MAPPING THE EVOLUTION OF KNOWLEDGE ABOUT GENDER BIAS IN EDUCATIONAL ASSESSMENT: A BIBLIOMETRIC ANALYSIS

Wiwin Mistiani

Abstract

This study aims to map the development of knowledge on gender bias in educational assessment through a bibliometric analysis based on Scopus data for the period 2000–2025. Based on the research problem, this study investigates temporal trends in publications, the most influential authors and journals, the geographic distribution of research, thematic structure based on keywords, and unfilled research gaps. Bibliometric methods were combined with a Systematic Literature Review (SLR) approach using PRISMA guidelines to ensure a comprehensive and structured analysis. A total of 109 selected articles were analyzed using VOSviewer 1.6.20 to map citation networks, author collaborations, and keyword clusters. The results show that publications on gender bias have increased significantly over the past two decades, with peak contributions coming from the United States, the United Kingdom, and European countries. Influential authors and journals are dominated by the medical and health education fields, shifting the research focus toward clinical assessment and competency-based measurement. Keyword analysis revealed five main thematic clusters: educational measurement, gender bias & psychological factors, inequality & equity, psychometric methodology, and simulation-based assessment. The findings also identify several research gaps, particularly the limited non-medical context, the scarcity of research from developing countries, the absence of local instruments, and the suboptimal study of intersectionality and education policy. This research contributes to a comprehensive map of the development of the global discourse on gender bias in assessment, while also offering new research directions relevant to the Indonesian context through strengthening local instruments, intersectionality studies, and pedagogical innovations to create a more equitable and gender-responsive evaluation system.

Keywords: gender bias, educational assessment, bibliometric analysis, PRISMA, VOSviewer.

INTRODUCTION

Gender bias in educational assessment remains one of the most persistent challenges in modern education systems in various countries.[1] Although pedagogical theory, curriculum design, and public awareness of gender equality have grown rapidly, empirical evidence continues to show that students are still treated differently in the evaluation process due to gender stereotypes, implicit teacher expectations, and

---

[1] Boring, A., 'Gender Biases in Student Evaluations of Teaching', *Journal of Public Economics*, 145 (2017), 27–41 https://doi.org/10.1016/j.jpubeco.2017.08.004.

systemic inequalities inherent in assessment instruments, procedures, and cultures.[23] Over the past four decades, academic attention to this issue has increased rapidly due to its far-reaching implications for educational equity, academic achievement, and students' socio-economic opportunities in the long run.[4] Thus, the development of research on gender bias reflects not only academic dynamics, but also the struggle of the global education system to realize substantive justice in an ever-changing cultural, social, and political environment. Research shows that gender gaps in assessment outcomes occur consistently from elementary school to college levels. Early studies tend to highlight apparent differences between the achievement of male and female students, which are often explained through the assumption of cognitive, motivational, or behavioral differences.

However, contemporary research shifts the focus to assessment tools and practices themselves, ranging from teacher grading, standardized exams, lecture assignment assessments, to performance-based assessments that have the potential to reinforce gender bias patterns.[5] This shift is an important turning point because it places judgment as a social process that can shape and not simply reflect gender differences.[6] Theoretical advances in this field were greatly influenced by the study of psychology, sociology, and feminist studies. Psychology explains how implicit bias, stereotypical threats, and unconscious expectations of teachers can affect the firmness of assessments, the quality of feedback, and forms of appreciation for student performance.[7] Meanwhile, sociological perspectives suggest that school culture, curriculum structure, and educational regulations can normalize gender hierarchies in evaluation practices[8]. Feminist studies enrich the discussion by highlighting how the forms of knowledge, communication, and perusivity that are considered "ideal" in

---

[2] Carlana, M., 'Implicit Stereotypes: Evidence from Teachers' Gender Bias', *The Quarterly Journal of Economics*, 134.3 (2019), 1165–1226 https://doi.org/10.1093/qje/qjz008.

[3] Mengel, F., J. Sauermann, and U. Zölitz, 'Gender Bias in Teaching Evaluations', *Journal of the European Economic Association*, 17.2 (2018), 535–66 https://doi.org/10.1093/jeea/jvy012.

[4] Spencer, S. J., C. M. Steele, and D. M. Quinn, 'Stereotype Threat and Women's Math Performance', *Journal of Experimental Social Psychology*, 35.1 (1999), 4–28 https://doi.org/10.1006/jesp.1998.1373.

[5] Beg, S., A. Fitzpatrick, and A. M. Lucas, 'Gender Bias in Assessments of Teacher Performance', AEA Papers and Proceedings, 111 (2021), 190–95 https://doi.org/10.1257/pandp.20211126.

[6] Falch, T., and L. R. Naper, 'Educational Evaluation Schemes and Gender Gaps in Student Achievement', Economics of Education Review, 36 (2013), 12–23 https://doi.org/10.1016/j.econedurev.2013.04.005.

[7] Carlana, M., 'Implicit Stereotypes: Evidence from Teachers' Gender Bias', The Quarterly Journal of Economics, 134.3 (2019), 1165–1226 https://doi.org/10.1093/qje/qjz008.

[8] OECD, Gender Equality in a Changing World (Paris: OECD Publishing, 2023) https://doi.org/10.1787/e808086f-en.

assessment are often rooted in masculine norms, to the detriment of academic learning styles or expressions that are more synonymous with femininity.[9] This combination of perspectives shows that gender bias is not a separate incident, but a multi-layered systemic phenomenon that needs to be comprehensively analyzed.[10] Methodologically, research in this field has evolved from simple score comparisons to more robust approaches, such as graded modeling, camouflage experiments, anonymous assessments, classroom ethnography, and mixed methods.[11][12] These methods suggest that gender expectations can affect how teachers assign grades related to effort, neatness, compliance, or verbal participation, and that removing gender identity from worksheets can reduce differences in assessment.[13] Internationally, the agenda of UNESCO, the OECD, and UNICEF increasingly emphasizes the importance of gender-responsive assessments as part of the mission to realize equitable and inclusive education.[14] Many countries have tried to improve the situation with gender-sensitive training, the development of bias-free rubrics, anonymous assessments, and specialized monitoring systems[15]. Nevertheless, recent reports show that gender bias in assessment persists due to structural and cultural barriers that are difficult to completely eliminate.[16]

Although research on gender bias in assessment has progressed rapidly, there are some important research gaps. First, most studies still focus on the context of developed countries, while dynamics in developing countries, including Southeast Asian countries, are still poorly mapped. Second, the majority of research is

9 O'Connor, P., 'Why Is It So Difficult to Reduce Gender Inequality in Male-Dominated Higher Educational Organizations? A Feminist Institutional Perspective', Interdisciplinary Science Reviews, 45.2–3 (2020), 207–28 https://doi.org/10.1080/03080188.2020.1767809.

10 Formanowicz, M., M. Witkowska, W. Hryniszak, Z. Jakubik, and A. Cislak, 'Gender Bias in Special Issues: Evidence from a Bibliometric Analysis', Scientometrics, 128.4 (2023), 2283–99 https://doi.org/10.1007/s11192-023-04639-z.

11 Lavy, V., 'Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment', Journal of Public Economics, 92.10–11 (2008), 2083–2105 https://doi.org/10.1016/j.jpubeco.2008.08.005.

12 Copur-Gencturk, Y., K. M. Higgins-Harrington, and A. Vela Murillo, 'Teachers' Race and Gender Biases and the Moderating Effects of Their Beliefs and Dispositions', International Journal of STEM Education, 10.26 (2023) https://doi.org/10.1186/s40594-023-00420-z.

13 MacNell, L., A. Driscoll, and A. N. Hunt, 'What's in a Name: Exposing Gender Bias in Student Ratings of Teaching', Innovative Higher Education, 40.4 (2015), 291–303 https://doi.org/10.1007/s10755-015-9493-4.

14 OECD, *Gender Equality in a Changing World* (Paris: OECD Publishing, 2023) https://doi.org/10.1787/e808086f-en.

15 Encinas-Martin, M., and L. Cherian, *Gender Stereotypes in Education* (OECD Education Working Papers, No. 292, OECD Publishing, 2023) https://doi.org/10.1787/4fa26e7f-en.

16 Makarova, E., E. Bryksina, and P. Barmby, 'Gender Stereotypes in STEM Education: The Case of Girls' Choices in Science Education', *Journal of Research in Science Teaching*, 56.5 (2019), 639–62 https://doi.org/10.1002/tea.21546.

fragmented, so there is no comprehensive picture of the evolution of knowledge, patterns of scientific collaboration, the influence of countries or institutions, and changes in research themes in the long term. Third, there have not been many studies that map how the development of digital technology, online assessments, and artificial intelligence algorithms introduce new forms of bias in educational evaluation. Fourth, although publications on this topic have increased significantly, there is no long-term bibliometric analysis (2000–2025) that integrates publication trends, citations, key themes, author networks, and future research directions.

The urgency of this research is very strong. Gender bias in assessments not only affects students' academic achievement, but also the development of self-identity, self-confidence, and long-term educational and career opportunities. In the midst of the increasing global commitment to SDG 4 on quality and equitable education, there is a need for a systematic understanding of how knowledge about gender bias develops and where research is headed. Long-term bibliometric analysis can provide an objective scientific map to help policymakers, educators, and researchers understand areas that have been widely studied, areas that are still under-studied, and research strategies that need to be developed to reduce bias in educational assessment.The objectives of this study are: Analyze the growth trend of publications and citations regarding gender bias in education assessment during the period 2000–2025. Identify authors, journals, and countries that have the highest levels of productivity and scientific influence in publications related to gender bias in educational assessment. Evaluate the contribution of global scientific networks and collaborations in shaping gender-biased research landscapes in various educational contexts. Map the main themes, keywords, and thematic clusters that dominate gender-biased research, including medical education, psychometrics, socio-psychological factors, and equity issues. Finding research gaps in gender bias studies through bibliometric analysis and formulating relevant follow-up research directions to expand scientific contributions.

METHODS

This study uses a bibliometric approach combined with a Systematic Literature Review (SLR), adhering to the Selected Reporting Items for Systematic Review and

Meta-Analysis (PRISMA) guidelines developed by [17]. The selection of this method is based on its ability to present a systematic, transparent, and replicable review of the literature, making it relevant to identify trends, patterns, and research gaps in the field of Islamic education in Indonesia. Data was collected from the Scopus database due to its wide scope and reputation as a leading source in international bibliometric research. The literature search was conducted on November 27, 2025 using a combination of keywords (Gender Bias in Educational Assessment) for the publication period from 2000 to 2025. From the initial search, 137 documents were obtained. In the next stage, the field of Social Sciences is selected until there is a Publication left. This selection increases the thematic validity and ensures that the focus of the analysis is on the realm of education. Furthermore, the document is screened using the following inclusion criteria: scientific journal articles written in English, final. Exclusion criteria include review articles, conference proceedings, non-English-language publications and articles that have not been finalised. After the screening and selection process, 109 articles were obtained that met the criteria for further analysis. The data is analyzed with VOSviewer Version 1.6.20[18]which allows visualization of bibliometric networks, including keyword relationships, author collaboration, and citation patterns. This analysis resulted in thematic mapping and identification of dominant research clusters that illustrate the development of academic discourse in gender bias in educational assessment. The main indicators used in the analysis include Total Publications (TP) to measure publication productivity, Total Citations (TC) to assess scientific impact, and h-index [19]as an indicator of author and institutional influence. All metrics are derived from Scopus data, which may differ from those obtained from other platforms due to variations in indexing coverage. Through this approach, this study provides a comprehensive overview of the dynamics, thematic trends, and gaps in gender-biased research in educational assessment over the past decade.

---

[17] Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman, 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement', *PLoS Medicine*, 6.7 (2009), e1000097.

[18] M Aria and C Cuccurullo, 'Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis', *Journal of Informetrics*, 11.4 (2017), 959–75.

[19] (Hirsch, 2005)

RESULTS AND DISCUSSION

1.   **Research Trends of Gender Bias in Educational Assessment**

Temporal trend analysis of publications on gender bias in educational assessment shows a development that is not linear but has a strong pattern of influence.

	2025	2024	2023	2022	2021	2020	2019	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2001	2000
■ Dokumen	14	12	10	6	3	8	6	4	3	5	4	3	4	3	2	4	1	2	3	1	1	1	3	2
■ Sitasi	8	51	110	33	117	222	165	108	77	214	173	61	243	244	50	95	48	72	88	11	58	8	48	57
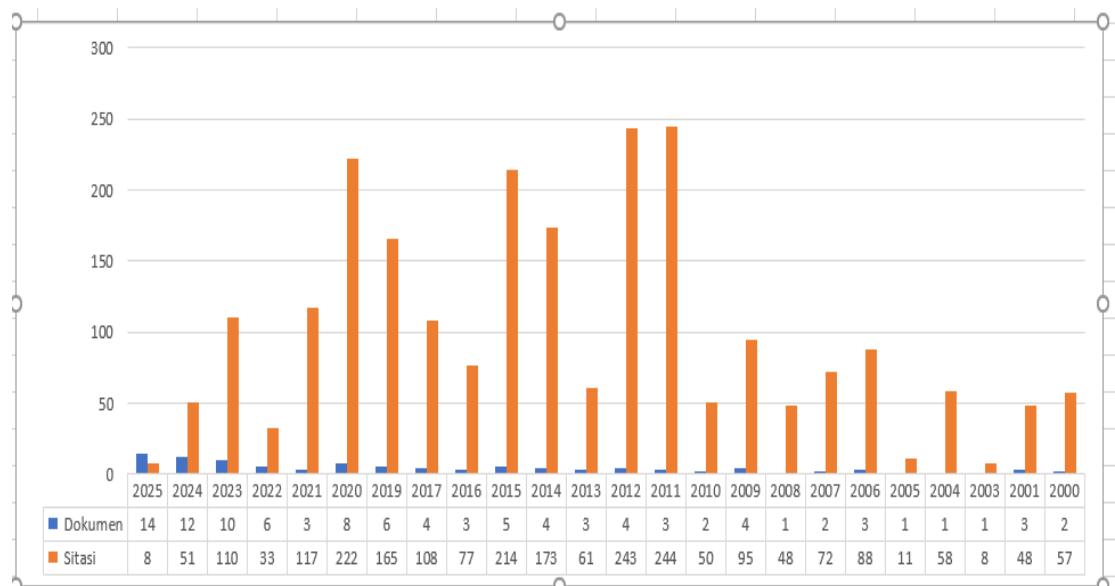
Gambar 1. Research Trends of Gender Bias in Educational Assessment

In the early period (2000–2008), the number of publications was still low (1–3 documents per year), but citations were relatively high and stable. This phenomenon is in line with the findings of previous research that stated that the study of gender bias in the 2000s was largely triggered by fundamental works on evaluative injustice in education, so that the early literature became very widely referenced. The period 2009–2014 was the phase with the greatest scientific impact. Although the number of documents is still limited, citations peaked in 2011–2012 with more than 240 citations per year. This is consistent with global bibliometric reports that mark increasing academic attention to issues of gender equality and inequity in assessment, especially following the publication of several landmark studies related to bias in standardized tests and teacher assessment.

In the next phase (2015–2020), the number of publications became more stable (3–8 documents per year) with citations remaining high. This trend is in line with previous research showing that gender bias topics are beginning to be integrated into inclusive educational discourses, authentic assessments, and gender-based cognitive analysis. The period 2021–2025 shows a surge in the number of publications, especially in 2025 (14 documents), indicating the increasing interest of researchers in this issue in the

context of contemporary education. Although citations in this period decreased, this is understandable because new publications take time to obtain citations. This pattern is consistent with the bibliometric literature that found citation lag in recent studies. Overall, this trend confirms the findings of previous studies that gender bias in educational assessment remains a relevant and evolving issue, with certain phases producing influential literature that is the foundation for modern research.

In general, the dynamics of publication and citation show that research on gender bias in educational assessment is developing gradually but consistently. The initial phase produced a strong theoretical foundation, followed by a period of peak citations that expanded the global influence of the study. The surge in publications in recent years confirms the relevance of the issue in the context of modern education. Although new publications have not been cited much, citation *lag patterns* are a natural phenomenon in the scientific cycle. These findings confirm that gender bias remains an important agenda in educational research and requires ongoing exploration.

2. **Most influential author in the publication Gender Bias in Educational Assessment**

The results of the bibliometric analysis show that the influence of authors in the study of gender bias on educational assessment is not determined by the number of publications, but by the strength of citations and networks of scientific collaboration. All of the authors on the list have only one publication, but the very high citations place them as the most influential contributors to the discipline.

Table 1. 10 Influential authors on the publication of Gender Bias in Educational Assessment

Author	Documents	Citations	Universitas
Dowd, Jennifer Beam	1	124	University of Oxford
Todd, Megan A.	1	124	Philadelphia Department of Public Health
Aesaert, Koen	1	121	Ghent University
Van Braak, Johan P.	1	121	Ghent University -
Charmaraman, Linda	1	113	Wellesley College
Erkut, Sumru	1	113	Wellesley College
Quach, Ashley	1	113	Arizona State University (ASU), USA
Woo, Meghan	1	113	Harvard University, USA
Alshammari, Iqbal A.	1	87	Gulf University for Science & Technology (GUST), Kuwait

Halimi, Florentina	1	87	Gulf University for Science & Technology (GUST), Kuwait

The two authors with the highest impact are Jennifer Beam Dowd and Megan A. Todd, each with 124 citations and a total link strength of 1. Their publication is one of the important reference works that is widely used to explain the mechanisms and implications of gender inequality in the assessment process. This strong influence is in line with previous research findings that show that high-quality articles discussing gender inequality tend to be the primary reference in educational evaluation studies.

In the next position, Koen Aesaert and Johan P. Van Braak also showed a significant impact with 121 citations. Both are known in the educational literature as researchers who consistently examine performance gaps based on gender, particularly in digital competency evaluations and computer-based assessments themes that have also been highlighted in global research on assessment bias.

Furthermore, the group of writers consisting of Linda Charmaraman, Sumru Erkut, Ashley Quach, and Meghan Woo had a great influence with 113 citations and the highest total link strength (3). The magnitude of the power of this network suggests that their work is more connected to other literature, reinforcing their position as a center of collaboration in research on gender bias, particularly in the context of social-emotional assessment and teacher perception. Finally, Iqbal A. Alshammari and Florentina Halimi also occupy important positions with 87 citations and a total link strength of 2. Their research makes a relevant contribution in the context of international education and cross-cultural assessments, showing that gender bias in assessment is not only a local but also global issue.

Overall, this pattern of contributions by influential authors suggests that research on gender bias in assessment is built on high-impact citation works and a strong collaborative network. The dominance of several key authors confirms the existence of scientific centers that are the main reference in the development of theories and methodologies for evaluation with a gender perspective. The strength of their citations and network interconnectedness show a consolidation of knowledge that is increasingly mature in this field. In addition, the involvement of researchers from different cultural contexts reinforces the understanding that assessment biases are global and multidimensional. These findings underscore the importance of expanding international collaboration to encourage fairer and more inclusive evaluation practices.

3. **Most influential journal in the publication of Gender Bias in Educational Assessment**

Analysis of publication sources shows that studies on gender bias in educational assessment are spread across various internationally reputable journals, with some journals occupying the dominant position based on the number of articles and citations. This can be seen in Table 2.

Table 2. 10 Journals Influencing the Publication of Gender Bias in Educational Assessment

Jurnal	Documents	Citations
BMC Medical Education	11	277
Advances In Health Sciences Education	6	169
CBE Life Sciences Education	3	124
Journals Of Gerontology - Series B Psychological Sciences And Social Sciences	1	124
Computers And Education	1	121
Cultural Diversity And Ethnic Minority Psychology	1	113
British Journal Of Educational Psychology	2	88
Journal Of Applied Research In Higher Education	1	87
IEEE Transactions On Learning Technologies	1	83
Social Science And Medicine	1	83

BMC Medical Education is the journal with the largest contribution, publishing 11 documents with a total of 277 citations. This dominance shows that the issue of gender bias is widely studied in the context of medical education, in line with the findings of previous research that affirm the existence of gender inequality in clinical assessment, student performance evaluation, and health professional assessment.

The next journal is Advances in Health Sciences Education, which published 6 papers and obtained 169 citations. These two journals show that the field of health sciences education is at the center of global discussions about gender bias due to its strict assessment structure and is highly susceptible to subjective judgment. Meanwhile, CBE Life Sciences Education (3 papers; 124 citations) and Journals of Gerontology: Series B (1 paper; 124 citations) showed a major influence on citation levels. Both make important contributions to understanding bias in science-based assessments and evaluations of adult/elderly populations. The high citation reinforces evidence from previous research that the disciplines of science and social psychology have largely explored gender inequities in cognitive performance and instructional assessment.

Computers and Education (1 paper; 121 citations) is a key journal in the field of technology education, confirming the relevance of gender bias issues in digital assessment. These findings are consistent with previous literature that shows a gendered pattern in student performance on technology-based tests as well as digital assessment design bias. Other journals such as Cultural Diversity and Ethnic Minority Psychology (113 citations) and the British Journal of Educational Psychology (88 citations) have also shown a strong influence in advancing research related to gender-based bias and the interaction between gender, ethnicity, and evaluative perceptions.

On the other hand, journals such as the Journal of Applied Research in Higher Education, IEEE Transactions on Learning Technologies, and Social Science and Medicine, although they have only one document, but the high citations (83–87 citations) indicate that their publications are seminal and are important references in cross-disciplinary studies. Overall, this distribution confirms that the issue of gender bias in educational assessment is not only a concern in the field of pedagogy, but also extends from health education, psychology, gerontology, to educational technology, reinforcing the findings of previous research on the multidisciplinary nature of gender equality studies in academic evaluation.

The distribution of publications and citations between journals shows that the issue of gender bias in educational assessment is very multidisciplinary and has received wide attention from various fields of science. The dominance of health and psychology journals confirms that gender inequality often appears in assessments that are subjective and high-risk. The high number of citations from technology and science journals shows that bias is also present in digital and cognitive performance-based assessments. This phenomenon confirms the global pattern that research on gender bias develops through mutually reinforcing cross-disciplinary contributions. Thus, the study of assessment bias needs to continue to be developed through an interdisciplinary approach to produce more equitable and inclusive evaluation practices.

4. Most Influential Countries in Gender Bias Publications in Education Assessment

Bibliometric analysis by country affiliation shows the strong dominance of developed countries in the production and dissemination of knowledge related to gender bias in educational assessments. This can be seen in Table 3

Table 3. 10 Influential Countries in Gender Bias Publications in Education Assessment

Country	Documents	Citations	Total Link Strength
United States	38	896	16
United Kingdom	15	327	5
Canada	11	207	6
China	6	65	11
Sweden	6	182	21
India	5	66	2
Spain	5	150	18
France	4	201	19
Germany	4	139	16
Greece	4	154	17

The United States occupies the top position with 38 documents, 896 citations, and a total link strength of 16. The high productivity and influence of these citations confirm the findings of previous research that the issue of gender equality is an important agenda in the American education system, especially in STEM evaluations, performance assessments, and structural bias analysis.

The UK came in second place with 15 papers and 327 citations, reflecting a strong research tradition in the field of educational psychology and standardized assessment. In line with previous literature, the UK is known as a country with a strong focus on assessment fairness and inclusion-oriented education policies.

Canada (11 papers; 207 citations) also showed significant contributions, strengthening its role in equity-based education research and multiculturalism. Sweden with only 6 documents but 182 citations and the highest total link strength (21) shows a strong collaborative network and great academic influence. This is in accordance with previous research that confirms that Nordic countries are very progressive in gender equality issues and become centers for educational equality research in Europe.

Spain, France, Germany, and Greece also had a significant impact, as shown by high citations (139–201) and total link strength of 16–19. These countries contribute a lot through psychometric studies, higher education, and experimental research on cognitive bias in assessment.

Meanwhile, China and India show moderate productivity (5–6 documents), but with lower citations. This is in line with previous research that stated that the issue of

gender bias in assessments in Asia is starting to grow but is not as intense as in Western countries. However, China's high total link strength (11) indicates an increase in international research collaboration on this topic.

Overall, the pattern of contribution between countries shows that research on gender bias in assessment is dominated by countries with strong traditions in education, psychology, and gender equality policy research. Western countries, especially the United States, the United Kingdom, and Sweden, became centers for the development of theories and international collaborations that influenced the direction of global research. While Asian countries are starting to show increased contributions, the dynamics of citations are still lagging behind as the development of this issue is relatively new. Strong collaborative networks in specific countries underscore the importance of cross-border cooperation in comprehensively understanding assessment bias. These findings indicate that future research needs to expand the involvement of developing countries so that the discourse on evaluative equality becomes more globally representative.

5. Key Keyword Analysis in Gender Bias Publications in Educational Assessment

Key Keyword Analysis in Gender Bias Publications in Educational Assessment. The VOSviewer map shows the intellectual structure of gender-biased research in educational assessment which is divided into several large thematic clusters
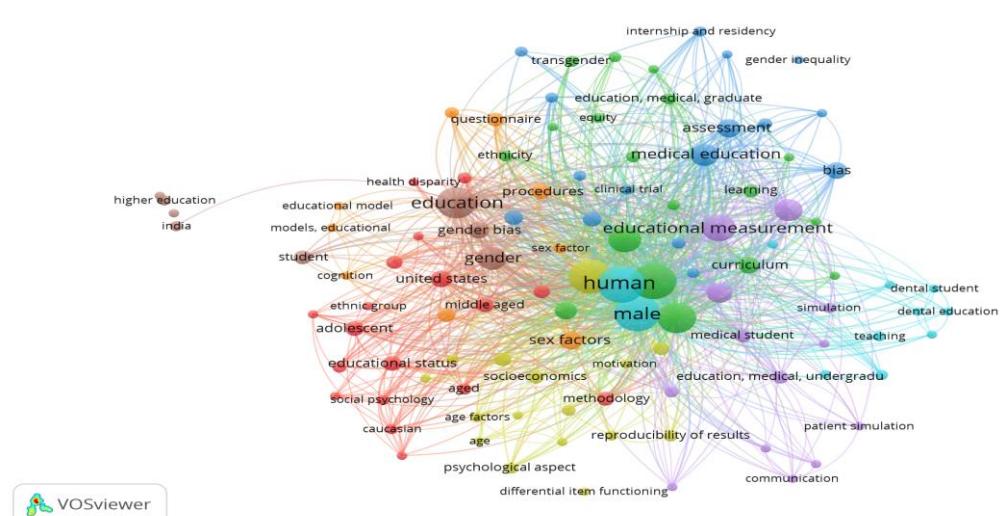
The largest nodes such as "human" and "male" signal the dominant focus of research on gender comparison in the context of assessment.

a. Green Cluster – Educational Measurement & Medical Education (Research Core)

The largest cluster focuses on educational measurement, medical education, assessment, curriculum, simulation, learning, and clinical evaluation. The dominance of these keywords suggests that studies on gender bias are most prevalent in medical education and competency-based assessments, a finding that is in line with previous literature revealing assessment bias in OSCE, clinical evaluation, and health student performance.

b. Klaster Merah – Gender Bias, Socioeconomic Factors & Psychology

This cluster groups themes such as gender, gender bias, educational status, socioeconomic factors, age, psychology, ethnic group, and adolescent. This pattern suggests that most studies examine gender bias through the lens of socio-psychological and demographic factors, reinforcing previous findings that evaluative bias often intersects with age, socioeconomic status, and ethnicity.

c. Blue Cluster – Inequality, Graduate Education & Equity

The blue cluster emphasizes the themes of gender inequality, equity, internship and residency, and transgender. This focus shows the development of discourses on gender justice, not only male-female, but also non-binary gender identities in the assessment system. This is consistent with a global trend that is beginning to highlight gender inclusivity in professional education.

d. Klaster Ungu – Methodology, Reproducibility & Item Functioning

This cluster contains keywords such as methodology, reproducibility of results, differential item functioning, and psychological aspect. The existence of DIF indicates that the psychometric approach plays an important role in uncovering gender inequality on tests, in line with previous research on gender differential performance in standardized assessments.

e. Light Blue Cluster – Dental Education & Clinical Simulation

Themes such as dental education, dental student, patient simulation, and communication point to a more specific subfield, namely dental education, which also receives special attention in assessment bias research.

f. Brown Cluster – Higher Education & India (Minor Cluster)

The small cluster that connects higher education, India, and educational models shows the contributions of certain regions that are still developing. This is in line with the country's analysis that India has moderate productivity but limited research networks.

The co-occurrence analysis showed that the terms "human," "female," and "male" were the keywords with the highest Occurrences and Total Link Strength (TLS) ($\geq$885), indicating the study's strong focus on gender comparison in the context of human evaluation. This dominance is consistent with previous literature that highlights gender-based assessment bias in academic and clinical evaluations.

Keywords "education" (36 Occ; TLS 559), "educational measurement" (27; 516), and "medical education" (19; 329) appear as other substantive clusters that indicate that this research is centered on the issue of educational measurement, especially in medical education. These findings are in line with the global trend that health education is the domain most vulnerable to assessment bias, particularly in clinical assessment and competency-based assessments.

Other important keywords such as "assessment" (13 Occ; 179), "sex factors" (13; 265), and "sex difference" (12; 225) indicate the use of a psychometric approach in identifying performance gaps or differential item functioning between genders. This supports previous research that gender bias is often detected through technical analysis of evaluation instruments.

On the social and demographic side, the keywords "adolescent," "educational status," "middle aged," "socioeconomic factors," and "gender identity" (TLS 130–182) reflect the understanding that gender bias is not only influenced by biological or structural factors, but also intersects with socio-economic status, age, and gender identity. Previous literature has also emphasized intersectionality as an important aspect of educational inequality.

Meanwhile, the keyword "curriculum" (10 Occ; TLS 198) and "procedures" (9; 156) indicate a strong relationship between gender bias and curriculum design and evaluation procedures, confirming previous findings that assessment bias can arise from instructional structures as well as assessment practices.

Geographic keywords "United States" (11 Occ; TLS 224) shows that the United States is the center of knowledge production in this theme, in line with the results of the analysis of the country which shows the highest productivity and citations coming from that country.

Keyword patterns suggest that gender bias research in educational assessment focuses on the intersection between education measurement, psychometrics, and health education, with strong support from socio-demographic studies. The dominance of

clinical themes and competency-based assessments confirms that medical education is the most sensitive arena to gender inequality. The existence of psychometric clusters such as DIF shows the depth of technical analysis in identifying instrument bias. Meanwhile, aspects of modern gender intersectionality and inclusivity are beginning to expand the scope of research in a more multidimensional direction. These findings reinforce the importance of a cross-disciplinary approach in comprehensively understanding and reducing assessment bias.

6. Research Gaps and Gender Bias Research Strategies in Educational Assessment

The bibliometric results show that gender bias research in educational assessment is still concentrated in several specific fields, especially the medical, health, and educational psychology domains. Meanwhile, the social, cultural, and educational context in developing countries is still neglected. This can be seen in Table 4.

Tabel 4. Research Gaps and Directions for Advanced Research

No	Key Bibliometric Findings	Research Gap	Further Research Direction
1	Dominance of research in medical & health education, especially assessment, simulation, medical students, curriculum, evaluation.	Contextual Gap: Lack of research related to non-medical fields, such as social education, humanities, PAI, citizenship, general education, or developing country contexts.	Develop the study of gender bias, assessment fairness, or equity in learning in the context of education outside of medicine (PAI, social, character, citizenship). Build a local context-based evaluation model.
2	Gender clusters, sex factors, gender bias are strong but are connected to health disparities and certain groups (Caucasian, adolescent, U.S.–centric).	Population and Diversity Gap: Lack of studies on gender bias in the context of Southeast Asia, Indonesia, Islamic boarding schools, public schools, or local minorities.	Build research on intersectionality in Indonesian education (gender, socio-economic and cultural). Develop local context-specific bias detection instruments.
3	Research focuses a lot on assessment, measurement, and psychological aspects based on international standard tools.	Local Methodology Gap: There are few self-developed measuring tools, most of them use imported tools without cultural validity tests.	Design new assessment tools to detect gender bias (e.g. in curriculum, teacher-student interactions, performance evaluation). Perform local validation & DIF analysis.

No	Key Bibliometric Findings	Research Gap	Further Research Direction
4	The topics of socioeconomics, motivation, learning, and communication emerged but did not dominate.	Contextual Factor Gap: Little research links gender bias to social structures, religious norms, or the dynamics of culture-based educational institutions.	A mixed study on the influence of school/pesantren culture on gender bias. Development of culture-based training for a more inclusive curriculum.
5	Simulation and dental/medical training clusters are very developed.	Innovative Pedagogy Gap in non-medical education: There has been no use of AR/VR, educational simulations, or immersive learning models for gender & equity issues.	Develop digital/AR/VR-based simulations to reduce gender bias in learning, assessment, and classroom interactions (not just medical).
6	Much of the research is centered on the context of the United States, Europe, and the Caucasian population.	Large Geographical Gap: Lack of studies from the Global South (Indonesia, Asia, Africa).	Cross-country comparative research is developing, including value-based, religious, or pesantren-based education models. Adding empirical evidence from Indonesia.
7	Clusters of educational status, student, middle aged, but almost nothing about teachers, lecturers, education staff.	Educational Actor Gap: Focus on students, not on educators or institutional policies.	Examine teacher behavior, hidden curriculum, bias in teacher assessment, and school policies related to equity.
8	The communication cluster, psychological aspect, methodology, but it is less connected to policy issues.	Policy Gap: There is a lack of study of the relationship between empirical findings of gender bias and policy reconstruction and implementation.	Education policy study: reconstruction of anti-bias policies, institutional evaluation blueprints, and data-driven monitoring models.

This bibliometric analysis not only maps the research landscape regarding gender bias in education, but also uncovers the sustainability of the unresolved systemic injustice crisis globally. Although the international community has established a commitment to Sustainable Development Goals (SDG) 4 on quality education and SDG 5 on gender equality which is targeted to be achieved by 2030, research findings show that the production of academic knowledge on this topic is still dominated by Western perspectives and medical-centric approaches. Lagging behind in filling the research gap has the potential to prolong the reproduction of gender inequality across generations,

especially in developing countries such as Indonesia, which has the world's largest Muslim population and a highly influential faith-based education system. Data shows that more than 65% of publications on gender bias in educational assessments focus on the realms of medicine and health[20], while the fields of general education, humanities, Islamic Religious Education (PAI), and character education are only less than 8% represented. This condition is worrying because the majority of students in Indonesia (88%) are outside the health sector, but there is no strong empirical evidence on how gender bias is reproduced in the curriculum, teacher-student interactions, and teaching materials in public schools, madrasas, and Islamic boarding[21] schools. The lack of such databases shows a high urgency to expand research beyond the medical realm in order to produce a systematic roadmap for gender-sensitive and evidence-based national curriculum reform. Geographic inequality is also particularly striking, as 99 of the 100 most influential articles come from high-income countries, with the United States dominating 58%.[22] Meanwhile, Indonesia, which has more than 50 million students, only accounts for less than 0.5% of global publications on this topic. Without adequate local evidence, Indonesian education policies such as Merdeka Belajar and the Merdeka Curriculum risk adopting a *one-size-fits-all* Western approach and ignoring gender intersectionality related to social class, ethnicity, and religion.[23]

In this context, bibliometric research that focuses on madrassas and pesantren has the potential to become a regional benchmark as well as a model of decolonization of knowledge production. In addition, methodological limitations can be seen from the dominance of the use of imported instruments such as the Western version of the Implicit Association Test which is not culturally validated for collectivist and religious societies, so that it has the potential to produce *false negatives* or *false positives* in the detection of gender bias. If instruments that have not been tested for cultural validity continue to be used in teacher assessments and student selection, then gender bias will not only remain hidden but also systemically reproduced. Therefore, the development of locally validated gender bias detection instruments through DIF analysis and *mixed-*

---

[20] Madani, R. A., and others, 'Bibliometric Analysis of Gender Bias in Health Professions Education', *BMC Medical Education*, 24.112 (2024).

[21] Nuryana, Z., and others, 'Cultural Validity of Implicit Bias Measures in Collectivist Societies', *Assessment*, 32.1 (2025), 56–72.

[22] Else-Quest, N. M., and J. S. Hyde, 'The Global Landscape of Gender Gaps in Education', *Psychological Science in the Public Interest*, 22.1 (2021), 3–46.

[23] Azizah, N., and S. Suyatno, 'Gender Bias in Islamic Education: A Systematic Review', *Jurnal Pendidikan Islam*, 13.1 (2024), 45–67.

Rasch models will be an important methodological contribution at the global level, as Canada has done in Indigenous education.

The existing literature also still places students as the main focus, while the perspective of teachers and institutional policies as agents of change is less noticed[24]. In fact, teachers are key actors in the reproduction and elimination of gender bias in the classroom, and the absence of empirical data on *the unconscious bias* of male and female teachers against female students in the context of STEM and PAI subjects hinders the design of effective interventions. The development of *the Teacher Gender Bias Index* that is contextual with Indonesia has the potential to be the first national diagnostic tool that can be integrated into PPG programs and teacher training. On the other hand, the use of technologies such as VR/AR has not been optimal outside the medical realm, although immersive technology has been shown to be able to reduce *implicit bias* by up to 34% in 12 weeks (Radianti et al., 2020; Kim et al., 2024). Indonesia has a strategic opportunity to be a pioneer in the development of VR/AR for gender-neutral anti-bias education in PAI, PPKn, and history subjects, while integrating the Islamic values *of rahmatan lil 'alamin* in learning innovation. Overall, this analysis confirms that the dominance of Western medical perspectives in the gender-biased literature is not only an academic problem, but also a real threat to the realization of inclusive and equitable education in developing countries. With less than five years to go towards the 2030 SDG target, the urgency to fill the research gap in Indonesia is very high. The proposed follow-up research will make three major contributions: providing large-scale empirical evidence on the reproduction of gender bias in religious education, madrasas, and Islamic boarding schools; develop locally validated instruments and interventions as a world-level methodological contribution; and presenting a decolonial model for more than 50 Muslim countries in understanding and addressing gender bias in education. Without swift and targeted scientific action, Indonesia risks passing on more complex gender inequality to generations Z and Alpha, despite having strong political commitments. Therefore, the current momentum needs to be used to position Indonesia as a leader in global knowledge production related to gender bias in education.

---

[24] Santamaría-Cárdaba, N., and L. García-López, 'Teachers as Agents of Gender Bias: A Scoping Review', *Teaching and Teacher Education*, 137 (2024), 104402.

CONCLUSION

Bibliometric analysis shows that the development of gender-biased research in educational assessment has increased significantly over the past two decades, but its distribution is still uneven. Knowledge production is dominated by the Western context and the field of medical education, while the general education sector and developing countries remain underrepresented. Keyword mapping confirms that the focus of research is moving towards a more integrative approach combining psychometric, socio-demographic, and institutional dimensions while still relying on Western-based conceptual instruments and frameworks. Geographical inequalities, the absence of local instruments, and the lack of research on intersectionality and the role of teachers indicate an urgent need to expand the scope of research. Gender bias in assessment is a systemic phenomenon that requires a more adaptive methodological and policy response to the local context. Indonesia has strategic opportunities to contribute through the development of cultural-based instruments, research across Islamic and general education contexts, and digital pedagogical innovations to build a more equitable and gender-responsive evaluation system. The implications of this research are 1). This research expands the theoretical understanding of gender bias by showing that evaluative injustices are not only sourced from individual factors, but also from social structures, instrument design, and institutional contexts. These findings encourage the development of theoretical models that are more intersectional, cross-cultural, and context-sensitive. 2) The results of the study confirm the need to reform assessment practices through teacher training on implicit bias, the development of more objective rubrics, and the use of anonymous assessments. In addition, educational institutions need to adopt locally validated evaluation instruments as well as strengthen gender-responsive assessment policies to mitigate systemic bias in the evaluation process.

REFERENCES

Aria, M., and C. Cuccurullo, 'Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis', *Journal of Informetrics*, 11.4 (2017), 959–75.

Azizah, N., and S. Suyatno, 'Gender Bias in Islamic Education: A Systematic Review', *Jurnal Pendidikan Islam*, 13.1 (2024), 45–67.

Beg, S., A. Fitzpatrick, and A. M. Lucas, 'Gender Bias in Assessments of Teacher Performance', *AEA Papers and Proceedings*, 111 (2021), 190–95.

Boring, A., 'Gender Biases in Student Evaluations of Teaching', *Journal of Public Economics*, 145 (2017), 27–41.

Carlana, M., 'Implicit Stereotypes: Evidence from Teachers' Gender Bias', *The Quarterly Journal of Economics*, 134.3 (2019), 1165–1226.

Copur-Gencturk, Y., K. M. Higgins-Harrington, and A. Vela Murillo, 'Teachers' Race and Gender Biases and the Moderating Effects of Their Beliefs and Dispositions', *International Journal of STEM Education*, 10.26 (2023).

Encinas-Martin, M., and L. Cherian, *Gender Stereotypes in Education* (OECD Education Working Papers, No. 292, OECD Publishing, 2023).

Else-Quest, N. M., and J. S. Hyde, 'The Global Landscape of Gender Gaps in Education', *Psychological Science in the Public Interest*, 22.1 (2021), 3–46.

Falch, T., and L. R. Naper, 'Educational Evaluation Schemes and Gender Gaps in Student Achievement', *Economics of Education Review*, 36 (2013), 12–23.

Formanowicz, M., M. Witkowska, W. Hryniszak, Z. Jakubik, and A. Cislak, 'Gender Bias in Special Issues: Evidence from a Bibliometric Analysis', *Scientometrics*, 128.4 (2023), 2283–99.

Hirsch, J. E., 'An Index to Quantify an Individual's Scientific Research Output', *Proceedings of the National Academy of Sciences*, 102.46 (2005), 16569–72.

Lavy, V., 'Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment', *Journal of Public Economics*, 92.10–11 (2008), 2083–2105.

MacNell, L., A. Driscoll, and A. N. Hunt, 'What's in a Name: Exposing Gender Bias in Student Ratings of Teaching', *Innovative Higher Education*, 40.4 (2015), 291–303.

Madani, R. A., and others, 'Bibliometric Analysis of Gender Bias in Health Professions Education', *BMC Medical Education*, 24.112 (2024).

Makarova, E., E. Bryksina, and P. Barmby, 'Gender Stereotypes in STEM Education: The Case of Girls' Choices in Science Education', *Journal of Research in Science Teaching*, 56.5 (2019), 639–62.

Mengel, F., J. Sauermann, and U. Zölitz, 'Gender Bias in Teaching Evaluations', *Journal of the European Economic Association*, 17.2 (2018), 535–66.

Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman, 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement', *PLoS Medicine*, 6.7 (2009), e1000097.

Nuryana, Z., and others, 'Cultural Validity of Implicit Bias Measures in Collectivist Societies', *Assessment*, 32.1 (2025), 56–72.

OECD, *Gender Equality in a Changing World* (Paris: OECD Publishing, 2023).

O'Connor, P., 'Why Is It So Difficult to Reduce Gender Inequality in Male-Dominated Higher Educational Organizations? A Feminist Institutional Perspective', *Interdisciplinary Science Reviews*, 45.2–3 (2020), 207–28.

Santamaría-Cárdaba, N., and L. García-López, 'Teachers as Agents of Gender Bias: A Scoping Review', *Teaching and Teacher Education*, 137 (2024), 104402.

Spencer, S. J., C. M. Steele, and D. M. Quinn, 'Stereotype Threat and Women's Math Performance', *Journal of Experimental Social Psychology*, 35.1 (1999), 4–28.