Journal of Applied Science, Engineering, Technology, and Education *ISSN: 2685-0591 (online)*
Vol. 7 No. 2 (2025)
https://doi.org/10.35877/454RI.asci4126

Artificial Intelligence in Educational Measurement: A Bibliometric Review (1997 to 2024)

Nursohana Othman[a], Mohd Effendi Ewan Mohd Matore[b,*], & Farahiyah Wan Yunus[c]

[a]*Faculty of Education, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia.*
[b]*Research Centre of Education Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia*
[c]*Center for Rehabilitation and Special Needs Studies (ICaRehab) Occupational Therapy Programme Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300, Kuala Lumpur, Malaysia*

Abstract

Artificial Intelligence (AI) is playing an increasingly important role in transforming measurement practices across domains such as health sciences, diagnostics, and psychology by improving precision, speed, and adaptability. In the field of education, however, the integration of AI into assessment remains limited, despite its potential to support adaptive learning systems and personalized feedback. This bibliometric review investigates 921 peer-reviewed articles published between 1997 and 2024 to explore how AI is shaping the development of educational measurement. The study employs citation analysis, co-citation mapping, and keyword co-occurrence techniques to identify influential publications, reveal the intellectual structure of the field, and uncover thematic trends. The findings show a notable increase in research activity after 2019, largely driven by developments in machine learning, natural language processing, and big data applications. Foundational insights from health and computer sciences, including deep learning algorithms and tools such as Scikit learn, are increasingly being adapted for educational use. Three major thematic areas emerge: technical foundations of AI, cross-sector applications in diagnostics, and ethical and policy considerations for education. International collaboration is expanding, with leading contributions from the United States and China and rising involvement from Malaysia, Pakistan and Nigeria. The review underscores the need for responsible and inclusive AI applications in educational assessment.

Keywords: Artificial Intelligence; Bibliometric Analysis; Educational Measurement; Global Collaboration; Machine Learning

Received: 9 July 2025 Revised: 9 August 2025 Published: 31 August 2025

1. Introduction

Artificial Intelligence (AI) is rapidly transforming numerous domains, including health sciences, diagnostics and psychology, by enhancing precision, efficiency, and personalization. In education, however, the integration of AI in measurement and assessment remains an emerging frontier. AI holds strong potential to revolutionize educational measurement by enabling adaptive systems, personalized feedback, and real-time learning analytics that can deepen insights into student progress (Bulut et al., 2024). Despite this potential, research on AI-driven educational assessment remains fragmented, calling for a systematic examination of its development and thematic evolution.

This study responds to that need by conducting a bibliometric analysis of peer-reviewed literature on AI in educational measurement, published between 1997 and 2024. Specifically, it utilizes citation analysis, co-citation analysis and keyword co-occurrence analysis to (1) identify influential publications, (2) map the intellectual structure of the field, and (3) forecast emerging trends and research directions. These methods allow for a comprehensive and evidence-based understanding of the scholarly landscape, offering insights into how AI applications are shaping educational assessment practices.

This research aligns with the broader agenda of the Fourth Industrial Revolution (4IR), which advocates for the integration of digital technologies in education to produce future-ready learners (Masdoki et al., 2021; Sovey, Osman, & Matore, 2022). Rather than focusing narrowly on traditional academic outcomes, 4IR emphasizes the importance of

* Corresponding author.
E-mail address: effendi@ukm.edu.my

Journal of Applied Science, Engineering, Technology, and Education is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

multidimensional skillsets, including adaptability, collaboration, and problem-solving. In this context, AI-supported educational measurement offers promising avenues for evaluating these broader competencies through data-driven and responsive systems. By mapping the current research terrain, this study contributes to advancing educational measurement theory and practice in ways that are inclusive, interdisciplinary and aligned with evolving technological landscapes.

2. Trends in AI in Measurement Across Disciplines

Artificial Intelligence (AI) is transforming measurement and assessment across various fields, offering greater accuracy, adaptability and efficiency. In healthcare, AI has advanced diagnostics and personalized treatment. These technologies are now being adapted in education to identify student learning patterns, provide real-time feedback and enable differentiated instruction, leading to more inclusive and effective assessment (Bulut et al., 2024).

Recent advancements in artificial intelligence, including deep learning, natural language processing and affective computing, have transitioned from use in medicine and commerce to educational environments. These technologies are now embedded in tools such as Intelligent Tutoring Systems and gamified platforms, which support personalized learning and enhance student engagement (Zhao & Zhong, 2024). However, the adoption of these tools in education brings new challenges, including concerns over data privacy, algorithmic bias and alignment with sound pedagogical practices (Swargiary, 2024). Addressing these issues is essential for ensuring responsible integration of AI that promotes fairness and supports meaningful learning outcomes. Bibliometric analysis offers a structured way to evaluate research trends, identify influential studies, and uncover conceptual linkages in this evolving field. By applying methods such as citation analysis, co-citation mapping, and keyword clustering, this study reveals the intellectual structure of AI applications in educational measurement and highlights opportunities for ethical and student-centered innovation (Wang et al., 2024). These insights contribute to a broader understanding of how AI can be effectively and responsibly integrated into educational systems.

3. Methodology

3.1. Overview of Research Design

This study utilized a structured, three-stage bibliometric research design to analyze the development of artificial intelligence (AI) in educational measurement. The approach was designed to address three core research objectives: identifying influential works, uncovering intellectual structures, and forecasting emerging trends.

The research process involved:

(1) a systematic literature search using well-defined keywords and Boolean logic;
(2) application of transparent inclusion and exclusion criteria to refine the dataset; and
(3) a triangulated bibliometric analysis comprising citation analysis, co-citation analysis, and keyword co-occurrence mapping.

These methods were selected for their capacity to synthesize large volumes of literature and visualize the evolving intellectual landscape of interdisciplinary fields such as AI in education. Detailed procedures for each phase are outlined in Sections 3.2 to 3.4.

3.2. Stage 1: Data Search Strategy

The literature search was guided by a strategic combination of primary keywords and synonymous terms, drawn from prior studies, thesaurus resources and expert defined classifications. The search was structured around four core concepts: (1) artificial intelligence, (2) education, (3) measurement and (4) screening. Keyword selection followed validated taxonomies and was informed by established bibliometric frameworks (Ishak et al., 2024; Ismail et al., 2024; Yang et al., 2025)

a) For artificial intelligence, commonly used terms included “AI,” “machine learning,” and “deep learning.”
b) For education, the search incorporated terms such as “e-learning” and “educational technology.”
c) For measurement, terms like “assessment” and “evaluation” were included.
d) For screening, the search considered terms such as “tool development” and “diagnostic tools.”

The search was conducted using the “Topic” field in the Web of Science database, which indexes titles, abstracts and author-supplied keywords. This method improved the precision of the dataset by limiting the retrieval to content where the terms were contextually relevant. The complete Boolean structure of the search query is presented in Table 1, which outlines how the four conceptual categories were operationalized into a reproducible search strategy. Similar practices have been recommended in prior bibliometric reviews in educational and multidisciplinary fields (Ishak et al., 2024; Ismail et al., 2024).

An initial total of 938 articles was retrieved. After applying refinement criteria, including relevance, publication type (journal articles only) and language (English only), the final dataset was narrowed to 929 publications as of August 2024. These records provided the empirical foundation for the bibliometric analyses, which were performed using VOSviewer software. These tools have been widely adopted for their capacity to visualize citation networks and extract thematic patterns from large academic datasets (Husaeni & Nandiyanto, 2021; Taherdoost et al., 2025). The Web of Science Core Collection was selected due to its extensive indexing capabilities and its widespread use in bibliometric research. The database includes more than 74.8 million scholarly records from over 34,000 journals across 25 academic disciplines, making it well-suited for comprehensive and high-quality bibliometric analysis (Zhao et al., 2024).

Table 1. The search string

WOS	ALL=((((“artificial intelligence” OR ai OR “machine learning”) AND (“education*” OR “e-learning*” OR “education* technolog*”) AND (“measurement* “OR “assessment*” OR “evaluation* “) AND (“screen*” OR “tool development*” OR “diagnostic tool*”))))

3.3. Stage 2: Inclusion and Exclusion Criteria

The dataset was retrieved from the Web of Science (WoS) Core Collection, chosen for its rigorous indexing standards and high-quality peer-reviewed content (Donthu et al., 2021). The following inclusion and exclusion criteria were applied to ensure consistency, relevance and reproducibility (see Table 2).

a) Inclusion criteria: Articles were included if they were written in English, published between 1997 and 2024, appeared in peer-reviewed journals (excluding reviews, conference papers and books) and were in final publication status. The year 1997 was selected as the starting point due to early applications of machine learning in education during that period.

b) Exclusion criteria: non-English articles, non-journal literature (e.g., grey literature, editorials, conference abstracts) and in-press or incomplete publications were excluded.

The initial search using the keywords (“artificial intelligence” OR “AI”) AND (“educational measurement” OR “assessment” OR “diagnostic”) yielded 1,237 records. After duplicate removal and manual screening based on the above criteria, 921 articles were retained for final analysis. It is noted that the apparent decline in 2024 publications reflects indexing lag rather than an actual drop in research output for that year.

Table 2. Inclusion and Exclusion Criteria for Literature Selection

Criterion	Inclusion	Exclusion
Language	English	Non-English
Time line	1997 – 2024	Outside time range
Literature type	Journal (Article)	Conferences, Books, Reviews, Editorials
Publication Stage	Final publication	In Press, early access, or incomplete data

3.4. Stage 3: Data Analysis

The third stage involved conducting a systematic bibliometric analysis to identify influential studies, thematic structures and emerging trends within the field of AI-driven educational measurement. Drawing upon the refined dataset of 921 publications, this analysis was performed using VOSviewer version 1.6.20 software, a widely adopted tool for science mapping and bibliometric visualization (Husaeni & Nandiyanto, 2021; Jia et al., 2022). The analysis was structured around three core techniques in bibliometric research:

3.4.1. Citation Analysis

Citation analysis was used to determine the most influential publications in the dataset by assessing citation frequencies and identifying key authors, journals and documents. This method highlights high-impact works that have significantly shaped the research landscape, offering insight into the foundational literature in the domain (Ishak et al., 2024; Passas, 2024). Using VOSviewer's normalization algorithm, a citation network was visualized to reveal clusters of highly cited papers and their interrelationships, providing a quantitative measure of scholarly influence.

3.4.2. Co-citation Analysis

Co-citation analysis was applied to uncover the intellectual structure of the research field by examining how often pairs of publications are cited together in subsequent literature. This method enables the identification of seminal works and the conceptual linkages among them, helping to map the thematic backbone of the domain (Kumar et al., 2024; Passas, 2024). VOSviewer version 1.6.20 was used to construct co-citation maps, which grouped documents into clusters based on their co-citation frequencies, representing key research areas and schools of thought in AI-based educational measurement.

3.4.3. Keyword Co-occurrence Analysis

Keyword co-occurrence analysis was employed to uncover dominant research themes and forecast emerging trends within the field of artificial intelligence (AI) in educational measurement. This technique evaluates the frequency and co-occurrence of keywords appearing in article titles, abstracts and author-designated keyword lists. It is particularly effective for visualizing thematic evolution and identifying conceptual structures in bibliometric landscapes (Bulut et al., 2024; Ishak et al., 2024; Ismail et al., 2024; Passas, 2024).

Using VOSviewer, keywords were extracted and mapped based on co-occurrence frequencies. A minimum threshold of five keyword occurrences was set to ensure analytical relevance. In the resulting network visualizations, the bubble size represents keyword prominence, while the thickness of connecting lines indicates the strength of association between terms (Passas, 2024). These clusters provide a data-driven foundation for tracing research trajectories, guiding future investigations and understanding emerging directions in AI-enhanced assessment practices.

This approach complements the study's multi-faceted bibliometric framework. The three analytical layers used in this study include citation analysis, which represents the performance dimension; co-citation analysis, which reflects the intellectual structure; and keyword co-occurrence analysis, which highlights the conceptual structure. Together, these dimensions offer a comprehensive perspective on the development and interdisciplinary scope of AI in educational measurement. The alignment between the research objectives, strategies and analytical methods employed in this study is summarized in Table 3.

Table 3. Summary of the research question, strategies and bibliometric techniques

Research Objective	Description	Strategy to Address RO	Bibliometric Technique
RO1	To identify the most influential articles shaping the field of AI in educational measurement.	Analyze highly cited publications to determine those with the greatest impact across disciplines.	Citation Analysis
RO2	To explore the intellectual structure of the domain and cross-disciplinary linkages in AI applications for measurement.	Investigate frequently co-cited documents to reveal conceptual relationships and disciplinary integration.	Co-citation Analysis
RO3	To forecast emerging research trends in AI-enhanced educational assessment.	Map frequently co-occurring keywords to identify dominant themes and new research directions.	Keyword Co-occurrence Analysis

4. Results and Findings

4.1. Influential Studies in AI-Driven Educational Measurement (Citation Analysis)

This section addresses the first objective of the study, which is to identify and analyze the most influential publications that have shaped the research landscape of artificial intelligence (AI) in educational measurement. To achieve this, a citation-based bibliometric analysis was conducted using data extracted from the Web of Science Core Collection. This approach provides insights into key contributions and the extent of scholarly impact across the field.

4.1.1. Dataset Overview and Citation Metrics

The initial search process retrieved 938 articles. Following a systematic screening procedure that excluded review papers, book chapters, editorials and conference proceedings, a final dataset of 921 peer-reviewed journal articles was established (Table 4). This curated dataset served as the foundation for the citation analysis.

Descriptive citation metrics from the dataset are summarized as follows:

a) Total citations: 21,744
b) Citations excluding self-citations: 21,679
c) Average citations per article: 24.46
d) h-index: 59

These metrics indicate a robust level of scholarly engagement and reflect increasing academic attention to AI applications in educational measurement. The citation frequency across publications suggests that this is a developing field with significant interdisciplinary influence, drawing from domains such as data science, healthcare and education.

4.1.2. Temporal Distribution of Publications

An analysis of publication trends on AI in educational measurement reveals a consistent increase in scholarly output from 1997 to 2018, followed by a significant surge beginning in 2019. This sharp rise, illustrated in Figure 1, coincides with rapid advancements in AI technologies and a growing emphasis on personalized and data-driven educational innovation. The number of indexed publications peaked in 2023 at 200 articles, signaling widespread academic and practical adoption of AI tools in education. Although a decline is observed in 2024 (with only 34 articles), this figure reflects data up to August only and may not represent the full year's output. Overall, the trend highlights AI's expanding role in adaptive assessments, personalized feedback systems and performance analytics, with increasing focus on education-specific applications beyond its earlier roots in healthcare and psychology.

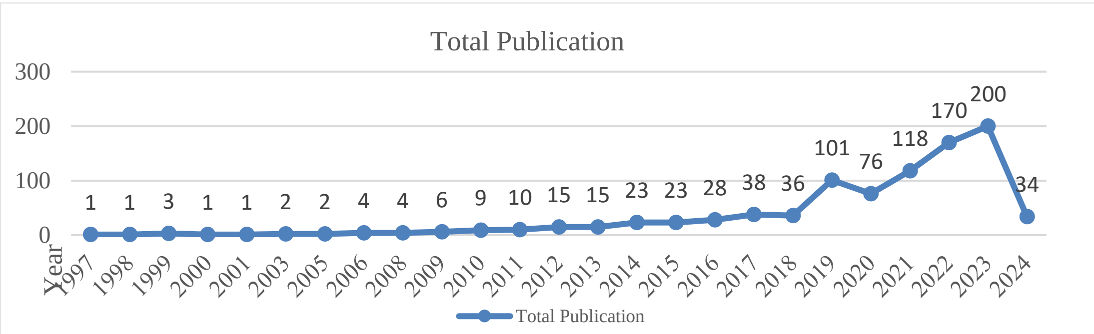

Fig. 1. Trend of Publications on AI in Educational Measurement (1997–2024).

**Note: 2024 data include publications up to August only*

4.1.3. Influential Publications

To identify the most impactful contributions in the field of AI-driven educational measurement, citation and co-citation analyses were employed. The analysis revealed five highly influential publications that have shaped the development of artificial intelligence applications in educational contexts, particularly in assessment, learning analytics and adaptive systems (Table 5).

Among the most cited is Falk et al. (2019), whose work on U-Net demonstrates the power of deep learning for image recognition and segmentation. Although developed within biomedical imaging, the methodology has had a significant methodological influence on educational technologies, especially in intelligent visual assessment systems. Xu and Ouyang (2022) emerged as a leading education-specific contribution. Their systematic review of AI's use in STEM education offers valuable perspectives on how artificial intelligence, encompassing machine learning, natural language processing and deep learning, is revolutionizing educational measurement. This includes advancements in automated grading, student modeling and personalized feedback systems.

Table 5. Top Cited and Co-Cited Studies in AI-driven Educational Measurement

Rank	Authors (Year)	Title / Source	Type	Citations	Link Strength
1	Falk et al. (2019)	U-Net: Deep Learning for Cell Counting and Morphometry – Nature Methods	Cited	947	–
2	Xu & Ouyang (2022)	AI Technologies in STEM Education: A Systematic Review – IJSTEM	Cited	208	–
3	Pedregosa et al. (2011)	Scikit-learn: Machine Learning in Python – JMLR	Co-Cited	36	46
4	Breiman (2001)	Random Forests – Machine Learning	Co-Cited	29	43
5	(Chawla et al., 2002)	SMOTE: Synthetic Minority Over-sampling Technique – Journal of Artificial Intelligence Research	Co-Cited	19	29

Note. *Type* indicates whether the article was highly cited or highly co-cited based on the analysis. *Link Strength* refers to how frequently the document was co-cited with others, indicating intellectual influence in the field.

The machine learning frameworks developed by Pedregosa et al. (2011) on Scikit-learn, Breiman (2001) on Random Forests and (Chawla et al., 2002) on Synthetic Minority Over-sampling Technique demonstrated strong co-citation link strength. These works provide the core computational architectures that support a wide range of AI applications in education, such as predictive analytics, performance modeling and intelligent tutoring systems. Collectively, these studies reflect both the interdisciplinary foundation and the practical convergence of AI techniques into educational measurement and innovation. The blend of domain-specific reviews and algorithmic foundations underscores how AI is not only reshaping technical systems but also redefining assessment practices in education.

4.2. Intellectual Structure of the Field (Co-Citation Analysis)

To address the second research objective, this section explores the intellectual structure of AI applications in educational measurement through co-citation analysis. This bibliometric method identifies references that are frequently cited together, revealing conceptual relationships and interdisciplinary linkages. Using VOSviewer, a total of 106 documents that met the threshold of at least five co-citations were analyzed. The resulting co-citation network revealed three major thematic clusters, each representing a distinct strand of research shaping the evolution of AI in educational measurement.

The first cluster emphasizes the technical foundations of AI, highlighting core machine learning algorithms and deep neural network models. These technologies underpin a range of AI-driven systems used in adaptive testing, automated scoring and learning analytics. Influential studies in this cluster include Srivastava et al. (2014), who introduced dropout as a method to prevent overfitting in neural networks and Pedregosa et al. (2011), who developed Scikit-learn, a widely used machine learning library that supports implementation of various algorithms in educational data analysis.

The second cluster reflects application-oriented research drawn from healthcare and medical education. These studies demonstrate how AI techniques are used in cognitive screening, diagnostic modeling and real-time performance feedback. Notably, Nasreddine et al. (2005) introduced the Montreal Cognitive Assessment (MoCA) as a screening tool for early detection of cognitive impairment, which has inspired similar approaches in educational settings where early identification and intervention are critical for student support.

The third cluster addresses socio-technical considerations, focusing on the ethical, pedagogical and policy-related challenges associated with AI integration in education. This includes concerns around algorithmic bias, transparency and responsible AI use in assessment practices. The presence of this cluster highlights the importance of aligning technological advancements with educational values and governance frameworks.

Together, these three clusters illustrate how research in AI-enhanced educational measurement is shaped by contributions from computer science, health informatics and education. The convergence of technical innovations, practical applications and human-centered considerations demonstrates AI's expanding role not only as a computational tool but also as a catalyst for pedagogical transformation. To further illustrate the intellectual foundation of the field, Table 6 presents the five most frequently co-cited documents, ranked by total link strength. These highly influential publications represent foundational methods such as random forests, support vector machines, ensemble learning and diagnostic screening models. For example, Breiman (2001) introduced the random forest algorithm, which has become a standard for predictive modeling in learning analytics, while Chawla et al. (2002) proposed SMOTE, a technique for handling imbalanced data that is widely applied in educational prediction models. These co-cited works collectively reveal the interdisciplinary roots and conceptual coherence that define the current knowledge structure of AI in educational measurement.

Table 6. Most Frequently Co-Cited Documents

No.	Authors	Source	Citations	Link Strength
1	Pedregosa et al., 2011	Journal of Machine Learning Research	36	46
2	Breiman, 2001	Machine Learning	29	43
3	Nasreddine et al., 2005	Journal of the American Geriatrics Society	24	20
4	Srivastava et al., 2014	Journal of Machine Learning Research	21	21
5	Chawla et al., 2002	Journal of Artificial Intelligence Research	16	29

Source: Author interpretation based on VOSviewer analysis.

Note. Link strength reflects the frequency with which each document was co-cited with others and indicates its centrality in the knowledge structure.

4.3. Emerging Research Trends in AI-Driven Educational Measurement (Keyword Co-Occurrence Analysis)

This section addresses the third objective of the study, which is to forecast emerging trends and future research directions in AI applications for educational measurement. To accomplish this, a keyword co-occurrence analysis was performed to identify dominant thematic areas and conceptual linkages within the literature. This technique enables the visualization of frequently co-occurring keywords, offering insight into the conceptual structure and research focus areas of the field.

4.3.1. Dataset and Analytical Parameters

Author-provided keywords were extracted from the 921 articles in the final dataset. A minimum occurrence threshold of five was applied to ensure analytical reliability. Of the 5,276 keywords initially identified, 222 met this criterion and were included in the co-occurrence analysis. The network visualization was generated using VOSviewer, allowing for cluster detection based on keyword proximity and total link strength.

4.3.2. Most Frequently Occurring Keywords

Table 7 presents the top 10 keywords based on frequency and total link strength. The most frequently used terms, including "machine learning," "artificial intelligence" and "diagnosis," reflect the interdisciplinary foundations of the field and emphasize the central role of computational methods in modern educational measurement.

These high-frequency terms indicate that current research is heavily oriented toward diagnostic modelling, performance prediction and classification tasks. Originally prevalent in healthcare domains, these approaches are now increasingly integrated into educational technology. The inclusion of keywords such as "dementia" and "prevalence" further highlights the cross-disciplinary application of AI, particularly in cognitive screening, which parallels the objectives of early detection and intervention within educational contexts.

4.3.3. Keyword Clustering and Thematic Interpretation

The six thematic clusters derived from keyword co-occurrence point to several emerging research trends in AI-driven educational measurement. First, the growth of natural language processing (NLP) reflects increasing interest in automated analysis of student-generated text and dialogue for formative feedback and engagement monitoring. Second, the integration of machine learning for performance prediction signals a shift toward continuous, personalized assessment rather than static, high-stakes testing. Third, the presence of terms related to cognitive screening and diagnostics (e.g., neurocognition, Alzheimer's, radiomics) indicates a growing cross-pollination of AI tools from medical and psychological diagnostics into educational settings. Finally, keywords related to simulation and cheminformati**cs** suggest the expansion of immersive, AI-powered environments for STEM assessment and skill transfer.

Table 7. Top 10 Most Frequently Occurring Keywords

Ranking	Keyword	Occurrences	Total link strength
1	machine learning	177	625
2	artificial intelligence	89	298
3	diagnosis	74	327
4	classification	69	277
5	deep learning	56	169
6	prevalence	50	190
7	prediction	45	213
8	dementia	42	235
9	risk	37	123
10	performance	31	125

These trends signal not only a diversification of AI applications in education but also an increasing convergence between assessment and real-time learning analytics. They highlight a growing demand for multimodal, adaptive, and inclusive assessment approaches aligned with 21st-century learning goals.

Figure 3 illustrates the co-authorship network at the country level, where node size represents publication volume and link thickness indicates the strength of international collaboration. The United States, China, and the United Kingdom form the central hubs, reflecting their dominant roles in AI and educational research. Emerging collaboration clusters involving Malaysia, Pakistan, and Nigeria suggest increasing regional engagement with AI in education, possibly driven by growing investments in digital learning infrastructure and international funding partnerships. This aligns with the third research objective of identifying global collaboration patterns shaping the field.

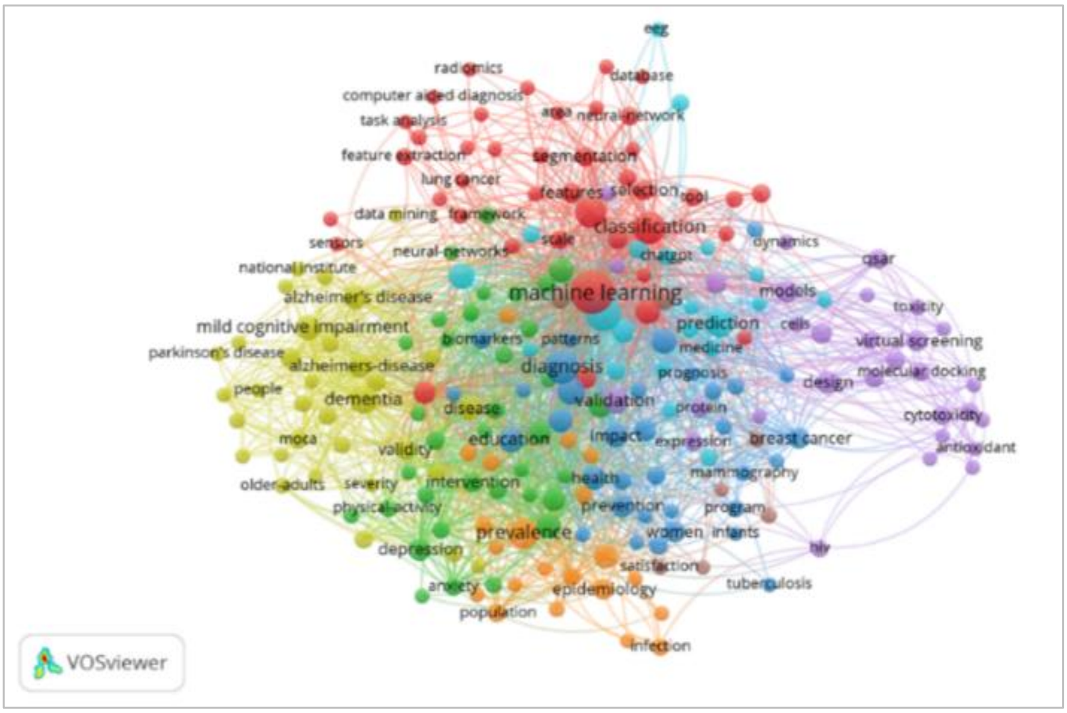

Fig. 3. Keyword Co-Occurrence Network Clusters in AI-Driven Educational Measurement.

Source: Author's visualization using VOSviewer.

5. Discussion and Implications

This section discusses the findings in relation to the study's objectives: identifying influential publications, mapping intellectual structures and understanding global collaboration trends in AI-driven educational measurement.

5.1. Influential Publications and Knowledge Transfer

Citation analysis revealed a marked increase in research output after 2019, reflecting the growing integration of AI methods into educational contexts. Influential studies such as Falk et al. (2019) demonstrate the application of deep learning for image segmentation, originally developed for biomedical imaging. These approaches are now being explored for use in intelligent assessment systems in education.

However, transferring AI techniques from clinical and diagnostic domains into education requires critical examination. The educational context presents unique challenges, such as diverse learning environments, developmental variability, and the need for pedagogical alignment. Tools like U-Net, widely used in radiological image segmentation, require substantial adaptation before being meaningfully applied to student behavior or performance data. Similarly, diagnostic instruments such as the Montreal Cognitive Assessment (MoCA), while effective in health screening, are not designed to support formative, learner-centered educational practices.

In education, the purpose of assessment extends beyond detection or classification. It is essential that assessment supports instructional planning, encourages learner growth, and ensures equitable access to learning opportunities. As such, the adaptation of artificial intelligence systems for educational purposes must be grounded in sound pedagogy, culturally appropriate, and sensitive to developmental stages. Researchers and practitioners should ensure that any model is validated within educational contexts to preserve relevance, fairness, and ethical standards.

5.2. Intellectual Structure and Thematic Patterns

The thematic clusters identified through keyword co-occurrence analysis indicate an increasing focus on multi-dimensional learning analytics that extend beyond traditional cognitive assessment. For example, the application of natural language processing (NLP) in analysing learner dialogue and reflective tasks provides valuable insights into emotional and social learning processes, supporting the development of emotional intelligence (EQ) and spiritual or social intelligence (SQ). Similarly, adaptive learning systems that utilize machine learning facilitate real-time feedback, enabling students to adjust to new learning scenarios and thereby enhancing their adaptability, which reflects the adversity quotient (AQ). Conventional assessments have typically concentrated on measuring intelligence quotient (IQ) through standardized formats; however, AI-powered educational tools allow for more authentic and context-sensitive evaluations of non-cognitive skills. This transformation aligns with global education reform efforts that emphasize the need for comprehensive and inclusive assessments. Within the context of the Fourth Industrial Revolution (4IR), such approaches support the development of learners who are not only intellectually competent but also emotionally resilient, socially aware, and capable of adapting to complex, evolving environments.

5.3. Global Collaboration and Emerging Contributors

Co-authorship analysis highlights that the United States and China are the leading contributors to research in AI-driven educational measurement. Their leadership is supported by large-scale national investments in artificial intelligence and education, along with strategic research partnerships involving global institutions. In Europe, countries such as the United Kingdom, Germany, and France are actively involved through initiatives like Horizon Europe, which encourage responsible and inclusive AI adoption in educational settings.

Within the Asia-Pacific region, Singapore, Japan, and South Korea are recognized as frontrunners in the development of smart education ecosystems. These include innovations in intelligent tutoring systems, adaptive learning platforms, and data-driven instructional design. Singapore's Smart Nation initiative, in particular, exemplifies a strong national commitment to digital transformation in education.

A number of emerging contributors are also gaining visibility. Countries such as Malaysia, Pakistan, Nigeria, and Saudi Arabia have demonstrated increasing research output and involvement in international academic collaborations. This growth is supported by public and private investments in educational technology, the establishment of national AI strategies, and the implementation of digital learning policies. For instance, Malaysia's National AI Roadmap and Pakistan's Digital Learning Initiatives have stimulated local research on AI-powered assessment tools and learner analytics. These efforts indicate a decentralization of research leadership and highlight new opportunities for contextual innovation in the Global South.

Figure 4 presents the international co-authorship network. Each node represents a country, and connecting lines indicate the strength of collaboration. Larger nodes reflect higher levels of publication output, while thicker lines denote more intensive partnerships.

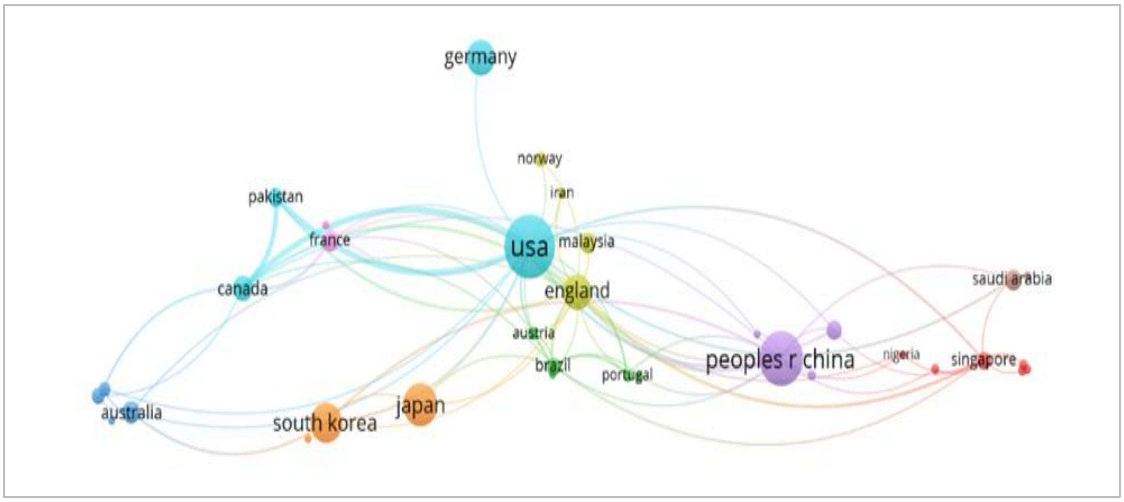

Fig.4. Global Co-authorship Network in AI-Driven Educational Measurement
Source: Author's visualization using VOSviewer

This global landscape reflects a vibrant and increasingly inclusive research community. Cross-border collaboration enhances the diversity of perspectives, ensuring that AI tools in education are not only technologically robust but also culturally and pedagogically relevant. Moving forward, the emphasis should be on scalable, ethical and locally adaptable AI applications that can address global and regional educational challenges.

5.4. Ethical Considerations in AI-Driven Educational Measurement

As AI continues to shape educational measurement, ethical considerations must be addressed to ensure that technological innovation supports rather than compromises equitable learning. One critical issue is algorithmic bias, where AI systems trained on incomplete or unrepresentative datasets may reinforce existing disparities. For example, predictive models based on high-performing learners may misclassify students from underserved backgrounds, resulting in inappropriate placement or missed intervention opportunities (Holmes et al., 2021).

Another concern is data privacy. AI-powered assessment tools often rely on real-time behavioral data, biometric indicators, or learning analytics, which may raise compliance issues under privacy regulations such as GDPR or national student data policies. Without transparent data governance, such systems risk violating learner autonomy and institutional accountability.

Pedagogical misalignment also arises when AI tools prioritize performance prediction at the expense of deeper learning goals. For instance, systems optimized for test score accuracy may neglect developmental aspects such as motivation, creativity, or critical thinking. If educators are not involved in the design and validation of these tools, the resulting assessments may fail to reflect curriculum values or learner diversity.

To address these challenges, future research must incorporate fairness audits, ethical-by-design frameworks, and teacher involvement throughout the AI implementation process. Responsible AI in education must be inclusive, interpretable, and aligned with educational values beyond efficiency and accuracy.

6. Conclusion

This bibliometric review synthesized 921 peer-reviewed articles on the role of artificial intelligence (AI) in educational measurement, published between 1997 and 2024. Guided by three core objectives, the study identified the most influential publications, mapped the intellectual structure of the field, and uncovered key research trends and global collaboration patterns. The findings reveal a marked increase in research activity after 2019, driven by advancements in machine learning, big data analytics, and diagnostic modeling. Widely used methods such as decision trees, support vector machines, and deep learning have been applied in areas including adaptive assessment, cognitive screening, and personalized learning pathways. Thematic analysis shows that AI in education intersects with multiple disciplines, including neurocognitive diagnostics, public health, and natural language processing, highlighting the field's interdisciplinary nature.

International collaboration continues to expand, with significant contributions from the United States, China, and several European countries. New research hubs are also emerging in Malaysia, Nigeria, and Saudi Arabia, reflecting a growing commitment to inclusive and regionally relevant innovation. These global networks suggest a collective interest in leveraging AI to promote equity, flexibility, and innovation in education. However, this study is limited to data extracted from the Web of Science database, which may not fully represent the broader research landscape. Furthermore, bibliometric analysis cannot capture the practical complexities of implementing AI in educational settings. To address these limitations, future research should integrate bibliometric insights with qualitative case studies, cross-database comparisons, and classroom-based evaluations to better understand how AI tools function in diverse educational environments.

Several directions for future inquiry have emerged. First, researchers should explore how natural language processing can be used to assess not only content knowledge but also learner sentiment, collaboration, and engagement. Second, there is a need to validate predictive models across various educational contexts, especially among underrepresented populations and in special education. Third, interdisciplinary collaboration will be essential to ethically adapt AI tools from domains such as health diagnostics to learner-centered educational settings. Lastly, future research should prioritize the development of AI-supported assessments that are interpretable, pedagogically aligned, and capable of nurturing key competencies associated with the Fourth Industrial Revolution (4IR), such as emotional intelligence, adaptability, and social reasoning. This review highlights the importance of developing AI applications for educational assessment that are equitable, ethically grounded, and accessible to diverse learners. These steps will help ensure that AI enhances, rather than displaces, human-centered educational practices.

Acknowledgements: Insert acknowledgements here (if any) We are really grateful for the assistance that the Federal Training Prize (HLP) scholarship from the Ministry of Education (MOE) has given us. We acknowledge the supervisor's and Kumpulan Penyelidikan Universiti (KPU), UKM for the knowledge and insights, which greatly aided this study. We also like to express our gratitude to our colleagues in the Measurement and Evaluation course for their support and helpful criticism, which has helped to make this manuscript better.

Funding Statement: Insert funding information here (if any) This research was funded by the Faculty of Education, Universiti Kebangsaan Malaysia (UKM) with grant number GG-2024-024 (Program STEM dan Minda).

References

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C. M., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J. M., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilova, P. (2024). *The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.18900

Chawla, N. V., Bowyer, K. W., Hall, L., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16, 321*. https://doi.org/10.1613/jair.953

Husaeni, D. F. A., & Nandiyanto, A. B. D. (2021). Bibliometric Using Vosviewer with Publish or Perish (using Google Scholar data): From Step-by-step Processing for Users to the Practical Examples in the Analysis of Digital Learning Articles in Pre and Post Covid-19 Pandemic. *ASEAN Journal of Science and Engineering, 2(1), 19.* https://doi.org/10.17509/ajse.v2i1.37368

Ishak, H., Matore, M. E. E. M., & Nor, H. A. M. (2025). Adaptive behavior assessment of individuals with intellectual disability: A comprehensive review. *Multidisciplinary Reviews*, *9*(1), 2026034. https://doi.org/10.31893/multirev.2026034

Ismail, S. A. S., Maat, S. M., & Khalid, F. (2024). 35 years of fraction learning: Integrating systematic review and bibliometric analysis on a global scale. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(12), em2376. https://doi.org/10.29333/ejmste/15657

Jia, K., Wang, P., Li, Y., Chen, Z., Jiang, X., Lin, C., & Chin, T. (2022). Research landscape of artificial intelligence and e-learning: A bibliometric research. *Frontiers in Psychology, 13*, **795039**. https://doi.org/10.3389/fpsyg.2022.795039

Kamaruddin, M., & Mohd Matore, M. E. E. (2022). Development and validation of psychometric properties of the 10 IB learner profile instrument: A combination of the Rasch and classical measurement model. *International Journal of Environmental Research and Public Health, 19*(10), 6123. https://doi.org/10.3390/ijerph19106123

Klarin, A. (2024). How to conduct a bibliometric content analysis: Guidelines and contributions of content co-occurrence or co-word literature reviews. *International Journal of Consumer Studies, 48*(2), 185–201. https://doi.org/10.1111/ijcs.13031

Kumar, M., Kumari, S., Guntipally, S. S., Chhabra, A., Kumar, A., & Kumar, P. (2024). A step-by-step guide of bibliometric study for healthcare and allied research. *Journal of Pharmacy and Bioallied Sciences, 16*(2), 145–151. https://doi.org/10.4103/jpbs.jpbs_1347_24

Ma, H., & Ismail, L. (2025). Bibliometric analysis and systematic review of digital competence in education. *Humanities and Social Sciences Communications, 12*(1), Article 37. https://doi.org/10.1057/s41599-025-04401-1

Masdoki, M., Din, R., & Mohd Matore, M. E. E. (2022). Teaching 4.0 competency in higher learning institutions: A systematic mapping review. *International Journal of Learning, Teaching and Educational Research, 20*(7), 217–231. https://doi.org/10.26803/ijlter.20.7.12

Mohd Noh, M. F., & Mohd Matore, M. E. E. (2022). Rater severity differences in English language as a second language speaking assessment based on rating experience, training experience, and teaching experience through many-faceted Rasch measurement analysis. *Frontiers in Psychology, 13*, 941084. https://doi.org/10.3389/fpsyg.2022.941084

Nasreddine, Z., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Passas, I. (2024). Bibliometric analysis: The main steps. *Encyclopedia, 4*(2), 1014–1024. https://doi.org/10.3390/encyclopedia4020065

Sovey, S., Osman, K., & Mohd Matore, M. E. E. (2022). Rasch analysis for disposition levels of computational thinking instrument among secondary school students. *Eurasia Journal of Mathematics, Science and Technology Education, 18*(3), em2088. https://doi.org/10.29333/ejmste/11794

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(1), 1929–1958. https://jmlr.csail.mit.edu/papers/volume15/srivastava14a/srivastava14a.pdf

Swargiary, K. (2024). *The impact of AI-driven personalized learning and intelligent tutoring systems on student engagement and academic achievement: Ethical implications and the digital divide* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4897241

Taherdoost, H., Le, T., & Slimani, K. (2025). Cryptographic techniques in artificial intelligence security: A bibliometric review. *Cryptography, 9*(1), 17. https://doi.org/10.3390/cryptography9010017

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications, 252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Yang, Y., Zhang, Y., Sun, D., He, W., & Wei, Y. (2025). Navigating the landscape of AI literacy education: Insights from a decade of research (2014–2024). *Humanities and Social Sciences Communications, 12*(1), Article 58. https://doi.org/10.1057/s41599-025-04583-8

Zhao, X., Wider, W., Jiang, L., Fauzi, M. A., Tanucan, J. C. M., Lin, J., & Udang, L. N. (2024). Transforming higher education institutions through EDI leadership: A bibliometric exploration. *Heliyon, 10*(4), e26241. https://doi.org/10.1016/j.heliyon.2024.e26241