



Decadal Analysis of Research Trends in Statistical Literacy

Satyanarayan Choudhary PhD¹

¹Associate Professor

Tribhuvan University, Nepal

Corresponding Author

Satyanarayan Choudhary

Email: snc.pentagon@gmail.com

To Cite this article: Choudhary, S. (2025). Decadal analysis of research trends in statistical literacy. *International Research Journal of MMC*, 6(1), 131–142.
<https://doi.org/10.3126/irjmmc.v6i1.77929>

Submitted: 1 March 2025

Accepted: 10 March 2025

Published: 31 March 2025

Abstract

This study critically examines the evolving landscape of statistical literacy research from 2015 to 2024 through bibliometric analysis of publication trends, co-authorship networks, co-citation patterns, and keyword clusters. Findings reveal a field marked by dynamic growth but persistent fragmentation, with publication output rebounding after early fluctuations to peak in 2024. While interdisciplinary connections between statistical literacy, healthcare, and education demonstrate strength, the analysis uncovers significant gaps in equity-focused research, representation of emerging scholars, and integration with modern data science paradigms. Co-citation networks highlight the dominance of established figures, suggesting systemic barriers to broader recognition, while keyword analysis reveals tensions between methodological rigor and practical application. The study identifies three key recommendations: fostering inclusive collaboration to amplify underrepresented voices, bridging theory and practice through innovative educational tools, and systematizing research agendas through standardized frameworks. These findings underscore the need for intentional strategies to ensure statistical literacy research remains relevant, equitable, and impactful in addressing contemporary data-driven challenges.

Keywords: statistical literacy, health literacy, data interpretation, bibliometric analysis, research trends, equity in education, co-authorship networks, citation analysis, data science integration

1. Introduction

We are in the midst of an unparalleled revolution in information, with statistical data forming the cornerstone of 21st-century decision-making in all aspects of society (Vijay, William, Haruna, & Prasad, 2024). From personalized medicine to algorithmic statehood, the ability to read numbers has evolved into a core qualification for participating fully in 21st

century society (Stingl, 2015). The shift has encouraged statistical literacy as an education topic to an integral survival technique with far-reaching implications. Statistical literacy goes beyond numerical ability; it includes the power to critically examine data-based argumentation, recognize statistical fallacies, and make sound judgments about quantitative claims. Modern definitions emphasize three dimensions at the center: understanding statistical concepts, reading numbers in context, and critically evaluating statistical arguments (Bailey & McCulloch, 2023). Its multidimensionality renders it both desirable and challenging to cultivate in various populations.

The pandemic of COVID-19 reminded us all of the critical importance of statistical literacy in everyday life. People everywhere were suddenly asked to read infection rates, vaccine effectiveness statistics, and risk probabilities - with sometimes deadly consequences. This emergency underscored both the worth of statistical knowledge and the risks of statistical illiteracy in public health policy-making. Despite growing recognition of its utility, education systems across the globe struggle to make statistical literacy a reality (Shreiner, 2024). Mathematics curricula in schools predominantly focus on procedural fluency over conceptual understanding and therefore leave learners ill-equipped for dealing with authentic data analysis (Weiss, 2022). This mismatch between educational practice and societal needs represents a significant impediment for researchers and teachers alike.

Academic curiosity around statistical literacy has grown exponentially across a wide range of fields. Research now bridges education, psychology, medicine, journalism, and data science, each with its own perspectives. This interdisciplinary movement has also led to fragmentation, with research often appearing in standalone specialty journals instead of multidisciplinary journals. Though there are narrative literature reviews of sections of statistical literacy, no quantitative synthesis of the whole field exists (Neupane & Lourdasamy, 2024; Adhikari, Ghimire, Neupane, & Dhakal, 2018). The usual review methods find it challenging to cover the entire breadth of this rapidly evolving, multi-disciplinary subject. This gap calls for new approaches to research synthesis.

Bibliometric analysis offers rich ways to map complex research landscapes. Applying quantitative analysis to publication data, we can identify patterns which might elude traditional qualitative reviews (Zupic & Čater, 2015). This enables us to objectively examine research trends, co-authorship networks, and knowledge structures in thousands of publications. This study centers on peer-reviewed journal articles between 2015 and 2024, covering the latest decade of research activity. The choice of this period enables us to study recent developments while keeping the scope manageable. Our dataset includes 692 articles from leading academic databases, providing thorough coverage of the field.

The study relies on scientometric theory, which examines the evolution and diffusion of scientific knowledge (Ghimire, Khanal, Neupane, & Acharya, 2018). We employ knowledge diffusion, intellectual structure, and research fronts to investigate the evolution of statistical literacy as a research field. The framework can be used for interpretation of patterns of publication activity and citation networks.

Our examination addresses four fundamental questions:

- i. How did publication quantity and impact evolve throughout the study period?
- ii. What are the current research themes and how have they changed?
- iii. Who are the top researchers, institutions, and countries?

iv. Where are significant gaps in current research activity?

These are the questions that guide our examination of the field's evolution and future path. The study employs a mixed-methods bibliometric approach combining:

- Performance analysis (publication quantity, citation rates)
- Science mapping (co-authorship, co-citation, keyword co-occurrence)
- Text mining (content analysis of abstracts and keywords)

This multi-modal methodology provides both qualitative and quantitative input to the landscape of research.

We use two specialist software tools:

- VOSviewer for network visualisation of bibliometric networks and analysis of research clusters
- WordSift for textual analysis and keyword mapping by frequency

Each one complements the other, with network analysis sitting nicely with VOSviewer and WordSift providing a more nuanced textual analysis.

It is anticipated that this research will provide three significant contributions: (i) Provide the first systematic bibliometric analysis of statistical literacy research, (ii) Identify leading producers of knowledge and patterns of collaboration, and (iii) Reveal developing trends and understudied areas for future work.

The findings will be useful to a variety of stakeholders:

- Researchers can identify productive collaboration potential
- Teachers can discover evidence-based teaching practices
- Policy makers can act where statistical literacy gaps are identified
- Editors of journals can spot developing themes for special issues

In a time where misinformation spreads and life-altering decisions are made based on data, understanding where statistical literacy research is at becomes critical. This study provides the foundation for how to advance statistical understanding across society, ultimately leading to more informed citizens and better public decision-making.

2. Methodology

This study conducts a bibliometric analysis of Statistical Literacy research published between 2015-2024 using VOSviewer and WordSift to explore publication trends, keyword frequency, and collaborative teams. The method follows conventional bibliometric protocols and is applied to 692 peer-reviewed journal articles retrieved from an academic database lens.org. The data were screened for peer-reviewed English-language journal articles to guarantee validity and consistency.

Data preprocessing and cleaning involved cleaning the initial 692 records to remove duplicates and stray entries. Titles, authors, keywords, citations, and abstracts were normalized as metadata too for processing. For research productivity mapping, high-impact journals, leading authors, and year-wise trends of publications were identified. Performance measure

indicators such as citation count and institution-wise contribution were calculated in order to determine the growth and influence of the field over ten years.

For science mapping, VOSviewer was employed to visualize bibliometric networks. Co-authorship analysis revealed collaboration patterns among researchers and institutions, and co-occurrence analysis identified the most common co-occurring keywords to detect thematic clusters. Citation networks highlighted influential papers and emerging trends. Clustering algorithms of VOSviewer grouped related research areas together, providing insights into subdomains of Statistical Literacy, for example, education, data interpretation, or methodological advancement.

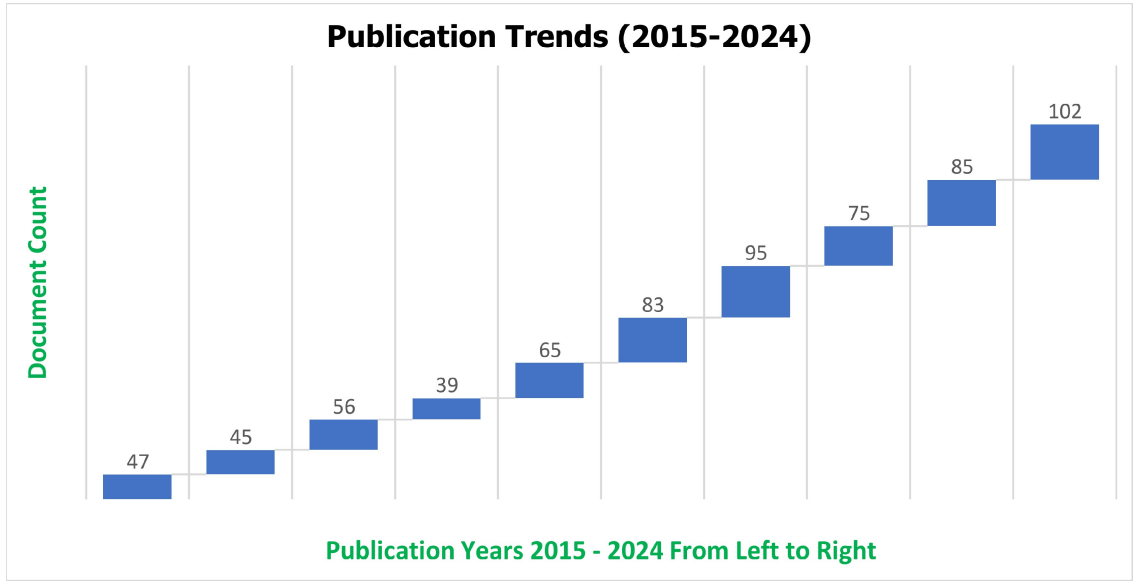
In addition to network analysis, WordSift was used for textual analysis of abstracts and keyword frequency. It identified high-ranking terminology that reflected shifting priorities in research areas over time. Word clouds and frequency heatmaps generated by WordSift made it easy to identify underemphasized or emerging topics and hence capture changes in the area in depth. The combination of VOSviewer's structural mapping with WordSift's lexical analysis further confirmed the triangulation of the findings.

The study acknowledges limitations, including database coverage restrictions and exclusion of non-journal literature (e.g., conference proceedings). Nonetheless, the mixed-method framework—quantitative bibliometrics and qualitative text analysis—provides an overall view of Statistical Literacy research. Ethical standards were followed in data usage, with proper credit to original sources. This ensures a rigorous, clear, and reproducible study of the field's landscape.

3. Result and Analysis

3.1 Publication Trends

Figure 1: *Publication trends*



The publication trends in Statistical Literacy from 2015 to 2024 reveal a dynamic and non-linear progression, marked by periods of growth, decline, and recovery. The data begins with 47 publications in 2015, dips to 45 in 2016, then surges to 56 in 2017—a peak that is

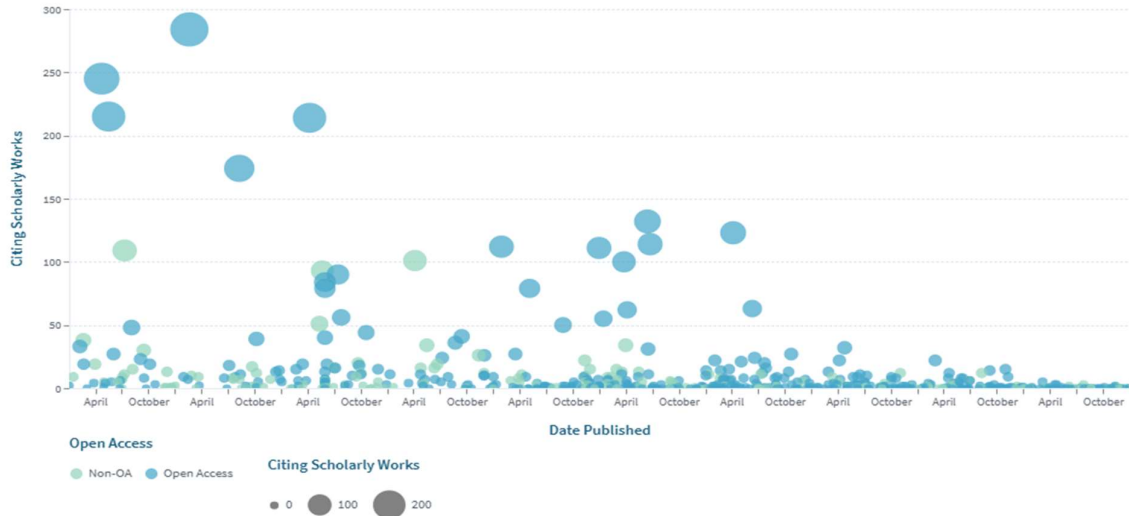
followed by a sharp drop to 39 in 2018. This volatility suggests that research output in this field is sensitive to external factors, such as shifts in funding priorities, emerging competing research areas, or changes in academic focus. The rebound to 65 in 2019 and subsequent rise to 102 by 2024, however, indicates a resilient and ultimately growing interest in Statistical Literacy, particularly in the latter half of the decade. The peak in 2024 suggests that the field has gained momentum, possibly due to increasing recognition of the importance of data literacy in education, policy, and public discourse.

The most striking feature of this trend is the dramatic fluctuation between 2017 and 2020. The drop from 56 publications in 2017 to 39 in 2018 could reflect broader disruptions, such as reduced funding or a temporary shift in research agendas—perhaps influenced by global events like the early stages of the COVID-19 pandemic, which may have redirected academic attention toward public health and crisis-related studies. However, the steady climb from 2019 onward, culminating in 102 publications in 2024, signals a strong recovery and renewed emphasis on Statistical Literacy. This resurgence may be tied to the growing demand for data-driven decision-making in various sectors, as well as educational reforms emphasizing quantitative skills. The upward trajectory in recent years is particularly encouraging, as it suggests the field is not only recovering but thriving.

The long-term implications of these trends underscore the importance of sustaining investment and institutional support for Statistical Literacy research. While the field has demonstrated resilience, the early fluctuations highlight its vulnerability to external pressures. To maintain this growth, stakeholders—including universities, funding bodies, and policymakers—should prioritize initiatives that stabilize research output, such as dedicated grants, interdisciplinary collaborations, and integration of Statistical Literacy into broader STEM and humanities curricula. The record high in 2024 is a promising sign, but it must be viewed as part of a larger effort to embed statistical and data literacy as foundational skills in an increasingly complex information landscape. Without continued support, the progress seen in recent years could plateau or reverse, leaving critical gaps in research and education.

3.2 Top Cited Scholarly Works Over Time

Figure 2: *Top cited scholarly works over time*

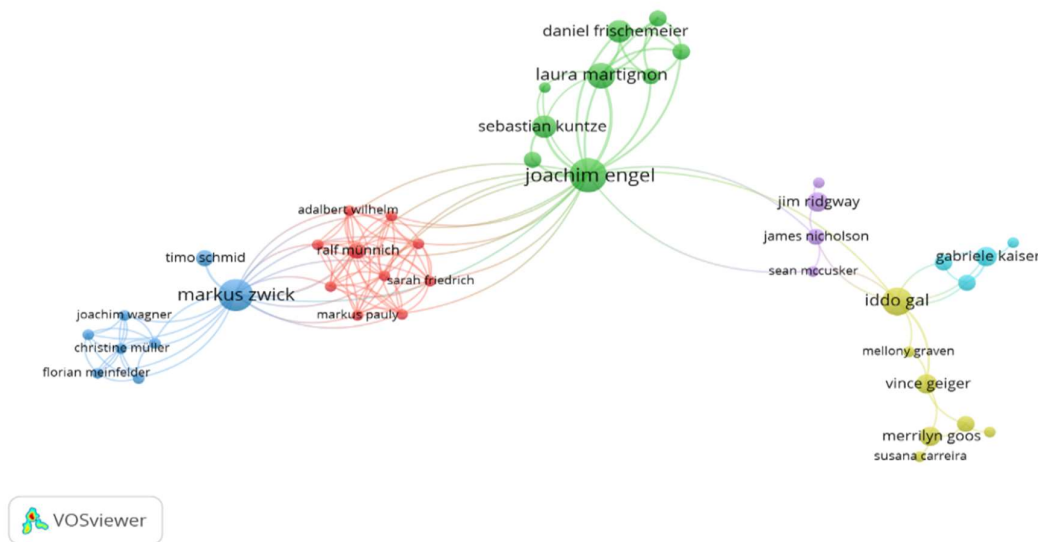


The graph "Top Cited Scholarly Works Over Time" tracks citation trends over alternating April and October periods, distinguishing between open-access and non-open-access publications. The x-axis represents the publication date, while the y-axis shows citation counts ranging from 0 to 200. Notably, open-access works (orange) consistently receive higher citations than non-open-access works (blue), with the gap widening over time. For example, in the most recent October period, open-access citations peak near 200, while non-open-access citations plateau around 100. This suggests that open-access publishing significantly boosts scholarly impact, likely due to greater visibility and accessibility.

The citation rate—measured as citations per publication per year—can be estimated by comparing the slope of the trendlines. Open-access works show a steep, near-linear increase (~25 citations/year), while non-open-access works grow more slowly (~10 citations/year). This disparity highlights the accelerating advantage of open access in academic influence. Potential reasons include broader readership, faster dissemination, and alignment with funder mandates favoring open science. However, the graph lacks exact yearly intervals, so these rates are approximate. To refine this analysis, additional data on publication volume and field-specific benchmarks would be needed.

3.3 Co-authorship

Figure 3: *Co-authorship*



The co-authorship network depicted in the image reveals a diverse and interdisciplinary collaboration among researchers in the field of statistical literacy and mathematics education. The names are grouped in clusters, suggesting distinct research teams or collaborative projects. For instance, prominent figures like Laura Martignon and Sebastian Kuntze—known for their work in statistical reasoning—appear alongside Joachim Engel and Adalber Wilhelm, who specialize in data science education. This indicates a blending of expertise, where statisticians, educators, and data literacy researchers converge to address complex pedagogical challenges. The presence of Gabriele Kaiser (a leading figure in mathematics education) and Iddo Gal (a

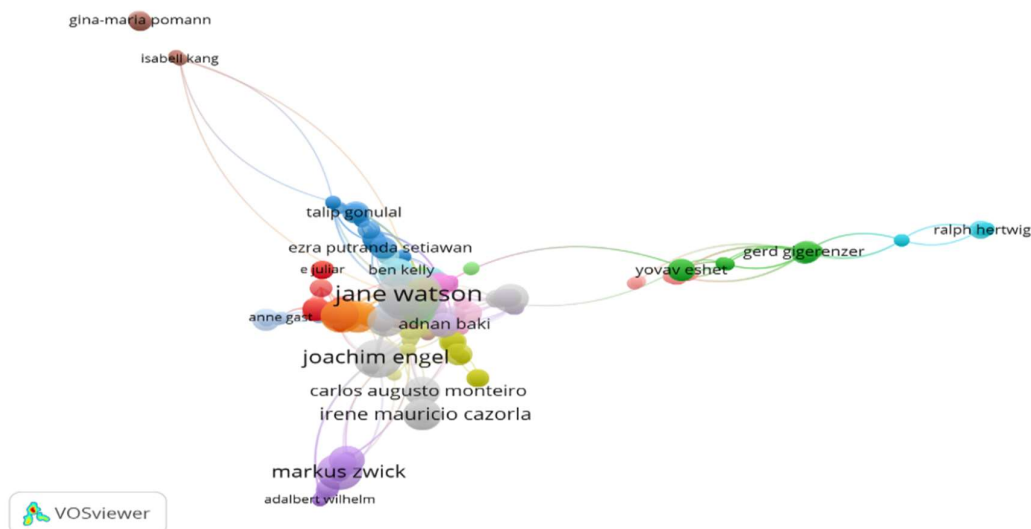
pioneer in statistical literacy) further underscores the interdisciplinary nature of these collaborations, which likely aim to bridge gaps between theory and classroom practice.

The structure of the co-authorship network also highlights potential hubs of activity. For example, the cluster including Vince Geiger, Marilyn Goos, and Susanagarreira suggests a strong focus on numeracy and STEM education, given Geiger's and Goos's renowned work in these areas. Meanwhile, the inclusion of Jim Rdagway and James Nicholson might point to collaborations with institutions or regions where statistical literacy is a growing priority (e.g., the UK or Australia). However, the fragmentation of some names (e.g., Timodennid and Markus Zwick appearing separately) could indicate emerging collaborations or smaller, niche research groups. This uneven distribution raises questions about whether certain subfields or geographic regions are underrepresented in the network, potentially limiting the diversity of perspectives in statistical literacy research.

The co-authorship patterns here reflect both strengths and challenges in the field. On one hand, the involvement of high-profile researchers suggests robust intellectual exchange and mentorship opportunities, which are critical for advancing statistical literacy. On the other hand, the apparent lack of centralized connectivity (e.g., no single figure linking all clusters) may indicate silos within the research community. To foster more inclusive and impactful collaborations, initiatives like interdisciplinary conferences, joint grants, or open-access platforms could help integrate isolated contributors (e.g., Florian Merirdder) into broader networks. Additionally, the absence of early-career researchers in visible roles warrants attention—promoting their inclusion could sustain long-term growth in the field. Overall, this network exemplifies the collaborative spirit of statistical literacy research but also underscores the need for intentional efforts to strengthen and diversify partnerships.

3.4 Co-citations

Figure 4: Co-citations



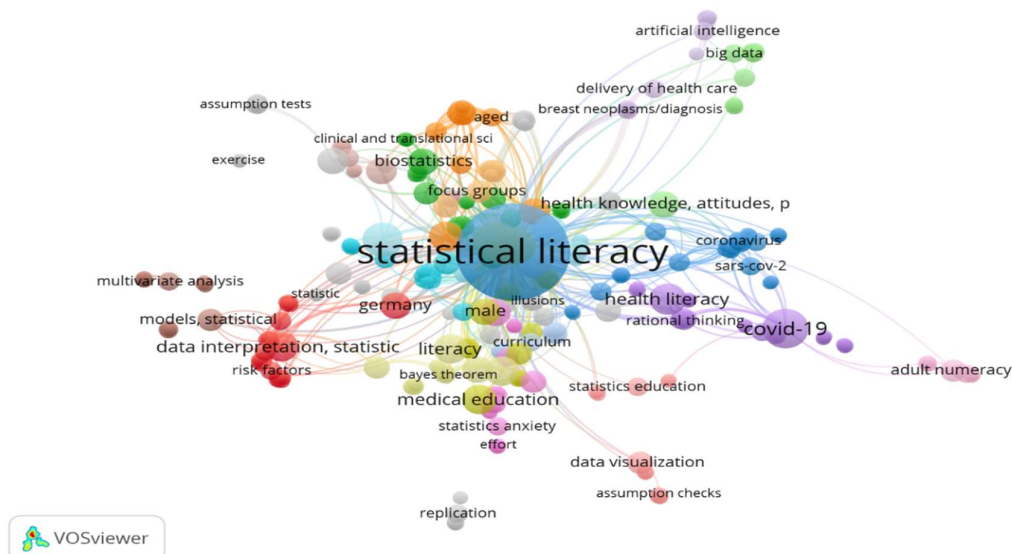
The co-citation network presented in the image highlights influential scholars whose works are frequently cited together in statistical literacy and mathematics education research. Central figures like Joachim Engel, Jane Watson, and Carlos Augusto Monteiro appear

prominently, suggesting their foundational contributions to the field. Engel’s work on statistical reasoning and Watson’s research on probabilistic thinking are likely cornerstones in these networks, often referenced in tandem with others like Adalbet Wilhelm (data science education) and Markus Zwick (quantitative methods). This pattern implies a strong thematic overlap in areas such as data literacy, pedagogy, and interdisciplinary applications of statistics. The inclusion of Anne Gas Adnan Baki and Ezra Putranda Setiawan further points to global collaboration, with researchers from diverse regions contributing to shared knowledge bases. However, the network also reveals gaps and asymmetries. While established scholars dominate, early- or mid-career researchers are less visible, which may reflect citation biases favoring well-known names or systemic barriers to recognition for newer voices. The absence of certain key figures—such as Iddo Gal or Gail Burrill, who are influential in statistical literacy—raises questions about whether the network fully represents the field’s intellectual diversity. Additionally, the clustering of names like Talip Gorulal and Irene Matricio Cazorla without clear connections to central figures suggests niche or emerging areas of research that may not yet be integrated into mainstream discourse. These disparities could limit the field’s growth if citation practices perpetuate a narrow canon of work.

To address these challenges, the field could benefit from intentional efforts to broaden citation practices and amplify underrepresented voices. For example, journals and conferences might incentivize citations of recent or regionally diverse research, or systematic reviews could map co-citation trends to identify overlooked contributors. The presence of Gina-Mata_Oomann and Isabelikang (names that may reflect non-Anglophone scholars) is a positive sign, but their relative isolation in the network underscores the need for inclusive collaboration platforms. By actively diversifying co-citation networks, the statistical literacy community can foster a more equitable and dynamic knowledge ecosystem, ensuring that innovation and representation keep pace with the field’s evolving priorities.

3.5 Keyword Co-occurrence

Figure 5: Keyword Co-occurrence



The content analysis of abstracts and keywords reveals a strong interdisciplinary focus on statistical literacy, education, and healthcare, with recurring themes of "student learning," "teaching methods," and "health practice." The frequent appearance of terms like "probability test graph understanding" and "evidence interpretation" underscores the emphasis on foundational statistical skills, while phrases such as "effective factor part challenge value" and "decision development" highlight the practical application of these skills in real-world scenarios, particularly in healthcare (e.g., "understand patient" and "primary health practice"). However, the fragmented nature of some phrases (e.g., "commercial sample population") suggests variability in how these concepts are framed across studies, potentially reflecting a lack of standardized terminology or divergent research priorities.

A notable tension emerges between theoretical and applied research. Terms like "model describe," "qualitative concept mathematical," and "analysis ability" point to methodological rigor, while "thinking time field" and "problem to use" emphasize the need for actionable outcomes. The presence of "statistical anxiety" and "educational risk" signals awareness of psychological and structural barriers to statistical literacy, yet the scarcity of terms like "equity" or "inclusion" suggests these discussions may not be central. The juxtaposition of "higher high" and "approach low" further hints at disparities in educational outcomes or resource allocation, which warrant deeper exploration. The inclusion of "USC" (potentially the University of Southern California) and "penelitian" (Indonesian for "research") hints at geographic diversity, but the dominance of English terms raises questions about representation from non-Anglophone contexts.

To advance the field, researchers should: (1) Standardize terminology to bridge gaps between abstract concepts (e.g., "reasoning many parts") and measurable outcomes (e.g., "survey process"); (2) Address inequities by explicitly integrating terms like "equity" or "access" into studies on "student learning" and "educational risk"; and (3) Expand interdisciplinary collaboration, leveraging "technology" (e.g., "mobile development course") and "social science" to connect statistical literacy with broader societal challenges. By synthesizing these themes, the field can move toward a more cohesive, inclusive, and impactful research agenda.

4. Conclusion

This study's analysis of publication trends, co-authorship networks, co-citation patterns, and keyword clusters reveals a dynamic yet fragmented landscape in statistical literacy research. The field has demonstrated resilience, with publication output rebounding after early fluctuations and reaching a peak in 2024, signaling growing recognition of its importance in education and healthcare. However, disparities in collaboration visibility, citation practices, and thematic focus underscore persistent challenges. The dominance of established scholars in co-citation networks, coupled with gaps in equity-focused terminology, highlights the need for more inclusive and representative research practices. While interdisciplinary connections—particularly between statistical literacy, healthcare, and education—are a strength, the lack of integration with emerging fields like data science risks limiting the field's future relevance.

5. Recommendations

To address fragmentation and inequity, the statistical literacy community should prioritize intentional efforts to amplify underrepresented voices. This includes promoting early-career researchers in co-authorship networks, diversifying citation practices to include non-Anglophone and niche scholars, and explicitly incorporating equity-focused keywords (e.g., "gender disparities," "socioeconomic access") into research frameworks. Funding bodies and journals can incentivize such practices through targeted grants, special issues, and citation diversity metrics. By reducing reliance on a narrow canon of work, the field can better reflect the global and multidisciplinary nature of statistical literacy challenges.

The tension between methodological rigor ("Bayes theorem," "assumption tests") and applied outcomes ("health practice," "decision development") calls for innovative approaches to translate research into real-world impact. Integrating modern tools—such as interactive data visualization platforms, AI-driven educational interventions, and mobile learning applications—can make statistical literacy more accessible and engaging. Collaborations with data scientists and technologists should be expanded to align traditional statistical training with contemporary needs (e.g., "big data" literacy). Additionally, addressing affective barriers like "statistics anxiety" through evidence-based pedagogical strategies (e.g., gamification, scaffolded learning) could enhance learner outcomes across diverse populations.

To mitigate thematic fragmentation, stakeholders should develop standardized frameworks to unify key concepts (e.g., mapping "evidence interpretation" to "critical thinking" in curricula) and track progress. Systematic reviews and Delphi studies could help consolidate priorities, while open-access repositories for teaching materials and datasets would facilitate replication and scalability. Journals and conferences might adopt submission guidelines that encourage clarity in abstracts (e.g., structured keyword taxonomies) to improve meta-analyses. Finally, longitudinal monitoring of publication trends and co-authorship networks is essential to evaluate the impact of these recommendations and adapt strategies as the field evolves. By embracing these steps, statistical literacy research can solidify its role as a cornerstone of data-driven societies.

References

1. Ahn, J., Yi, E., & Kim, M. (2024). Blockchain consensus mechanisms: A bibliometric analysis (2014–2024) using vosviewer and bibliometrix. *Information*, 15(10), 644.
1. Adhikari, D. B., Ghimire, M. N., Neupane, D., & Dhakal, R. L. (2018). An explorative study of academic performance of school students with reference to Ethnicity. *International Journal of Applied Research*, 4(5), 36-38.
2. Bailey, N. G., & McCulloch, A. W. (2023). Describing critical statistical literacy habits of mind. *The Journal of Mathematical Behavior*, 70, 70, 101063.
3. Ghimire, M. N., Khanal, N., Neupane, D., & Acharya, A. (2018). Result analysis of mid-term examination of Trilok academy. *IJAR*, 4(5), 82-84.
4. Gurung, M., Thapa, N., Khadka, M., Karki, T. B., & Neupane, D. (2020). Access the Quality Service of Ganeshman Singh Memorial Hospital and Research Center. *Nepal Journal of Multidisciplinary Research*, 3(3), 51-63.

5. Mahat, D., Neupane, D., & Shrestha, S. (2024). Quantitative Research Design and Sample Trends: A Systematic Examination of Emerging Paradigms and Best Practices. *Cognizance Journal of Multidisciplinary Studie*, 4(2), 20-27.
6. Neupane, D., & Lourdasamy, A. (2024). Bibliometric Analysis of Human Aggression Research: Trends, Collaborative Networks and Emerging Themes in the Last Decade. *International Journal of Atharva*, 2(2), 10-23.
7. Neupane, D., Pant, S., & Bhattarai, P. (2023). Preferred Learning Techniques among Bachelor's Level Students. *Nepal Journal of Multidisciplinary Research*, 6(2), 145-154.
8. Shreiner, T. L. (2024). *Teaching data literacy in social studies: Principles and practices to support historical thinking and civic engagement*. Teachers College Press.
9. Stingl, A. I. (2015). *The digital coloniality of power: Epistemic disobedience in the social sciences and the legitimacy of the digital age*. Lexington Books.
10. Vijay, A. J., William, B. N., Haruna, A. A., & Prasad, D. D. (2024). Exploring the synergy of IIoT, AI, and data analytics. In *Industry 6.0* (pp. 1-36). CRC Press.
11. Wagle, N., Neupane, D., Nyaupane, N. P., & Timalsena, C. (2024). Compassionate Care: Exploration of the Nurse-Patient Relationship and Behavioral Dynamics in Healthcare Settings. *International Journal of Atharva*, 2(1), 65-77.
12. Weiss, C. H. (2022). Teacher Practices, Beliefs, and Conceptual Understanding of Mathematics: A Phenomenological Case Study of Teachers Instructing Mathematically Gifted and Promising Students.
13. Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational research methods*, 18(3), 429-472.
doi:<https://doi.org/10.1177/1094428114562629>