International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 5, pp. 283-303, May 2025
https://doi.org/10.26803/ijlter.24.5.15
Received Mar 25, 2025; Revised May 15, 2025; Accepted May 20, 2025

Performance Evaluation for Educators in Higher Education from Bibliometric Analysis Views

Nor Azarif Sofia Hussain
Faculty of Education, The National University of Malaysia (UKM),
Bangi, Selangor, Malaysia

Mohd Effendi@Ewan Mohd Matore* **and Mohd Izham Mohd Hamzah**
Research Centre of Education Leadership and Policy, Faculty of Education,
The National University of Malaysia (UKM), Bangi, Selangor, Malaysia.

Abstract. Educator performance evaluation in higher education is an evolving process that is shaped by pedagogical advances and institutional needs. This study uses bibliometric analysis to explore research trends and emerging themes in educator performance evaluation. Using Scopus data, a comprehensive analysis of 1,173 publications from 2000 to 2024 was conducted, examining the evolution from traditional student evaluation to competency-based frameworks that integrate artificial intelligence and data-driven methodologies. Key findings indicate a shift toward technology-enhanced evaluation, integrating AI-driven analytics and real-time feedback. The study highlights key contributors, including leading countries, institutions and authors, providing insights into global research collaborations. The analysis identified five dominant research clusters: technology-driven evaluation systems, competency-based faculty evaluation, policy frameworks, faculty evaluation in medical education, and post-pandemic teaching evaluation. The increasing use of machine learning, data mining and performance indicators reflects a shift toward structured and objective evaluation methods. However, challenges such as implicit bias, inconsistent rubrics and an overreliance on quantitative metrics persist. Findings from this study offer valuable insights for policymakers, administrators, and educators seeking to refine evaluation frameworks in alignment with emerging trends. The research underlines the necessity for a hybrid evaluation method that balances quantitative performance measurements with qualitative evaluation strategies, ensuring a thorough understanding of teaching efficacy. For higher education evaluation systems to be more equitable, dependable, and inclusive, future studies should investigate different approaches, interdisciplinary evaluation frameworks, and the moral ramifications of AI-driven performance evaluations.

*Corresponding author: Mohd Effendi @ Ewan Mohd Matore; *effendi@ukm.edu.my*

Keywords: performance evaluation; educator; bibliometric analysis; Higher Education Assessment; Sustainable Development Goals (SDGs)

1. Introduction

The transformation towards the implementation of appraisal systems for higher education educators involves changes in the field pedagogy and institutional restructuring, as well as a universalistic concern for sustainable and just educational development. Aligned with Sustainable Development Goal 4 (Quality Education), continual evaluation of educators' competencies enhances the quality of teaching and learning outcomes (Pifer et al., 2024; Hao, 2024). This aligns with the fact that educator performance is often positively impacted by emotional and professional support systems. At the same time, this standard also in line with the needs of new competencies of 4.0 era (Masdoki et al., 2021). Educator performance is often positively impacted by emotional and professional support systems, with factors such as work happiness mediating the effects of occupational stress on teaching outcomes (Zhu & Chang, 2025).

At the same time, SDG 10 (Reduce Inequality) specifies that there is a need for unbiased and holistic evaluation systems. Newer models like 360-degree appraisal (Ugwoke et al., 2023; Rahimi et al., 2024) and multi-criteria decision models (Mamatha et al., 2024) appear to be more popular because of their multicentral approach to fairness across a variety of institutional settings and educator types. These methods counteract the effects of institutional hierarchy, resource inequality, or other subjective biases common in appraisals. Concerning SDG 16 (Peace and Justice and Strong Institutions), work geared towards the development of non-subjective, automated, and integrity-based evaluation systems has increased. Formatting-based models (Hou, 2024), fuzzy analysis approaches (Geng et al., 2024), integrated performance dashboards (Karimi et al., 2024) are necessary for improving institutional and governance credibility.

As pedagogies and institutional needs evolve, educator evaluation in higher education progresses alongside it. This change from subjective evaluations to structured, data-driven approaches accompanies the growing international concern for accountability and transparency (Hou, 2024; Karimi et al., 2024). Site-based assessment strategies combine qualitative and quantitative evaluation so that holistic college and faculty performance measures are achieved to support institutional accreditation and teaching excellence (Pifer et al., 2024; Ugwoke et al., 2023). Furthermore, systematic evaluation of performance achievements helps develop and enhance educator's professional competencies, thus nurturing reflective practice and academic honesty (Hao, 2024; Prieto et al., 2023).

Because of the changing paradigms in teaching and learning, institutions of higher learning globally have instituted comprehensive evaluation systems to assess educators appropriately. Professional development programs that allow educators to adopt digital-centric teaching approaches are one of the components of these frameworks (Musid, Matore, & Hamid, 2024). These include performance appraisal tools such as student evaluations, peer reviews, and self-assessments through which a educator's effectiveness can be assessed accurately. As Altassan

(2022) highlights, transparency and sustainability in appraisal systems are crucial for fostering academic excellence and institutional credibility.

The higher education performance evaluation for educators has gone through three predominant phases: traditional, developmental, and competency-oriented. The traditional one based on student evaluation was problematic because of its dependability on satisfaction surveys which was highly subjective and biased (Prieto et al., 2023). Later developmental models added mentorship along with formative evaluation for sustaining growth (Pifer et al., 2024). The emergence of competency-based models aligned educator performance with institutional goals and industry standards (Fan et al., 2022). Technological advancements, including AI and data analytics, now enhance objectivity in evaluations (Zheng & Sun, 2022; Liu, 2024). Bibliometric analysis serves as a systematic method to trace this scholarly progression (Janavi & Abdi, 2024).

It is possible to gain a comprehensive view of the scholarly environment by looking at the configuration of citation clusters and collaboration of authors thanks to bibliometric analysis. It reveals key works, spots inadequate areas, and shows innovation in methods of educator performance evaluation (Janavi & Abdi, 2024). This research combines bibliometric tools with the evolution of performance appraisal in higher education to track its history in alignment with the new requests for analyzing evidence (Mamatha et al., 2024; Geng et al., 2024). Its goal is to understand the prominent participants, principal sources of publications as well as the dominant institutions' discourses while observing changes in keywords and co-author patterns (Hou, 2024; Karimi et al., 2024).

The evolution of performance evaluation for educators in higher education reflects a dynamic interplay between traditional evaluation methods, technological advancements, and shifting academic priorities. By leveraging bibliometric analysis, this study seeks to contribute to the ongoing discourse on refining evaluation frameworks to enhance educational quality and institutional accountability. The findings will offer valuable insights for policymakers, administrators, and educators striving to optimize evaluation mechanisms in alignment with best practices and emerging trends.

Thus, this article will analysis and explore the evolution of performance evaluation for educators in higher education, identifying key developments and emerging trends relevant to our research objectives. This study addresses the following main research questions:

RQ1. How has the evaluation of educator performance in higher education changed over time based on the year of publication?
RQ2: Which countries, institutions and authors have made contributions to research in this field?
RQ3: What are the prevailing themes and most frequently used keywords in performance evaluation research?

2. Literature Review

Historically, the evaluation of educators in higher education was premised on standardized test scores and grades, with an emphasis on student evaluation and feedback (Zaabalawi & Zaabalawi, 2024). They gave little consideration to qualitative aspects of teaching. Initially, evaluations consisted largely of final semester evaluations, without much attention to pedagogical enhancement.

By the early 21st century, institutions started employing a wider range of evaluative instruments that included evaluation of teaching, curriculum development, and the level of participation of the educator (Do et al., 2024). Self and peer assessments gained prominence, focusing on reflective elements and professional growth. Technological evaluation, like the Learning Management System (LMS) enabled the capture of interaction and engagement data by students within the course, thus creating the possibility of determining the effectiveness of teaching in a more comprehensive manner (Liu, 2024). The implementation of formative evaluations made it possible to change many of the teaching methods that were taught, resulting in a more flexible educational system (Zhu et al., 2024). In response to the need for more integrated evaluation systems that focus on competencies, new models of evaluation, like project and rubric type evaluation emerged (Zaabalawi & Zaabalawi, 2024). These shifts highlight the need for valid, reliable, and fair evaluation methodologies (Do et al., 2024).

Schools offer both qualitative and quantitative approaches to evaluation. Open-ended student comments and narrative evaluations, for example, are qualitative methods which capture deeper details about teaching effectiveness (Yılmaz, 2024). Cross peer reviews enable professional collaboration while offering diverse opinions about the educator's performance (Pifer et al, 2024). Self-reviews allow faculty members to think about the strengths and weaknesses they possess in relationship to the institution's standards. Students reviews of educators still stand as a fundamental concern, but controversies on whether it is done in an objective manner still exists (Chen & Liu, 2023).

This goes with the comprehensive change in the evaluation of performance where educators influence the level of resilience and adaptability of the students. Therefore, linking the adversity quotient (AQ) with students' academic performance is relevant and challenges the National Philosophy of Education (FPK) in its efforts to enhance value addition for graduates (Matore et al, 2017). Data analytics enabled by technology have now assumed a fundamental position in the processing of feedback. Still, there should be a balance between numbers and descriptive comments so that evaluation is fair (Khan & Badulescu, 2025).

AI has redefined educator evaluations from a qualitative approach to a more objective and systematic one. E-learning platforms and AI-enabled services track and evaluate student participation and learning outcomes (Rogers et al., 2025). AI assists in providing precise guidance which helps in improving teaching methods (Chen et al., 2025). Transparent and efficient criteria for evaluation can be achieved through real-time data analytics and automated rubrics (Maturana et al.,

2025). Machine learning models further improve evaluations by anticipating performance trends and advising on appropriate professional development (Ramasamy & Cheng, 2024). AI-based evaluations risk repeating errors of existing systems by focusing on quantifiable metrics while teaching qualitative aspects are ignored (Bennett, 2025). Institutions have to connect AI based evaluations to the various demands of education to ensure human control concerning the accuracy and trustworthiness of such evaluations. Besides, they can connect with out of the box approach in education setting such as six sigma. This will give more colors in educators transformation quality (Sabtu & Mohd Matore, 2024).

All issues related to prejudice in evaluations, unclear criteria for evaluation, and insufficient evaluation training profoundly complicate the evaluation of educators (Darling-Hammond, 2012). Gender, ethnicity, or institutional bias can blind and lead to unfair performance appraisal results. Therefore, failing to use set rubrics leads to arbitrarily different evaluations of the same educators. Many institutions do not offer sufficient training to evaluators, which leads to poor or no feedback (Mahmud et al., 2024). Mentorship and engagement by students are often lost due to an overemphasis on quantitative metrics (Chen & Chen, 2024). Due to misplaced institutional culture, research output is often prioritized over teaching effectiveness. Addressing this problem by implementing a mechanistic, transparent process with qualitative and quantitative measures restores fairness and further professional development (Velázquez et al., 2024). The use of digital tools in evaluation processes can enhance fairness in practical evaluations, offering a promising approach to reduce bias based on gender, ethnicity, or institutional structures (Firmansyah et al., 2022).

Different regions have their own evaluation practices. In the United States, there is extensive use of multi-source feedback systems comprising of student feedback, peer reviews, and self-assessment (Sabharwal & Miah, 2024). On the other hand, Southeast Asian countries place a disproportionate emphasis on student evaluations, which can be problematic because it can introduce bias and grade inflation (Liu & Chen, 2024). Finnish and Dutch models are more focused on competency evaluation; for example, they assess collaborative teaching, research participation, and other aspects of professional development (Regmi et al., 2024). However, poorer countries have less resources and therefore have to depend on a narrower range of evaluations (Alshaikhi & Khasawneh, 2025). Regardless of these regional differences, evaluation based on technology is being adopted in most regions around the world. Almost half of the developed countries have provided online evaluation systems for their institutions, thus ensuring the efficiency and accuracy of data (Zhu, 2024). A common trend that all these countries share is the shift towards more systematic and transparent methods for gauging performance to ensure accountability and equity (Alshaikhi & Khasawneh, 2025).

3. Methodology

This study examines the evolution of performance evaluation for educators in higher education through a bibliometric analysis framework. Building on the insights of (Aria & Cuccurullo, 2017), bibliometric methods are pivotal in

addressing three fundamental inquiries: first, they delineate the core knowledge and intellectual structure of a research domain; second, they explore the cutting edge or conceptual framework of the field; and third, they map the social network of the scholarly community. As depicted in, this methodology has enabled the extraction of essential elements such as the most cited documents, the leading authors, emergent trend keywords, and the driving forces as well as the dimensions of artificial intelligence applications in human resource management. The process is organized into five distinct stages study design, data collection, data analysis, data visualization, and interpretation in accordance with the bibliometric mapping standards set forth (Zupic & Čater, 2015).

3.1 Study Design

The study design phase lays the foundation for the research by defining the guiding questions and determining the methodological approach necessary to address them (Aria & Cuccurullo, 2017). During this phase, strategic decisions are made regarding several critical aspects: establishing the temporal boundaries of the study, selecting the keywords that will direct the literature review, choosing the appropriate databases for sourcing relevant literature, and identifying the software tools required for data analysis. Each decision is carefully calibrated to ensure alignment with the overall objectives of the study. Figure 1 shows the study design of this study.

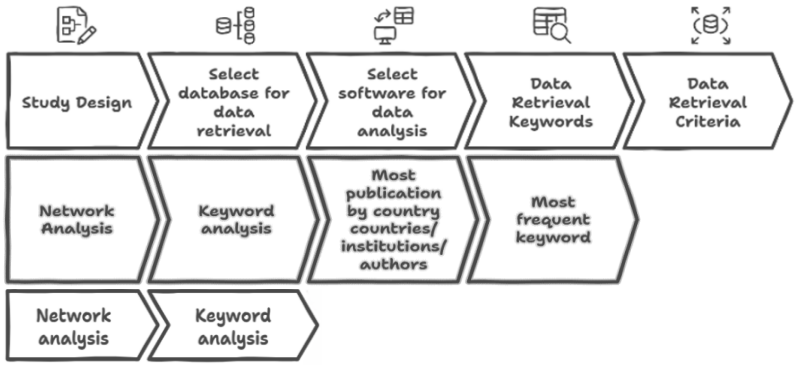

Figure 1: Study design

3.2 Data Collection

Data for the analysis were sourced from the Scopus database, widely recognized for its extensive coverage and frequent application in bibliometric research (Mongeon & Paul-Hus, 2016). Despite the acknowledged advantages of other databases such as PubMed, Web of Science, and Google Scholar (Harzing & Alakangas, 2016), Scopus was preferred due to its comprehensive repository of peer-reviewed publications. The literature search was executed using targeted keywords ("performance evaluation*" OR "performance assessment*" OR "performance appraisal*" OR "teach* evaluation*" OR "teach* assessment*") AND ("educator*" OR "lecturer*" OR "teacher*" OR "scholar*" OR "tutor*" OR "supervisor*" OR "mentor*") AND ("higher education" OR "universit* student*" OR "tertiary education" OR "college*" OR "graduate school*" OR "institute*") applied to titles, abstracts, and keyword sections, with the timeframe limited to

2015–2025 to capture the most recent developments as showed in Table 1 and Table 2.

In recognition of the fact that journal articles typically offer the most current and detailed research findings especially in contrast to conference proceedings, which have been more severely impacted by disruptions such as those experienced during the COVID-19 pandemic (Jin et al., 2018) the study exclusively considered journal articles. An initial pool of 1356 documents were refined by filtering according to publication year, document type, and language, resulting in a final selection of 1173 relevant articles.

Table 1: The search string

Database	Keyword Used
Scopus	TITLE-ABS-KEY ((("performance evaluation" OR "performance as sessment" OR "performance appraisal" OR "performance review" O R "performance feedback") AND ("educator*" OR "instructor*" OR "faculty" OR "teacher*") AND ("higher education" OR "tertiary edu cation" OR "postsecondar*" OR "universit*")))

Table 2: The exclusion and inclusion criteria

Criterion	Inclusion	Exclusion
Year of publication	2000 − 2024	< 2000
Language	English	Non-English
Publication Stage	Final	In Press

Overall, this bibliometric approach integrates quantitative techniques to systematically navigate scientific databases via Scopus, utilizing analytical tools such as VOSviewer. Although the Web of Science is also a major bibliometric resource supporting diverse analytical outputs in conjunction with VOSviewer version 1.6.19, the adopted method proved effective for identifying evolving research trends in performance evaluation. The entire process was conducted in five sequential steps as outlined by Donthu et al. (2021): (1) defining the study's aim and scope, (2) selecting appropriate bibliometric techniques, (3) collecting the bibliometric data, (4) executing the bibliometric analysis which includes both performance analysis and science mapping, and (5) presenting the findings along with a discussion of their implications for future research.

3.3 Data Analysis

VOSviewer is employed to visualize bibliometric networks using data imported from the Scopus database. Figure 2 illustrates the bibliometric analysis process applied to the Scopus-derived data through VOSviewer (Eun & Sanghee, 2022).

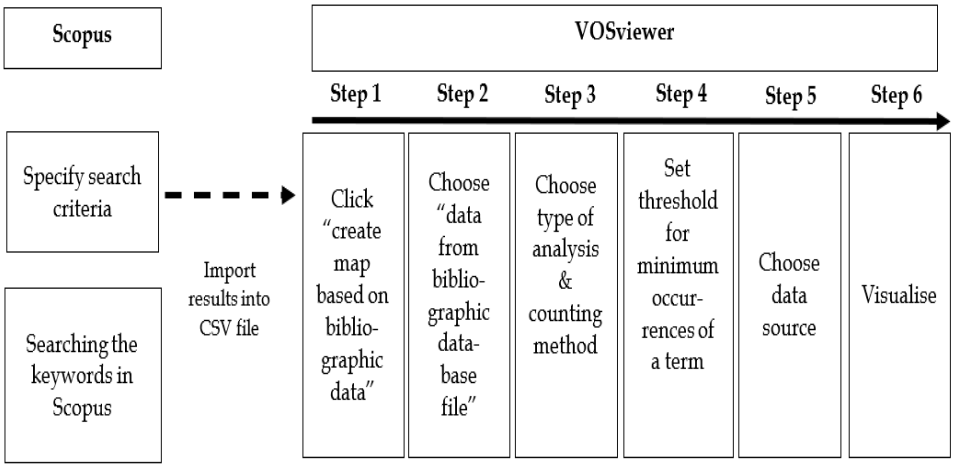

Figure 2: The process of the bibliometric analysis using the VOSviewer

The dataset, comprising publication year, title, author names, journal, citations, and keywords in CSV format, was extracted from the Scopus database, covering the period 2000 to 2024. The collected data was then analyzed using VOSviewer software (version 1.6.15) to visualize research trends and collaborations in the field of performance evaluation for educators in higher education. This software employs VOS clustering and mapping techniques, enabling the creation of bibliometric networks that illustrate the relationships between research topics, authors, and institutions (Van Eck & Waltman, 2010).
Data analysis in bibliometric research entails systematic examination and interpretation of collected information to discern patterns and thematic structures in performance evaluation research. Citation analysis identifies influential works by tracking citation counts, while co-citation mapping examines relationships among referenced studies (Epstein, 2018; Székely & Knirsch, 2005). Keyword co-occurrence analysis reveals dominant research themes and their interconnections, highlighting trends in evaluation methodologies and faculty evaluation strategies (Aguinis, 2019).

Using visualization techniques, essential bibliometric analyses were conducted, including keyword co-occurrence analysis, citation analysis, and co-citation analysis. Keyword co-occurrence analysis was employed to recognize dominant research themes and monitor the development of performance evaluation research in higher education. Keyword co-occurrence analysis has been instrumental in recognizing key topics and trends in faculty performance evaluation (Janavi & Abdi, 2024).

Similarly, citation analysis has played a crucial role in identifying key publications and influential authors, providing insights into historical and emerging research trajectories (Hou, 2024). Moreover, co-citation analysis, a widely used bibliometric method, mapped scholarly relationships to highlight interdisciplinary research connections and the diffusion of knowledge in the field of educator performance evaluation (Mamatha et al., 2024). These bibliometric techniques collectively provide a comprehensive understanding of the evolving landscape of performance evaluation research.

4. Result and Discussion

4.1 How has the evaluation of educator performance in higher education changed over time based on the year of publication?

The evolution of educator performance evaluation in higher education has undergone significant transformation over the past two decades, as reflected in the increasing volume of research publications. Figure 3 shows the number of publications per year from 2000 to 2024. In the early years (2000–2004), research on performance evaluation was relatively limited, with publications ranging between 9 and 16 per year. During this period, traditional evaluation methods such as peer assessments and student feedback were widely used, with fewer empirical studies exploring innovative evaluation techniques. However, as higher education institutions placed greater emphasis on institutional development and accountability, research interest began to grow. From 2005 to 2012, the number of publications steadily increased, peaking at 49 in 2012. This period saw the rise of student-centered learning approaches and institutional policies aimed at enhancing teaching effectiveness, which contributed to the growing focus on educator performance evaluation.

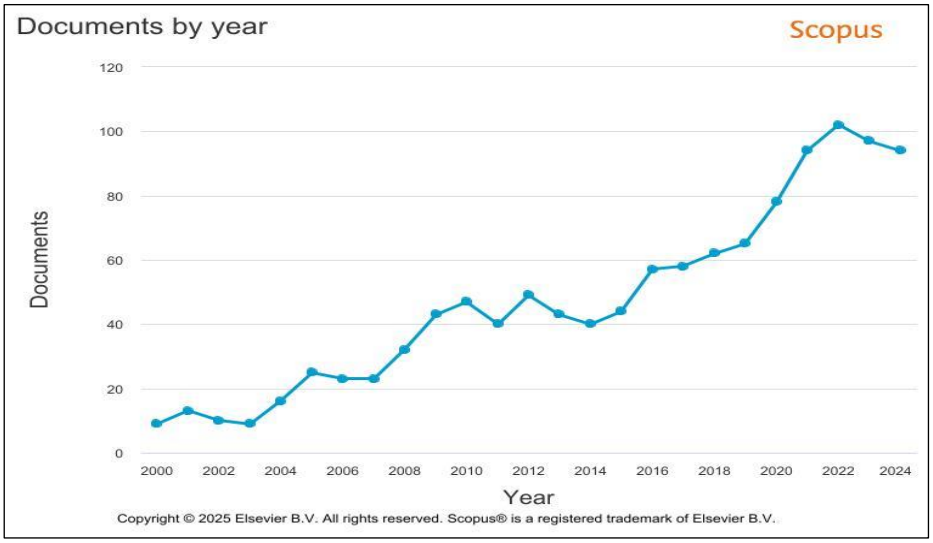

Figure 3: Trend of performance evaluation research publications by years.

Research activity stayed consistent from 2013 to 2018, with minor variations but a yearly publication count of over 40. Learning analytics, structured feedback systems, and institutional performance reviews are just a few examples of the data-driven methodologies that were integrated into institution evaluation during this phase. From 2019 onward, there was a notable change, with a strong rise in publications that peaked in 2022 at 102, the greatest number in this dataset. The growing use of technology-enhanced evaluation systems and the COVID-19 pandemic, which called for different evaluation frameworks for online and hybrid learning contexts, are both responsible for this spike in research effort. A reevaluation of educator performance was prompted by the pandemic-driven shift to virtual learning, with a focus on engagement metrics, the efficacy of digital teaching, and online student interaction.

Research interest in educator performance evaluation is still high, as seen by the 97 publications in 2023 and the 94 in 2024, despite a minor decline in publications after 2022. This ongoing interest points to the need for more research into cutting-edge evaluation techniques like competency-based institutional evaluation models, real-time performance tracking, and evaluations powered by artificial intelligence (AI). Future research is anticipated to concentrate on automated evaluation tools, adaptive learning analytics, and real-time feedback mechanisms that can offer a more thorough and dynamic evaluation of educator performance as higher education institutions adjust to changing teaching modalities and technological advancements. The increasing significance of creating strong, technologically advanced frameworks to improve teaching quality and institutional efficacy in higher education is highlighted by the consistent rise in scholarly contributions throughout time.

4.2 Which countries, institutions and authors have made contributions to research in this field?

The analysis of country-wise contributions to research on educator performance evaluation in higher education indicates that certain nations have played a more dominant role in advancing this field. The United States, China, Australia, India, the United Kingdom, Canada, Turkey, Indonesia, Iran, and Saudi Arabia are the top ten contributors to research in this area. The data can be seen in Figure 4. These countries have demonstrated a significant scholarly impact through a substantial number of research publications.

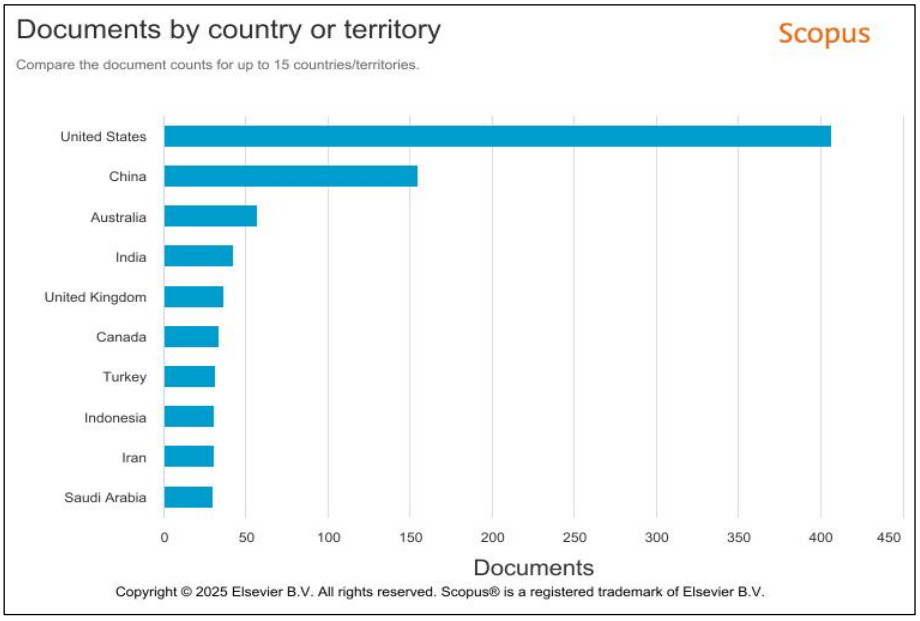

Figure 4: Trend of performance evaluation research publications by country (Source: Adapted from Scopus analysis)

Among these nations, the United States leads in publication output, highlighting its well-established higher education system, strong research funding, and emphasis on institutional performance evaluation. Many American universities implement rigorous educator performance evaluation models, including student feedback, peer review, and learning analytics-driven evaluations, which

contribute to the extensive research in this field. China follows as the second-largest contributor, reflecting its increasing focus on education quality and academic performance. China's higher education reforms, including educator evaluation systems, pedagogical innovations, and competency-based evaluations, have driven substantial research interest in educator performance evaluation. The United Kingdom, Canada, and Australia also rank high in research contributions, largely due to their structured quality assurance frameworks and accreditation policies that mandate continuous institution performance reviews.

Emerging economies such as India, Turkey, Indonesia, Iran, and Saudi Arabia have also shown significant contributions to the field. India's growing interest in faculty evaluation aligns with its push for higher education reforms, accreditation processes, and the adoption of digital tools for educator evaluation. Similarly, Turkey and Indonesia have seen increasing research outputs, likely due to national policies emphasizing educator training, teaching effectiveness, and institutional accountability. Iran and Saudi Arabia, though relatively newer contributors, are witnessing a rise in research on faculty evaluation, possibly due to national efforts to improve university rankings and academic performance.

Figure 5 shows the volume of publications by institutions. The top institutions contributing significantly to research on educator performance evaluation in higher education include leading universities and medical schools primarily based in North America. The University of Michigan, Ann Arbor, University of Toronto, and Stanford University are among the most influential research centers in this domain, alongside prestigious medical schools such as the University of Michigan Medical School and Northwestern University Feinberg School of Medicine.

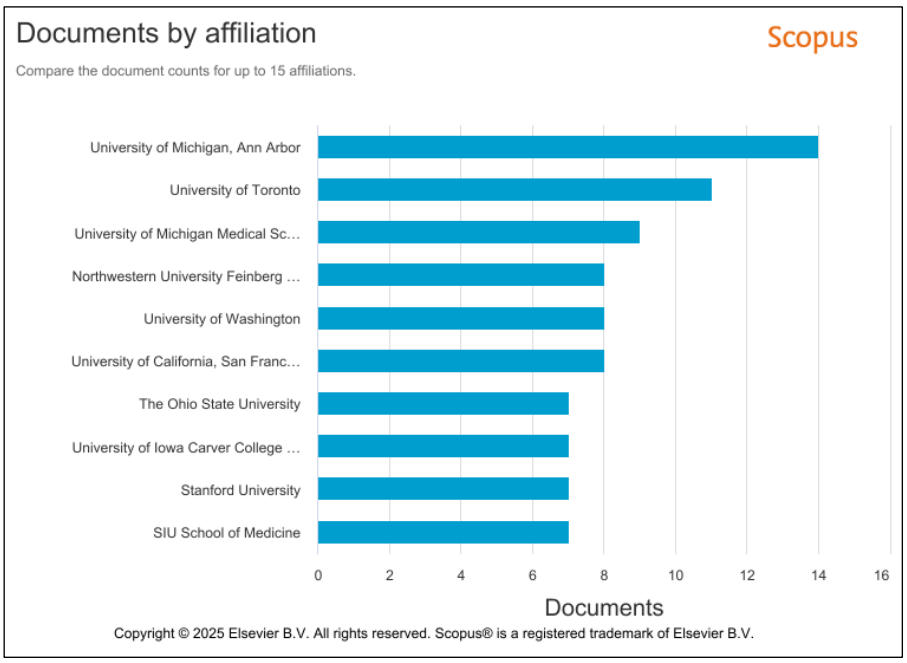

Figure 5: Trend of performance evaluation research publications by institution (Source: Adapted from Scopus analysis)

Among these institutions, the University of Michigan, Ann Arbor, appears as a dominant research hub, consistently producing high-impact studies on faculty evaluation, teaching effectiveness, and performance appraisal frameworks. The University of Toronto is also a leading institution, with a strong emphasis on faculty development, innovative evaluation methodologies, and higher education policy research. These universities, along with the University of Washington and the University of California, San Francisco, play a crucial role in advancing knowledge on data-driven faculty evaluation models, learning analytics, and competency-based evaluation frameworks.

A noteworthy trend is the presence of medical schools in this list, including the University of Michigan Medical School, Northwestern University Feinberg School of Medicine, University of Iowa Carver College of Medicine, and SIU School of Medicine. This suggests that educator performance evaluation research is not limited to general higher education but is also a crucial focus within medical education. Given the highly structured nature of medical training, institutional evaluation in medical schools often incorporates clinical teaching effectiveness, student competency evaluation, and evidence-based instructional strategies.

Figure 6 shows the number of publications by authors. The analysis of top contributing authors in educator performance evaluation research reveals that several scholars have made significant contributions to the field. The top 10 most influential authors include Thuc Doan Do, Pamela Morgan, Doreen Cleave-Hogg, Brian George, Haydu, Verônica Bender Haydu, Monalisa Muniz, Sally Santen, Neal Seymour, Kulno Türk, João dos Santos Carmo. These researchers have been instrumental in advancing knowledge on faculty evaluation, performance appraisal models, and educator feedback mechanisms in higher education.

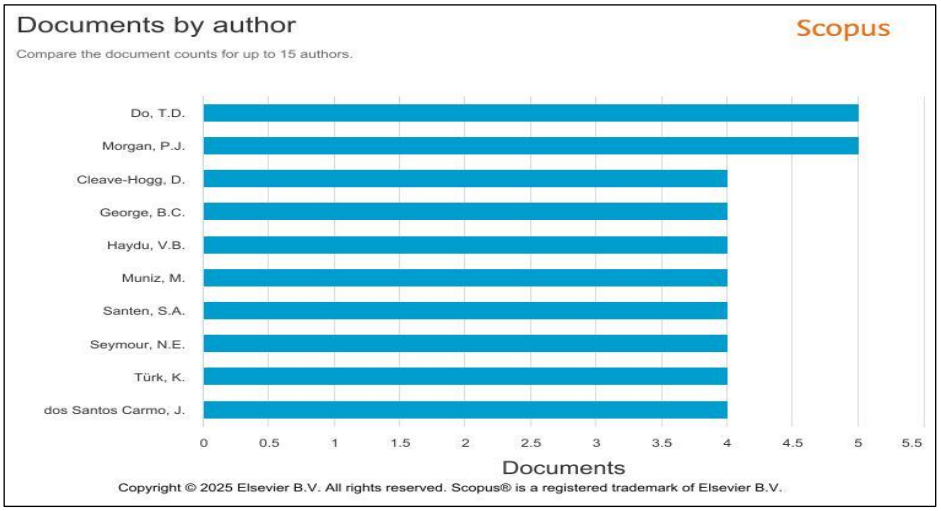

Figure 6. Trend of performance evaluation research publications by authors (Source: Adapted from Scopus analysis)

Among them, Thuc Doan Do and Pamela Morgan are leading contributors, with multiple publications focusing on faculty evaluation methodologies, instructional feedback, and teaching effectiveness metrics. Doreen Cleave-Hogg and Brian George have also played crucial roles, particularly in research related to peer

assessments, competency-based evaluations, and structured faculty development programs. The presence of Verônica Bender Haydu and Monalisa Muniz suggests a strong focus on quantitative faculty evaluation techniques and the integration of data analytics in educator performance evaluation. Additionally, Sally Santen, Neal Seymour and Kulno Türk have contributed significantly to the development of faculty evaluation frameworks, medical education evaluation models, and evidence-based teaching performance indicators. The inclusion of João dos Santos Carmo highlights the role of research in institutional evaluation strategies, student-centered evaluation techniques, and continuous professional development for educators.

4.3 What are the prevailing themes and most frequently used keywords in performance evaluation research?

The Figure 7 below illustrates the network visualization of co-occurrence in performance evaluation research in higher education. The relationships between keywords are represented by clusters of nodes (colors), with proximity indicating their level of connection within the research field. The visualization identifies five primary clusters, each reflecting distinct thematic areas in institutional performance evaluation.

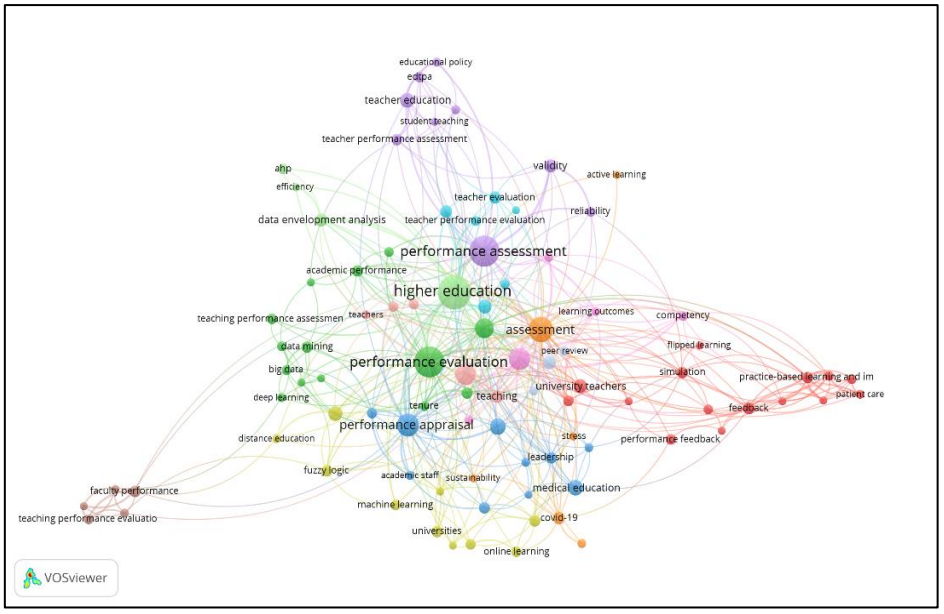

Figure 7. Network Visualization of co-occurrence of keywords used by authors

The green cluster includes 18 keywords associated with technology-driven performance evaluation, comprising keywords such as "performance evaluation", "academic performance", "teaching performance assessment", "data mining", "big data", and "deep learning". This suggests that institutions are integrating AI-driven methodologies to enhance accuracy, efficiency, and predictive capabilities in performance evaluations. The connections between data mining, deep learning, and academic performance imply that machine learning models are being employed to analyze faculty effectiveness, student outcomes, and instructional quality. Furthermore, the presence of teaching performance evaluation within this cluster suggests a growing trend in using data-driven tools

to track teaching effectiveness, engagement levels, and student learning progress. This cluster underscores the transformation of performance evaluation from traditional subjective evaluations to objective, AI-enhanced systems, leveraging big data analytics to predict and improve faculty performance. Institutions adopting these approaches aim to develop more transparent, evidence-based, and dynamic evaluation frameworks that align with technological advancements in education.

The red cluster with 25 keywords represents faculty evaluations that are competency-based and student-feedback oriented, indicating a shift toward a student-centered evaluation model in higher education. Key terms such as “competence,” “flipped learning,” “simulation,” “performance feedback,” and “practice-based learning” highlight the increasing emphasis on faculty performance evaluation methods that prioritize student engagement, hands-on learning experiences, and outcomes-based evaluation. The presence of simulation and practice-based learning suggests a particular focus on professional and medical education, where faculty effectiveness is measured through interactive teaching methodologies. This trend is moving away from traditional lecture-based evaluation models and instead emphasizing practical application, critical thinking, and student skill development. Additionally, the inclusion of performance feedback and flipped learning reinforces the importance of structured feedback mechanisms and student participation in faculty evaluation. This indicates that faculty evaluations are increasingly incorporating direct student input, ensuring that teaching effectiveness is measured based on its impact on student learning and competency. Overall, this cluster outlines the evolution of an institution's evaluation system, aligning it with a competency-based education framework that emphasizes interactive learning, skill mastery, and student-driven evaluation models.

The purple cluster with 14 keywords highlights research on performance evaluation in educator education and policy development, emphasizing the role of institutional frameworks and accreditation standards in shaping educator evaluations. Key terms such as "teacher education," "educational policy," "teacher performance assessment," "evaluation," "competency," "learning outcomes," "validity," and "reliability" indicate a strong focus on developing standardized and evidence-based faculty evaluation tools. This suggests that research in this area is dedicated to improving the accuracy and effectiveness of educator evaluation methodologies by ensuring that evaluation tools are both valid and reliable. The validity and reliability of the instrument by using classical and modern theory will help to empower the quality of measurement including the criterion validity (Mohd Matore, et al., (2021).

The connection between educator education and educational policy implies an emphasis on preparing future educators through structured faculty training programs while aligning evaluation practices with national and institutional accreditation requirements. The presence of competency and learning outcomes reinforces the shift towards outcome-based evaluation models, ensuring that educator evaluations measure not just instructional delivery, but also their impact

on student learning. Additionally, the focus on validity and reliability suggests that ongoing research is dedicated to refining evaluation instruments to provide consistent, fair, and data-driven assessments of faculty performance. By using modern theory like item response theory or Rasch models with teaching evaluation instrument will also can solve psychometric problems in measurement such as previous researches (Mohd Noh & Mohd Matore, 2022; Sovey, 2022).

The blue cluster indicates 19 keywords is centered on performance appraisal in "universities", incorporating "faculty evaluation", "medical education", "program evaluation", and "professional development". These keywords suggest a focus on assessing faculty performance, particularly in medical education, where clinical teaching effectiveness and leadership development are crucial. Key terms such as "academic staff", "competencies", "human resource management", and "performance management" highlight that faculty evaluation extends beyond traditional teaching evaluations to include skills development, administrative responsibilities, and institutional contributions. The presence of "teaching evaluation" and "leadership" underscores the emphasis on mentorship, career progression, and structured professional growth initiatives within universities. The link to "universities" and "research" reinforces that performance appraisal in higher education is evolving to integrate data-driven decision-making, leadership training, and competency-based evaluations.

The orange cluster with 16 keywords reflects research on post-pandemic faculty evaluation trends, emphasizing the impact of COVID-19 on faculty performance evaluation. Key terms such as "COVID-19," "online learning," "universities," and "distance education" highlight the shift towards digital learning environments and the growing need for new evaluation models in higher education. The presence of online learning and distance education suggests an increased reliance on virtual teaching platforms, necessitating faculty evaluation frameworks that consider engagement, adaptability, and instructional effectiveness in digital classrooms. This transition has reshaped traditional faculty evaluations, prompting research into hybrid evaluation models that integrate both online and in-person teaching metrics. Additionally, the connection to universities signifies an institutional focus on adapting faculty performance appraisal to remote teaching conditions. Future research is likely to explore the long-term impact of online education on institutional effectiveness, the development of digital teaching competencies, and the refinement of evaluation criteria that reflect the new learning landscape. This cluster underscores the urgency for higher education institutions to establish robust, adaptable faculty evaluation strategies that align with the evolving demands of hybrid and online education models.

Based on the network visualization, technology-driven evaluation, competency-based faculty evaluations, educator education policies, medical education faculty evaluation, and post-pandemic teaching evaluation emerge as dominant themes in performance evaluation research. The increasing focus on AI-driven assessment tools, data-informed teaching feedback, and student-centered evaluation models suggests that future studies will continue to advance towards

automated, competency-based, and technology-enhanced performance evaluation methods in higher education.

The Figure 8 below illustrates the density visualization of keyword co-occurrence in performance evaluation research within higher education. This mapping provides insight into the research intensity of various topics, where brighter areas indicate high research activity, while darker regions represent relatively underexplored themes (Van Eck & Waltman, 2010).

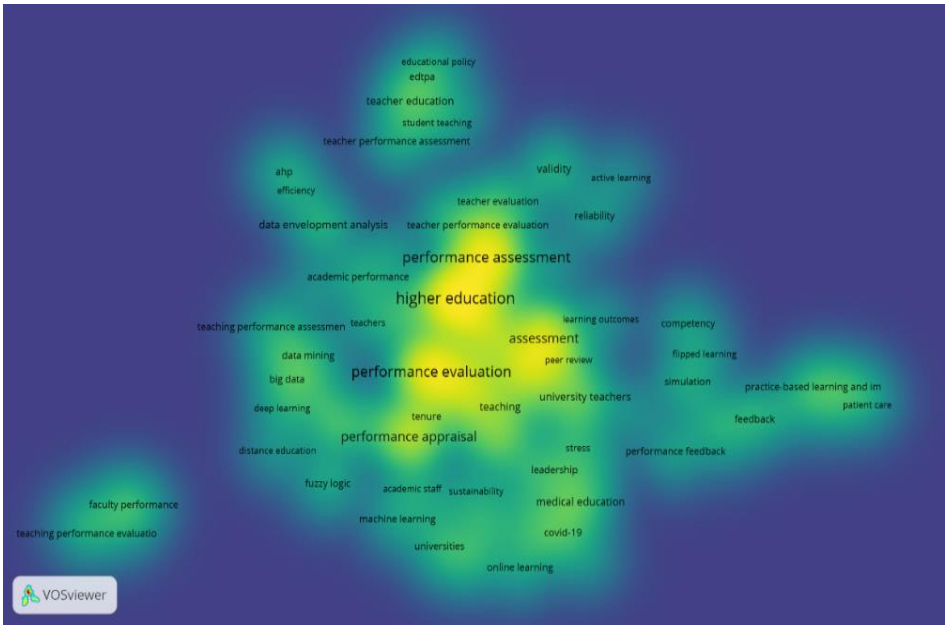

Figure 8. Density visualization of co-occurrence of keywords used by authors

The brightly colored regions represent frequently researched topics, including "performance evaluation", "performance assessment", "assessment", "performance appraisal", "higher education" and "university teachers". The prominence of these keywords suggests a significant focus on an institution's evaluation frameworks, teaching effectiveness, and structured evaluation methodologies in higher education. Additionally, themes related to "medical education", "competency-based assessment", "leadership", and "peer review" also appear in bright colors, indicating a high volume of studies in these areas.

The moderately bright regions encompass keywords such as "data mining", "big data", "deep learning", "fuzzy logic", and "machine learning", reflecting the increasing adoption of technology-driven evaluation methods. The integration of artificial intelligence and learning analytics in faculty evaluation is becoming a growing area of research, as institutions aim to enhance the accuracy and efficiency of performance evaluations. Similarly, topics like "online learning", "distance education", and "COVID-19" highlight the recent shift in faculty evaluations toward virtual teaching effectiveness and digital education frameworks due to the pandemic's impact. In contrast, darker areas represent research gaps, indicating themes that have not been extensively explored. Keywords such as "faculty performance", "teaching performance evaluation",

"performance feedback", including "self-essessment", and "peer assessment" in dim colors, suggesting that further investigations are needed in these areas. Although the emphasis on institutional evaluation is structured, research on individual institutional development, specific teaching performance evaluation, and evaluation models are among the disciplines that are still limited.

5. Conclusions

This research examined 1,173 publications on performance evaluation for educators in higher education from 2000 to 2024, covering 86.5% (1,173 out of 1,356) of the total studies on performance evaluation available in the Scopus database between 1973 and 2024. The initial phase of metadata trend contributors, identifying year metadata, identifying public contributors. countries, influential institutions, leading authors, leading journals, and dominant research themes. Findings indicate a gradual shift from conventional evaluation methods to more structured, data-driven, and competency-based evaluation models. The integration of artificial intelligence, learning analytics, and sustainability-driven performance indicators is becoming increasingly prevalent, reflecting efforts to increase transparency, accountability, and objectivity in faculty evaluation.

Despite its contributions, this study is subject to certain limitations. Data collection was limited to Scopus-indexed publications, which may have excluded relevant studies from databases such as Web of Science, Google Scholar, or institutional archives. Bibliometric approaches focus on quantitative patterns, citation structure, and keyword co-occurrence, potentially overlooking qualitative insights and practices practiced in the evaluation of a particular institution. In addition, regional gaps in higher education policies and institutional performance evaluation frameworks are not fully captured in global bibliometric analyses.

The results of this study have significant implications for higher education administrators, policymakers, and academic researchers seeking to refine faculty performance evaluation systems. The increasing reliance on technology-enhanced evaluation mechanisms, including real-time feedback tools, AI-powered evaluation models, and adaptive learning analytics, underscore the need to develop faculty evaluation strategies that are fair, transparent, and equitable. These findings suggest that institutions should adopt a hybrid evaluation model which is by integrating quantitative performance indicators with its own qualitative institutional development initiatives to achieve a comprehensive understanding of teaching effectiveness.

Future research should explore alternative educator evaluation methodologies, emphasizing mixed methods and interdisciplinary approaches to bridge the gap between quantitative data analysis and qualitative institutional development insights. The role of machine learning in predictive performance evaluation, bias reduction, and personalized faculty development pathways requires further investigation. Additionally, an in-depth examination of regional and cultural variations in faculty evaluation frameworks can contribute to context-specific best practices for performance evaluation in different higher education systems. With the increasing focus on competency-based evaluation, AI-driven institutional monitoring, and sustainability-related evaluation frameworks, further research

should explore how these emerging trends are shaping the future of educator performance evaluation in higher education.

6. Implications and Recommendations

The results of this study have important ramifications for enhancing higher education frameworks for evaluating the effectiveness of educators. First, institutions must incorporate analytics-driven evaluation tools and artificial intelligence as a result of the move toward competency-based and data-driven assessments. This change addresses historical constraints related to subjectivity and inconsistency while guaranteeing objective and up-to-date insights into faculty performance. Second, frameworks for performance evaluation that are both standardized and flexible must be established by academic leaders and legislators. The findings highlight the need for a hybrid evaluation strategy that incorporates both qualitative and quantitative performance indicators. This method guarantees a fair evaluation system that accounts for both professional development and instructional efficacy. Third, the paper notes how AI is becoming more and more involved in faculty evaluations, which raises ethical questions about algorithmic bias and impartiality. To reduce biases and make sure AI-driven assessments are in line with institutional values and teaching philosophies, institutions should put strict oversight procedures in place. Fourth, more effort has to be done to improve competency-based frameworks to take into consideration disciplinary variations and different teaching approaches. Institutions of higher learning must offer faculty development programs that match institutional goals and instructional innovations with performance reviews. In conclusion, it is recommended that future studies investigate multidisciplinary evaluation models that integrate cultural influences on performance evaluations, adaptive learning technology, and sustainability measures. Higher education institutions can improve faculty accountability, teaching quality, and institutional credibility in the global education scene by promoting fair, open, and technologically advanced evaluation processes.

7. Acknowledgments

We are really grateful for the assistance that the Federal Training Prize (HLP) scholarship from the Ministry of Education (MOE) has given us. This research was funded by the Faculty of Education, Universiti Kebangsaan Malaysia (UKM) with grant number GG-2024-024 (Program STEM dan Minda). 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.

8. References

Abubakar, U., Ogunlade, O. O., & Ibrahim,H. A. (2024). The influence of technology-integratedcurriculum resources on student engagement and aca-demic achievement in higher education. *Advances in Mobile Learning Educational Research, 4(2)*, 1208-1223. https://doi.org/10.25082/AMLER.2024.02.014

Aguinis, H. (2019). *Performance management for dummies*. Wiley.

Alshaikhi, T., & Khasawneh, M. A. S. (2025). Enhancing teacher competence in differentiated instruction for English language learners with disabilities: A professional development intervention. *World Journal of English Language, 15(1)*, 101-115. https://doi.org/10.5430/wjel.v15n1p101

Altassan, M. A. (2022). Advancing sustainable development through performance appraisal systems: Critiques, innovations, and future trends. *WSEAS Transactions on Environment and Development, 19*, 1160–1178. https://doi.org/10.37394/232015.2023.19.107

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. https://doi.org/10.1016/j.joi.2017.08.007

Bennett, R. E. (2025). Rethinking equity and assessment through opportunity to learn. *Assessment in Education: Principles, Policy & Practice*, 1–28. https://doi.org/10.1080/0969594X.2025.2462549

Chen, X., & Chen, R. (2024). Examining the influence of the performance of mobile devices on teaching quality based on the technological acceptance mode. *Library Hi Tech*, 42(2), 670–695. https://doi.org/10.1108/LHT-12-2021-0489

Chen, Y., Pan, Y., Liu, H., Wu, H., & Deng, G. (2023). Efficiency analysis of Chinese universities with shared inputs: An aggregated two-stage network DEA approach. *Socio-Economic Planning Sciences, 90*, 101728. https://doi.org/10.1016/j.seps.2023.101728

Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA: Stanford Center for Opportunity Policy in Education. https://edpolicy.stanford.edu

Do, Q. H., Tran, V. T., & Tran, T. T. (2024). Evaluating lecturer performance in Vietnam: An application of fuzzy AHP and fuzzy TOPSIS methods. Heliyon, 10(11). https://doi.org/10.1016/j.heliyon.2024.e30772

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

Eun Joo Park, & Sanghee Lee. (2022). Creative thinking in the architecture design studio: Bibliometric analysis and literature review. https://doi.org/10.3390/buildings12060828

Epstein, M. J. (2018). *Making sustainability work: Best practices in managing and measuring corporate social, environmental, and economic impacts (2nd ed.)*. Routledge.

Firmansyah, J., Suhandi, A., Setiawan, A., & Permanasari, A. (2022). Evaluation of virtual workspace laboratory: Cloud communication and collaborative work on project-based laboratory. *Journal on Efficiency and Responsibility in Education and Science, 15*(3), 131–141. https://doi.org/10.7160/eriesj.2022.150301

Hao, H. (2024). The relationship between faculty emotional intelligence and student engagement in blended learning environments. *Journal of Education and e-Learning Research, 11*(2), 159–167. https://doi.org/10.20448/journal.509.2024.112.159.167

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics, 106,* 787-804. https://doi.org/10.1007/s11192-015-1765-5

Hou, W. (2024). The establishment of a performance evaluation model using education informatization to evaluate teacher morality construction in colleges and universities. *Open Computer Science, 14(1)*. https://doi.org/10.1515/comp-2023-0101

Janavi, E., & Abdi, S. (2024). Faculty members' performance appraisal system: A bibliometric analysis of the scientific literature. *International Journal of Information Science and Management, 22(3)*, 63–83. https://doi.org/10.22034/ijism.2024.2013962.1281

Jin, R., Gao, S., Cheshmehzangi, A., & Aboagye-Nimo, E. (2018). A holistic review of off-site construction literature published between 2008 and 2018. In *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2018.08.195

Khan, F. U., & Badulescu, D. (2025). Sustainability pressures unveiled: Navigating the role of organizational sustainable culture in promoting sustainability performance. *Sustainability (Switzerland), 17(3)*. https://doi.org/10.3390/su17031322

Liu, C. (2024). Automating student assessment using digital data to improve education management effectiveness in higher education institutions. *Education and Information Technologies, 29*(2), 1885–1901. https://doi.org/10.1007/s10639-023-11898-z

Liu, Y., & Chen, D. (2024). Design of a virtual reality serious game for experiencing the colors of Dunhuang frescoes. *Heritage Science, 12(1)*. https://doi.org/10.1186/s40494-024-01477-x

Mahmud, A., Susilowati, N., & Sari, P. N. (2024). Assessment methods and learning styles for learning satisfaction: the role of self-efficacy. *International Journal of Evaluation and Research in Education, 13(6)*, 515-526. https://doi.org/10.11591/ijere.v13i6.29757

Mamatha, H. K., Sridhar, R., & Somanatha, B. (2024). Multicriteria decision analysis tool for faculty performance evaluation in higher education institutions. *Multidisciplinary Science Journal, 6(7)*, 0-0. https://doi.org/10.31893/multiscience.2024118

Masdoki, M., Din, R., & Mohd Matore, M. E. E. (2021). Teaching 4.0 competency in higher learning institutions: A systematic mapping review. *International Journal of Learning, Teaching and Educational Research, 20*(10), 217–231. https://doi.org/10.26803/ijlter.20.10.12

Maturana, M., et al. (2025). Exploring the role of role-theoretical concepts: Determining dimensionality of difficulty in cross-occupational collaboration. *Vocations and Learning, 18(1)*. https://doi.org/10.1007/s12186-025-09361-w

Mohd Matore, M. E. E., Khairani, A. Z., Mohamad, S. M., Alias, N., & Mohd Mokhtar, E. R. (2018). The influence of intellectual quotient (IQ), emotional quotient (EQ) and spiritual quotient (SQ) against adversity quotient (AQ) on polytechnic students in Malaysia. *Journal of Engineering Science and Technology, 13*(Special Issue on ICETVESS 2017), 83–91.

Mohd Matore, M. E. E., Zainal, M. A., Mohd Noh, M. F., Khairani, A. Z., & Abd Razak, N. (2021). The development and psychometric assessment of Malaysian Youth Adversity Quotient Instrument (MY-AQi) by combining Rasch model and confirmatory factor analysis. *IEEE Access, 9*, 13314–13329. https://doi.org/10.1109/ACCESS.2021.3050311

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*, Article 941084. https://doi.org/10.3389/fpsyg.2022.941084

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics, 106, 213-228.

Musid, N. A., Matore, M. E. E. M., & Hamid, A. H. A. (2024). A scoping review on mapping the digital leadership constructs for educational settings: What we can learn? *International Journal of Evaluation and Research in Education, 13(1)*, 111-121. https://doi.org/10.11591/ijere.v13i1.26760

Pifer, M. J., Logsdon, M. C., Ibarra, M., & Gardner, K. (2024). Advancing scholarly excellence at midcareer: The influence of exceptional others on faculty professional growth. *International Journal of Mentoring and Coaching in Education, 13*(2), 230–245. https://doi.org/10.1108/IJMCE-08-2023-0076

Rahimi, Z., Badiyepeymaiejahromi, Z., & Taghizadeganzadeh, M. (2024). 360-degree evaluation of nursing students' performance. *Shiraz E Medical Journal, 25*(11), e148872. https://doi.org/10.5812/semj-148872

Rogers, J. K. B., Mercado, T. C. R., & Decano, R. S. (2025). Moodle interactions and academic performance: Educational data mining in a Philippine university. *Journal of Education and Learning, 19(1)*, 542–550. https://doi.org/10.11591/edulearn.v19i1.21549

Sabharwal, R., & Miah, S. J. (2024). The role of social media in spreading the culture of volunteer work among university students. *Studies in Media and Communication, 12(4),* 201–213. https://doi.org/10.11114/smc.v12i4.6941

Sabtu, S. H., & Matore, M. E. E. M. (2024). Systematic literature review on the implementation of the Six Sigma approach in education. *International Journal of Evaluation and Research in Education, 13*(1), 262–270. https://doi.org/10.11591/ijere.v13i1.26196

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), Article em2088. https://doi.org/10.29333/ejmste/11794

Székely, F., & Knirsch, M. (2005). Responsible leadership and corporate social responsibility: Metrics for sustainable performance. *European Management Journal, 23(6)*, 628–647.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84(2),* 523–538. https://doi.org/10.1007/s11192-009-0146-3

Velázquez, M. del R. H., Macias, M. L. C., Soraida, Q. C. L., & Figueroa, M. S. Q. (2024). Evaluation of Competencies in Nursing Students. a Cross-Sectional Study. *Revista De Gestão Social E Ambiental*, *18*(8), e08219. https://doi.org/10.24857/rgsa.v18n8-167

Ye, Y., Zhu, H., Cui, T., Yu, R., Zhang, L., & Xiong, H. (2024). University evaluation through graduate employment prediction: An influence-based graph autoencoder approach. *IEEE Transactions on Knowledge and Data Engineering, 36*(11), 7255–7267. https://doi.org/10.1109/TKDE.2024.3402234

Zaabalawi, R.S., Zaabalawi, J. Portfolios versus exams: a study to gauge the better student assessment tool. *Lang Test Asia* 14, 28 (2024). https://doi.org/10.1186/s40468-024-00296-y

Zheng, Y., & Sun, C. (2022). Construction scheme of quantitative assessment of college teachers' performance based on big data analysis. *Mobile Information Systems, 2022*, Article 8373164. https://doi.org/10.1155/2022/8373164

Zhu, J. (2024). Key challenges and countermeasures: A review of undergraduate teaching of neurology in outpatient settings. *BMC Medical Education, 25(1).* https://doi.org/10.1186/s12909-024-06601-w

Zhu, Z., & Chang, Y. (2025). The influence of work stress on the work performance of university teachers in Zhejiang Province, China: The mediating role of work happiness. *Journal of Educational and Social Research, 15(1)*, 42-55. https://doi.org/10.36941/jesr-2025-0004

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*. https://doi.org/10.1177/1094428114562629