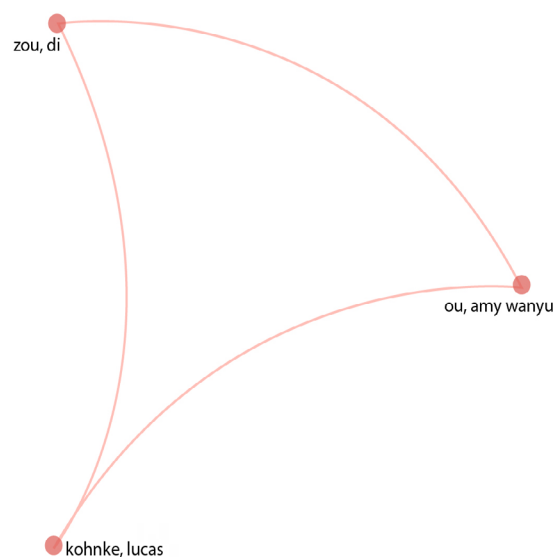


Figure 6. Co-authorship network of authors in artificial intelligence and educational assessment research (2015–2025)

Note: Node size represents the number of publications; connecting lines indicate co-authorship ties; line thickness reflects collaboration strength.

Overall, the bibliometric analysis highlights four broad phases in the literature. The exploratory phase (2015–2017) tested feasibility, followed by a formative phase (2018–2019) that refined early applications. An expansion phase (2020–2022) then engaged with institutional and policy concerns, before a diversification phase (2023–2025) shaped by generative AI. Collectively, these phases show how research has progressed from technical trials to broader discussions on governance, ethics and inclusion.

At the same time, the evidence reveals persistent gaps. The authorship patterns show fragmented networks, with limited collaboration across institutions and disciplines. Keyword clusters confirm the dominance of Western perspectives, while contributions from emerging economies remain underrepresented. Taken together, these findings suggest that the field is growing in scale and scope but has not achieved the integration and diversity needed for global relevance.

5. Discussion

The objective of this discussion is to move beyond

reporting the findings to offering a critical interpretation of their significance for the future of educational assessment. It explores the evolution pattern of this field by combining a bibliometric approach with insights from the wider literature to understand the emerging scope and implications of AI in assessment.

The findings from this study demonstrate that research on AI in educational assessment has developed through a series of distinct but overlapping phases. Early studies (2015–2019) were exploratory and technical, reflecting feasibility studies in domains such as medical education (Chan & Zary, 2019). Between 2020 and 2022, bibliometric evidence shows a shift toward institutional engagement and policy alignment, from works such as Ifenthaler and Yau (2020) and Miao *et al.* (2021). The literature linked AI to systemic questions of fairness and inclusivity, which represents a broadening of scope from classroom applications to governance and curriculum design.

However, the generative AI surge from 2023 to 2024 marked a decisive turning point. Citation mapping highlights Chiu (2024) and Lim *et al.* (2023) as shaping debates on integrity, personalised feedback and responsible adoption. UNESCO (2023) underscored the risks of rapid implementation outpacing governance, raising global concerns about equity and cultural bias. By 2025, the literature diversified into AI literacy, fairness in grading and cross-cultural perspectives (Alkouk & Khlaif, 2024; Kohnke *et al.*, 2025). Most importantly, co-citation analysis revealed that these new directions remain grounded in established theoretical traditions, particularly feedback literacy, evaluative judgement and curriculum design (Bearman *et al.*, 2024; Boud & Dawson, 2021). The evolution of AI in educational assessment can also be interpreted through the concept of maieutics. This concept emphasises the co-construction of knowledge through conversational exchange. Maieutic systems differ from purely artificial control mechanisms by fostering adaptive, interactive and co-constructed learning processes. In educational contexts, generative AI tools increasingly function as dialogic partners that scaffold student thinking rather than simply evaluating outcomes. This reflects a shift in educational aims from summative measurement toward formative, inquiry-based learning. In this sense, AI-supported assessment becomes a maieutic process in which feedback emerges through iterative interaction between learner and intelligent system, promoting metacognition, self-regulation, and deeper conceptual understanding. Therefore, the integration of maieutic principles advances traditional assessment theory by positioning AI as a facilitator for dialogic and learner-centred pedagogical transformation and not merely as an

automation tool.

Nevertheless, one of the notable gaps lies in regional representation. While Asia is present through contributions from Hong Kong, Singapore, and China, Malaysia is notably absent from the bibliometric mapping despite being an active participant in exploring digital transformation in education. The absence of Malaysian scholarship signals a missed opportunity to contextualise AI adoption within local assessment practices, cultural expectations and policy frameworks. Given Malaysia's emphasis on aligning education with the aspirations of the National Education Blueprint and SDG4, future research must address this gap to ensure that the global discourse on AI in assessment does not remain dominated by Western and East Asian perspectives.

Two limitations are acknowledged in the current study. First, restricting the dataset to open-access articles may have excluded influential work in subscription-based journals. Second, bibliometric analysis emphasises scale and connectivity, rather than the depth or nuance of individual studies. Nevertheless, these limitations do not undermine the study's value. By integrating bibliometric mapping with a targeted qualitative review, this methodology provides a transparent, reproducible and comprehensive account of how AI in educational assessment has evolved between 2015 and 2025.

6. Conclusion and future directions

This study demonstrates that AI in educational assessment has evolved from a peripheral technological enhancement into a structurally disruptive force that is reshaping the epistemology of evaluation itself. While early research focused on feasibility and automation, the emergence of generative AI has shifted the discourse toward questions of responsibility, ethics and the redefinition of assessment practices within complex educational systems.

The findings further reveal that, despite rapid expansion, the field remains fragmented, with limited collaboration across regions and a persistent dominance of Western-centric perspectives. This fragmentation constrains the development of coherent, globally relevant frameworks and highlights the need for more inclusive and context-sensitive scholarship, particularly from underrepresented regions such as Malaysia.

Importantly, the evolution of AI in educational assessment can be interpreted as a dynamic socio-technical process rather than a linear progression. The interaction between technological innovation, institutional policy, pedagogical adaptation and ethical critique reflects a continuous feedback loop that shapes the direction of

the field. In this sense, AI does not simply transform assessment practices but co-evolves with the systems that govern education.

This transformation can be further understood through the lens of control system theory. The maieutic perspective is consistent with contemporary educational goals that emphasise dialogic learning, personalised feedback and learner agency. This frames AI as a facilitator of knowledge co-construction rather than a predefined assessment tool. The progression of AI in assessment reflects a shift from artificial control systems, characterised by automated grading and rule-based evaluation, toward more maieutic forms of interaction, where generative AI enables dialogic, adaptive and co-constructed feedback processes. This shift redefines assessment from a mechanism of measurement to a process of guided learning, where control is no longer imposed but negotiated through human–AI interaction.

Considering these findings, several directions for future research are proposed. There is a clear need for more cross-regional and comparative studies to address the current imbalance in global research representation and to ensure that AI-driven assessment practices are relevant across diverse educational contexts. Future research should also prioritise ethical governance, with a focus on algorithmic fairness, transparency and accountability in AI-supported assessment systems.

Additionally, the development of AI literacy and capacity-building frameworks for educators and students is essential to support informed and responsible use. Research should explore how such competencies can be embedded within curriculum design and professional development initiatives. There is also a pressing need for context-specific studies, particularly within countries such as Malaysia, which are underrepresented, to better understand local challenges, policy alignment and implementation strategies.

Finally, future research should adopt a more integrated, systems-oriented perspective that considers the interaction between technology, pedagogy and governance. Such an approach will be critical in advancing the field beyond isolated innovations toward more sustainable, equitable and socially impactful models of AI in educational assessment.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Premalatha Perumal, Asokan Vasudevan

Formal analysis: Premalatha Perumal, Asokan Vasudevan, Aruna Dev Rroy

Investigation: Premalatha Perumal, Asokan Vasudevan, Soon Eu Hui

Methodology: Premalatha Perumal, Ganesan Palanisamy

Writing–original draft: Premalatha Perumal

Writing–review & editing: Asokan Vasudevan, Soon Eu Hui, Ganesan Palanisamy, Aruna Dev Rroy

Availability of data

Not applicable.

References

- Alkoul, W. A., & Khlaif, Z. N. (2024). AI-resistant assessments in higher education: Practical insights from faculty training workshops. *Frontiers in Education, 9*, 1499495. <https://doi.org/10.3389/educ.2024.1499495>
- Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability, 15*(17), 12983. <https://doi.org/10.3390/su151712983>
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893–905. <https://doi.org/10.1080/02602938.2024.2335321>
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering, 137*, 85–86.
- Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education, 5*(1), e13930. <https://doi.org/10.2196/13930>
- Chiu, T. K. F. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments, 32*(10), 6187–6203. <https://doi.org/10.1080/10494820.2023.2253861>
- Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative

- AI. *Computers and Education: Artificial Intelligence*, 6, 100197.
<https://doi.org/10.1016/j.caeai.2023.100197>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296.
<https://doi.org/10.1016/j.jbusres.2021.04.070>
- 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>
- Khlaif, Z. N., Alkhouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2), 174.
<https://doi.org/10.3390/educsci15020174>
- Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14(10), 1090.
<https://doi.org/10.3390/educsci14101090>
- Kohnke, L., Zou, D., Ou, A. W., & Gu, M. M. (2025). Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration. *Computers and Education: Artificial Intelligence*, 8, 100398.
<https://doi.org/10.1016/j.caeai.2025.100398>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790.
<https://doi.org/10.1016/j.ijme.2023.100790>
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317–323.
- Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences*, 14(6), 569.
<https://doi.org/10.3390/educsci14060569>
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. United Nations Educational, Scientific and Cultural Organization.
<https://doi.org/10.54675/PCSP7350>
- Ou, A. W., Stöhr, C., & Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. *System*, 121, 103225.
<https://doi.org/10.1016/j.system.2024.103225>
- Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 25(4), 348–349.
- Tai, J., Ajjawi, R., Bearman, M., Boud, D., Dawson, P., & Jorre de St Jorre, T. (2023). Assessment for inclusion: Rethinking contemporary strategies in assessment design. *Higher Education Research & Development*, 42(2), 483–497.
<https://doi.org/10.1080/07294360.2022.2057451>
- UNESCO. (n.d.). Monitoring SDG 4. *Global Education Monitoring Report*. Accessed November 28, 2025. <https://www.unesco.org/gem-report/en/monitoring-sdg4>
- Van Eck, N. J., & Waltman, L. (2018). *VOSviewer manual* (Version 1.6.10). Leiden University. Accessed September 29, 2025. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.10.pdf
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

Appendix

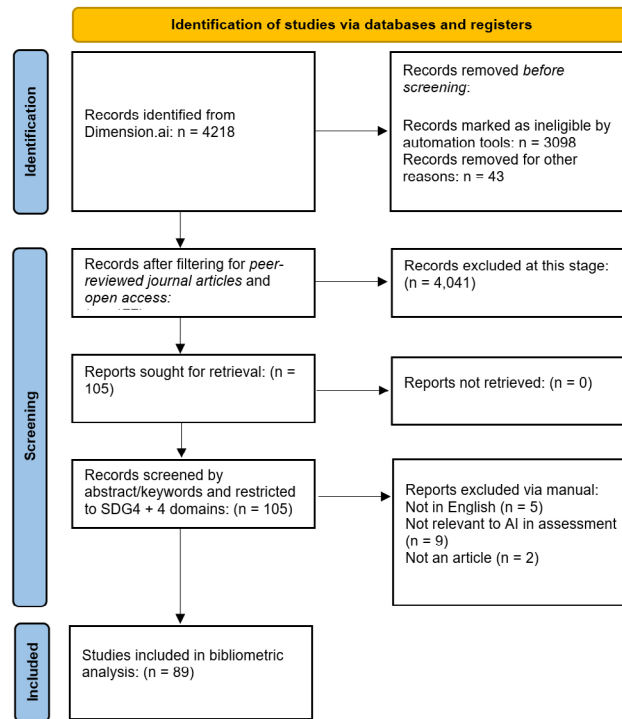


Figure A1. PRISMA flow diagram of data selection (2015–2025)

Table A1. Most prolific institutions in AI and educational assessment research (2015–2025)

Institution	Country/Region	Publications	Notes (specialisation/cluster)
Stockholm University	Sweden	5	Digital learning, ed-tech
An-Najah National University	Palestinian Territories	3	Regional leadership
Chinese University of Hong Kong	Hong Kong	3	Higher education focus
Education University of Hong Kong	Hong Kong	3	Pedagogy and assessment
RMIT University	Australia	3	Cross-continental collaborations
University of Wollongong	Australia	2	Applied educational tech
Birzeit University	Palestinian Territories	2	Local/regional contribution
Arizona State University	USA	2	AI and assessment integrity
Dhofar University	Oman	2	Regional interest
Iraqi University	Iraq	2	Assessment in HE
University of Babylon	Iraq	2	Local ed-tech studies
University of Warwick	UK	2	Assessment ethics and policy
Xi'an Jiaotong–Liverpool University	China–UK partnership	2	Cross-cultural collaboration
Monash University	Australia	2	AI and pedagogy
Arden University	UK	2	Online learning and AI
Hong Kong Polytechnic University	Hong Kong	2	Applied AI in HE
American University of Sharjah	UAE	2	AI in tertiary education

Abbreviations: AI: Artificial intelligence; UAE: United Arab Emirates; USA: United States of America; UK: United Kingdom.