

Bibliometric and Network Analysis of Arabic Language Resources and Digital Tools

Mohamed Uwais Mohamed Ameen

Islamic University of Maldives. Kulliyyah of Education.
Male, Maldives.

Abstract

Purpose : This study offers a bibliometric and network analysis of global research on Arabic language resources and digital tools, identifying key publication trends, collaboration patterns, and thematic structures. It addresses the gap in systematic mapping by providing an overview of the field's growth and conceptual development from 2000 to 2024.

Design/methodology/approach: A bibliometric and science mapping approach was applied using data retrieved from the Scopus database. On 23 August 2025, a comprehensive search identified studies on Arabic language resources and digital tools published between 2000 and 2024. The search produced 3,405 records, of which 674 were English-language, peer-reviewed journal articles that met the inclusion criteria. Data were analyzed using descriptive bibliometric indicators in Microsoft Excel and network visualizations in VOSviewer to map co-authorship, co-citation, and keyword relationships.

Findings : Results reveal steady growth in Arabic language research since 2010, with sharp expansion after 2015 alongside advances in digital humanities and natural language processing. The United States, the United Kingdom, Saudi Arabia, and Egypt were the leading contributors, forming strong transnational collaboration networks. Nine country clusters were identified, reflecting diverse regional partnerships across Europe, the Commonwealth, and emerging South–South alliances in Asia and Africa. The intellectual structure displayed seven author clusters, whereas six conceptual clusters underscored the amalgamation of pedagogy, technology, and evaluation themes.

Originality/value : Arabic language resource research is a growing global field, and this study maps its knowledge structure, offering clear insights and directions for future research.

Keywords: bibliometric analysis, Arabic language resources, digital tools, corpus linguistics, science mapping

1 Introduction And Background Literature

The increasing digitization of language and communication has revolutionized linguistic research, paving the way for computational tools and digital resources that enhance the study, preservation, and dissemination of world languages. Among these, Arabic, spoken by more than 420 million people across twenty-five nations, occupies a central position, both for its historical significance and for its linguistic complexity (Filiu, 2011). Despite its global relevance, Arabic language resources have historically lagged behind English and other European languages in terms of digitization, standardization, and accessibility (Awal, 2025).

Over the past two decades, global interest in Arabic language technologies has grown substantially. The advent of natural language processing (NLP), corpus linguistics, and digital humanities has shifted Arabic studies from a primarily theoretical field to one increasingly characterized by data-driven and technology-enhanced approaches. Researchers now employ computational tools for corpus annotation, automatic speech recognition, sentiment analysis, and machine translation, expanding the utility of Arabic in digital contexts.

Bibliometric analysis has emerged as a powerful method to trace these scholarly evolutions (Donthu et al., 2021). Through quantitative mapping of publication metadata, co-authorship, and thematic keywords, it

enables the identification of influential authors, collaboration networks, and emerging research fronts (Wallin, 2005). Addressing this gap is crucial, as understanding the trajectory of Arabic digital scholarship provides insights into how linguistic technologies develop within multilingual and multicultural ecosystems.

This study therefore aims to systematically examine the global research landscape on Arabic language resources and digital tools between 2000 and 2024. Specifically, it seeks to (i) identify publication and citation trends; (ii) examine collaboration networks among countries and institutions; (iii) analyze the geographical distribution of contributions; (iv) reveal the intellectual structure through author co-citation patterns; and (v) explore the conceptual structure via keyword co-occurrence mapping. By integrating bibliometric indicators with network visualizations, this research contributes a comprehensive understanding of how Arabic language technology has evolved, diversified, and globalized over the past twenty-five years.

2. Methods

Bibliometric analysis has become one of the most systematic and widely adopted methods for examining extensive collections of scientific literature. It offers a quantitative lens through which researchers can trace the developmental trajectories of a discipline while uncovering emerging themes, intellectual structures, and research frontiers (Donthu et al., 2021). Its growing prominence is closely associated with the advancement of analytical tools such as VOSviewer, Leximancer, and Gephi, together with the continuous enhancement of scientific databases including Web of Science, Scopus, and PubMed (Fauzi, 2024). Although bibliometric studies have gained renewed attention in the era of digital scholarship, their conceptual roots can be traced back to the 1950s, when systematic methods for measuring scholarly communication began to emerge (Wallin, 2005). This longstanding methodological tradition, revitalized by modern visualization technologies and expansive indexing platforms, positions bibliometric analysis as a cornerstone for understanding how knowledge domains evolve across time and geography (Fauzi, 2024).

2.1 Searching and Retrieving Data

On 23 August 2025, a comprehensive search of the *Scopus* database was conducted to identify academic contributions addressing Arabic language resources and digital linguistic tools published between 2000 and 2024. The search string applied was as follows:

TITLE-ABS-KEY (“Arabic language resources” OR “Arabic digital tools” OR “Arabic corpus” OR “language technology”) AND PUBYEAR > 1999 AND PUBYEAR < 2025 AND (LIMIT-TO(DOCTYPE, “ar”)) AND (LIMIT-TO(SRCTYPE, “j”)) AND (LIMIT-TO(LANGUAGE, “English”))

This query generated an initial dataset of 3,405 records. The inclusion criteria specified English-language, peer-reviewed journal articles with clear focus on Arabic language resource development or digital applications. Exclusion criteria involved conference proceedings, non-linguistic studies, or publications with incomplete bibliographic metadata. After removing duplicates ($n = 2$) and irrelevant articles, a total of 674 records were retained for bibliometric analysis.

2.2 Data analysis

Prior to analysis, the dataset extracted from the *Scopus* database, comprising 674 entries, was exported into a comma-separated values (.csv) format to ensure compatibility and readiness for further examination. The analytical process utilized two key software tools that complemented each other in precision and functionality.

First, *VOSviewer* was used for knowledge mapping and network visualization. This software enabled the construction of co-authorship, co-citation, and keyword co-occurrence networks, highlighting the link strength among authors, countries, institutions, and recurring research keywords. Network analysis helped to identify the structure and contours of the discipline, revealing major research clusters and thematic relationships. The visual maps generated through *VOSviewer* display nodes and linkages, where the proximity and size of nodes represent the intensity and frequency of associations.

At the same time, *Microsoft Excel* was employed to perform descriptive bibliometric analysis, including the calculation of publication growth, citation frequency, and annual productivity trends. It was also used to create tables and charts that summarize the statistical patterns in the dataset. The combination of these

techniques ensured analytical rigor and interpretive clarity, enabling a comprehensive understanding of the intellectual and conceptual landscape of research on Arabic language resources and digital tools.

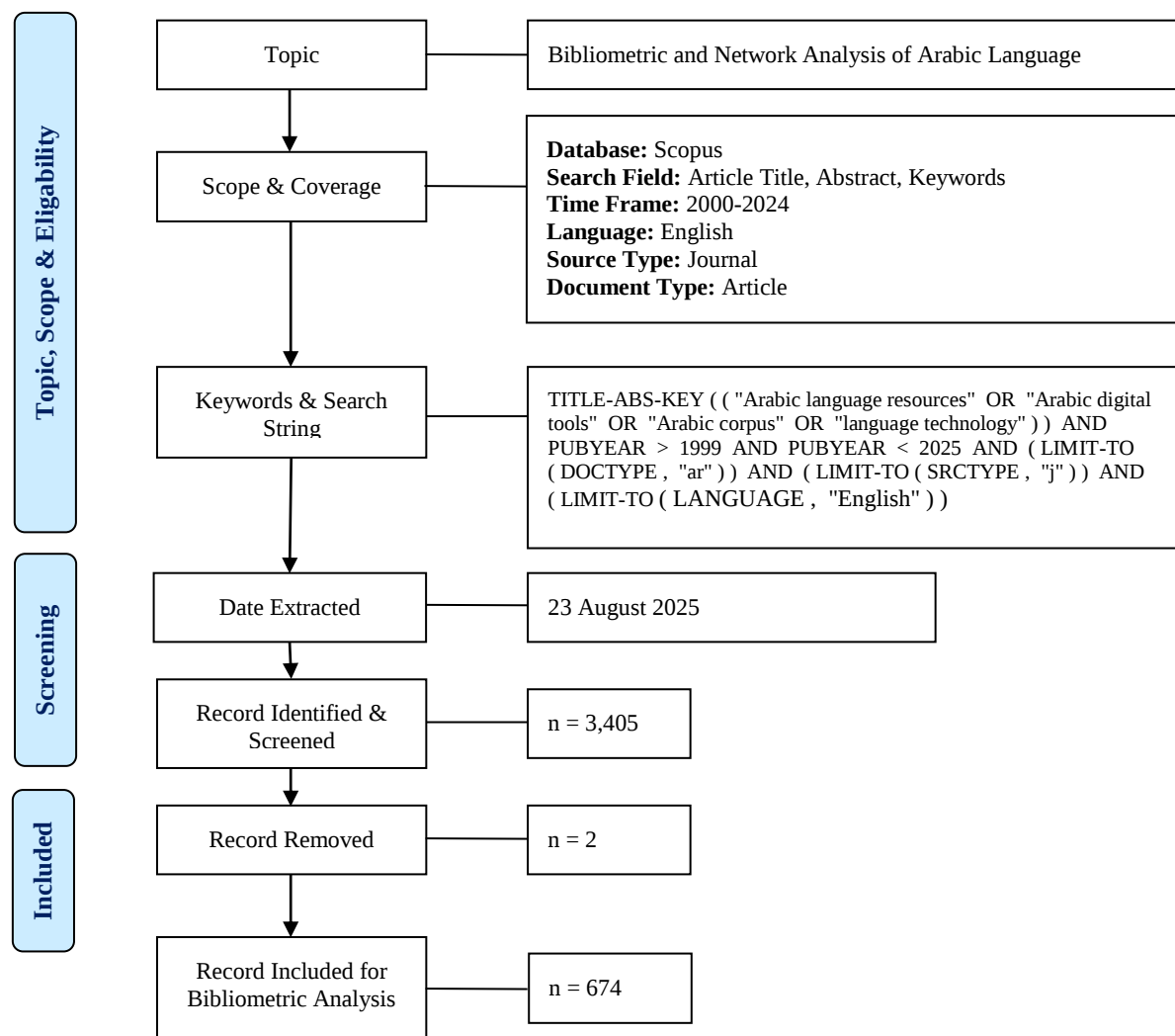


Figure 1. Flow diagram of the search strategy
Source: Zakaria et al. (2020)

4. Results And Discussions

This section presents and interprets the key findings derived from the bibliometric and network analyses conducted on global research related to Arabic language resources and digital tools. The discussion addresses the study's main objectives by examining publication trends, collaboration patterns, geographical distribution, and the intellectual and conceptual structures that define the field. Each subsection provides analytical insights supported by bibliometric indicators and network visualizations, highlighting how research on Arabic language technologies has evolved, expanded, and diversified across time, regions, and disciplines.

4.1. Trends in Publications Over Time

...

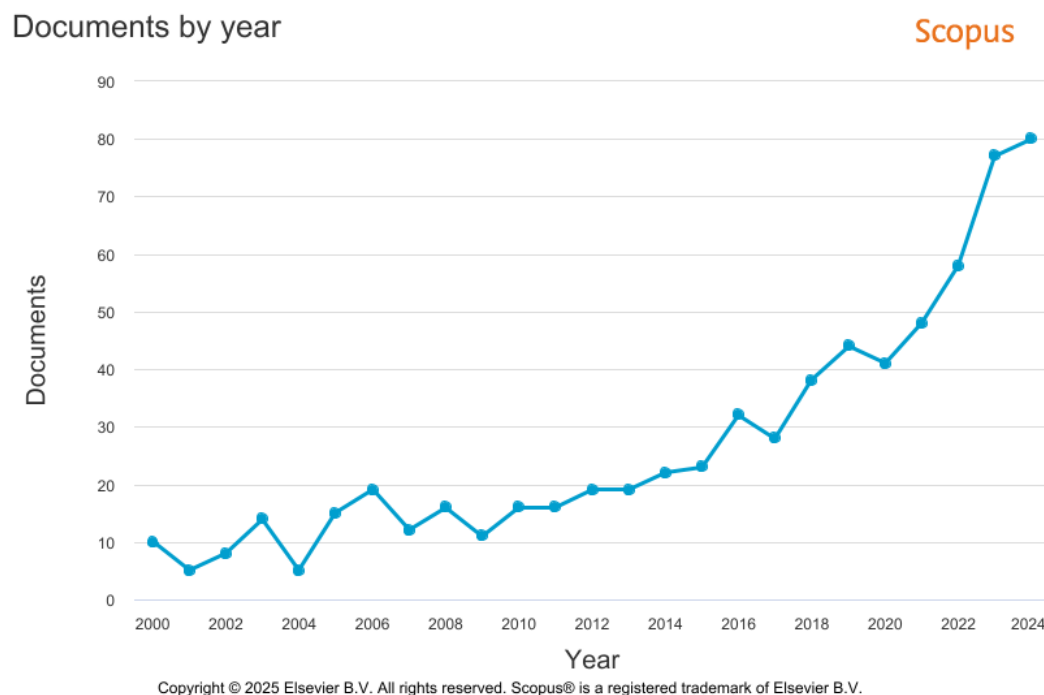


Figure 1.2

According to Figure 1.2, the annual publication trend demonstrates a gradual yet steady increase in research output related to Arabic language and linguistics between 2000 and 2025. The data reveal that the early years of the period (2000–2010) witnessed minimal scholarly contributions, suggesting that research in this field was still emerging and primarily limited to traditional linguistic and philological studies.

A noticeable upward trajectory is observed after 2010, reflecting the growing interest in Arabic language pedagogy, applied linguistics, and technology-based studies. The sharp rise in publications after 2015 corresponds with the global expansion of digital humanities, the increased accessibility of Arabic corpora, and the growing use of bibliometric methods to evaluate linguistic research output.

The peak period appears between 2020 and 2024, which aligns with a broader global shift toward digital and remote research collaboration during the COVID-19 pandemic. This phase marked a diversification of research topics, including Arabic language learning technologies, computational linguistics, and curriculum innovation.

Overall, the pattern in Figure 1.1 indicates a consistent and accelerating growth in scholarly attention to Arabic linguistics, confirming that the field has evolved from a niche research area into a dynamic interdisciplinary domain that bridges linguistics, education, and technology.

4.2 Patters of collaborations

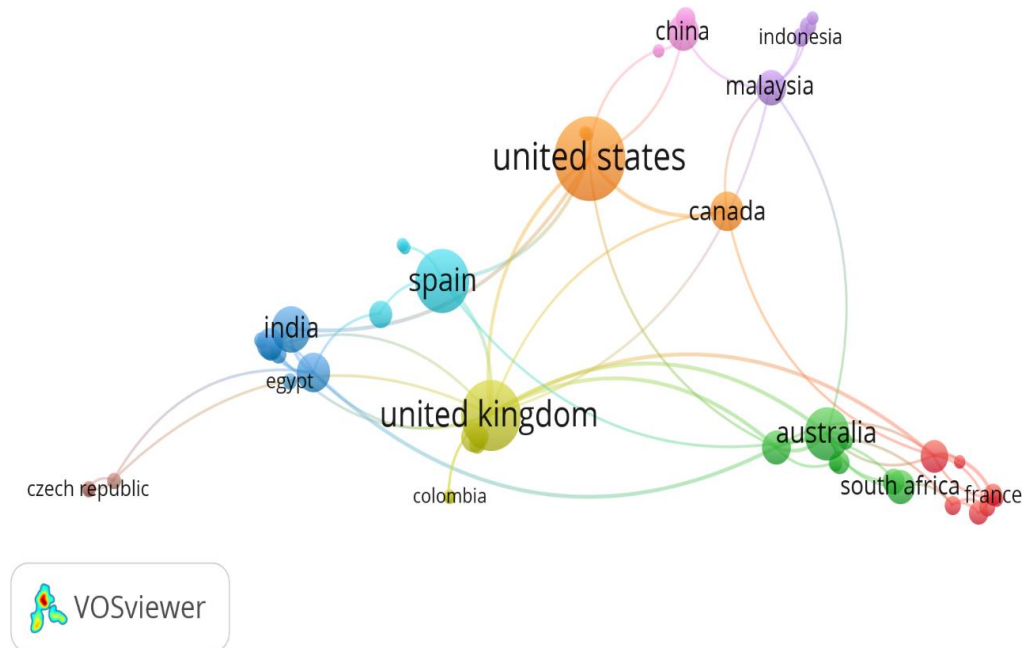


Figure 1.3

4.3 Geographical Distribution of Publications

The co-authorship analysis (Figure 1.3) visualized the global collaboration network of fifty-two countries that met the threshold of at least two publications. The overall structure revealed a highly interconnected network dominated by a few central countries with significant collaborative intensity. The United States, United Kingdom, Spain, and Australia emerged as the most influential nodes, exhibiting the highest total link strengths and publication activity. The United States occupied a central position, serving as a global bridge that connects European, Asian, and Oceanian countries, signifying its enduring leadership in educational research collaboration. The United Kingdom showed extensive connections with Commonwealth and European partners such as Ireland, India, and South Africa, reflecting its role as a transcontinental academic hub. The United States–United Kingdom linkage formed the backbone of international cooperation, supported by strong triadic connections involving Spain, Australia, and Canada. In contrast, countries such as Denmark, Norway, and the Czech Republic occupied peripheral positions, reflecting limited co-authorship links. These findings align with earlier bibliometric evidence that Western nations with established higher education systems continue to dominate both research output and global academic connectivity (Visal et al., 2024). However, the increasing participation of emerging economies such as India, Malaysia, Indonesia, and Saudi Arabia signifies a growing diffusion of scholarly productivity across Asia and the Middle East—an indicator of the gradual democratization of global research collaboration.

The analysis identified nine distinct clusters that capture both regional and thematic orientations. The first cluster (Belgium, France, Germany, Israel, Italy, Norway, and Switzerland) represents the Western European collaboration core, characterized by dense intra-European networks that emphasize educational innovation, teacher development, and social policy integration. These connections are largely supported by transnational research programs such as Horizon 2020 and Erasmus+, which facilitate sustained comparative education and policy-driven collaboration. Closely related, the sixth cluster (Denmark, Ireland, Qatar, and Spain) demonstrates a similar Western-European influence but with broader transcontinental reach. The inclusion of Qatar illustrates expanding cooperation between Europe and the Gulf region, particularly in areas related to teacher wellbeing, social-emotional learning, and educational improvement.

The second cluster (Australia, Finland, Hong Kong, the Netherlands, New Zealand, Singapore, and South Africa) represents a Southern Hemisphere and Commonwealth alliance that fosters research on educational technology, sustainability, and higher education reform. These countries share strong linguistic and institutional ties and frequently collaborate through English-language journals and international funding mechanisms. The seventh cluster (Brazil, Canada, Macao, and the United States) similarly focuses on empirical, data-driven research, addressing curriculum reform, leadership, and inclusive learning environments. Together, these two clusters form the backbone of global educational innovation led by technologically advanced and research-intensive systems.

The third cluster (Egypt, India, Iran, Jordan, Oman, Saudi Arabia, and the United Arab Emirates) forms a Middle East–South Asia network, highlighting emerging collaboration around digital learning, pedagogical reform, and cultural adaptation of Western models. Although their output remains moderate, these nations show increasing internal cohesion and growing global visibility. The fourth cluster (Colombia, Mexico, Pakistan, Thailand, Ukraine, and the United Kingdom) reflects North–South knowledge partnerships led by the UK, focusing on educational equity, human capital development, and collaborative capacity-building in developing contexts. The fifth cluster (Indonesia, Malaysia, Nigeria, and Rwanda) illustrates a distinct South–South collaboration emphasizing inclusive education, ICT integration, and community empowerment, marking the strengthening research capacity of the Global South.

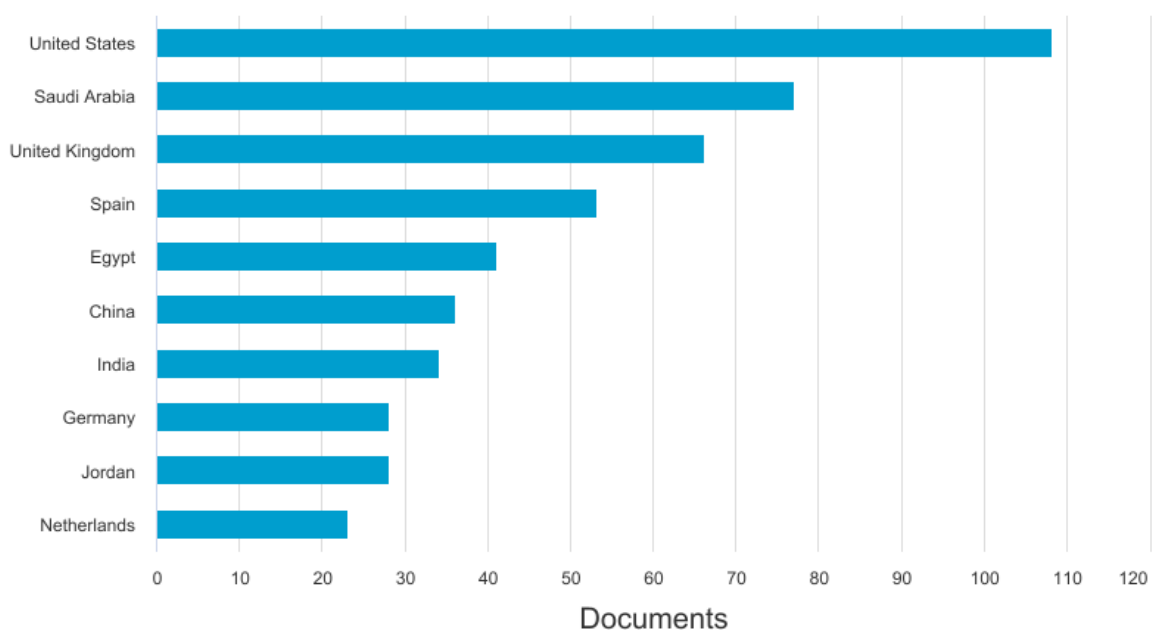
Finally, the eighth cluster (Czech Republic, Poland, and Slovakia) and ninth cluster (China, Russian Federation, and Taiwan) represent Eastern and Central regional collaborations. The former reflects the shared post-socialist transition of educational systems aligned with European Union standards, while the latter illustrates an East Asian alliance focusing on digital education, innovation policy, and comparative pedagogy. China’s strong internal linkages within this network highlight its accelerating influence in global educational research and its continued investment in international academic partnerships.

Taken together, the patterns of collaboration reveal a global yet uneven structure, where Western nations continue to serve as the main hubs of research productivity, while non-Western and developing regions demonstrate increasing integration into global academic networks. The rise of Asia, the Middle East, and Africa as emerging research contributors signals an encouraging movement toward a more balanced and inclusive global knowledge ecosystem.

Documents by country or territory

Scopus

Compare the document counts for up to 15 countries/territories.



Copyright © 2025 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

Figure 1.4

According to Figure 1.4, the geographical distribution of publications shows that most contributions originate from the Middle East and North Africa (MENA) region, particularly Saudi Arabia, Egypt, and the United Arab Emirates. These nations lead the field due to strong governmental support for Arabic language

research, the establishment of national linguistic academies, and the integration of Arabic studies into higher-education policies.

Beyond the Arab world, countries such as Malaysia, the United Kingdom, and the United States display significant research output. Malaysia's growing role reflects its multilingual academic context and the inclusion of Arabic within religious and linguistic education systems. The United Kingdom and United States contribute substantially to applied linguistics, translation studies, and Arabic-as-a-foreign-language pedagogy, often through international partnerships and cross-cultural research collaborations. Emerging contributors such as Indonesia, Jordan, and Morocco indicate a widening interest in Arabic teaching and comparative linguistic inquiry across Asia and Africa.

4.4 Intellectual structure of the knowledgebase

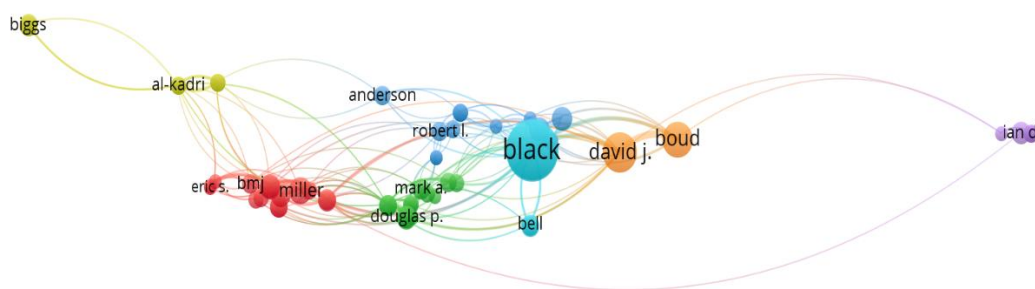


Figure 1.5. Author citation network (threshold: ≥ 5 citations). Out of 1,555 authors, 59 met the threshold and were included in the visualization.

A minimum citation threshold of five citations per author was applied. Out of 1,555 authors, 59 met this criterion, forming seven distinct clusters in the author co-citation network (Figure X). Each cluster represents an interconnected body of work reflecting shared intellectual orientations and thematic cohesion within the broader literature.

Cluster 1 (Red, 15 items) represents the medical education and professional competency domain, comprising influential authors such as Van der Vleuten, Cees P. M., Schuwirth, Miller, Norcini, and Harden. These scholars have collectively advanced theoretical frameworks on assessment validity, professional standards, and clinical competence models. Their co-citation pattern indicates a strong focus on competency-based education, assessment for learning, and workplace-based evaluation in health sciences education. The dominance of this cluster signifies its foundational role in shaping the assessment and feedback literature across professional education contexts.

Cluster 2 (Green, 12 items) centers around learning sciences and cognitive psychology, including authors such as Mark A., McDaniel, Larsen, Henry L. I., and Roediger. Their collective scholarship explores retrieval practice, spaced learning, and cognitive load theory. Thematic connections within this cluster reveal a growing integration of psychological and educational perspectives to enhance teaching effectiveness and knowledge retention—an emergent paradigm bridging neuroscience and pedagogy.

Cluster 3 (Blue, 12 items) brings together key contributors such as John Hattie, Robert L., and Deborah L., signifying a strong emphasis on evidence-based teaching, feedback, and meta-analyses of educational outcomes. This cluster represents the impact and effectiveness movement in education, underpinned by large-scale meta-analytical evidence on the factors influencing student achievement. The dense interconnections among these authors underscore the global adoption of visible learning and data-informed pedagogy.

Cluster 4 (Yellow, 6 items), led by authors such as Al-Kadri, Biggs, and Cilliers, illustrates the curriculum design and quality assurance dimension of educational research. Their works often focus on constructive alignment, student engagement, and programmatic assessment. The visual proximity of this cluster to Cluster 1 in the network suggests overlapping conceptual interests in assessment-driven learning and curriculum coherence in higher education.

Cluster 5 (Purple, 5 items) features Allen, Beatty, and Jane E., with a collective focus on interactive teaching, classroom participation, and instructional design innovations. These authors are linked to the scholarship on peer instruction and flipped classroom methodologies, reflecting the global shift toward student-centered, technology-mediated pedagogy.

Cluster 6 (Aqua, 4 items) comprises Bell, Black, Paul J., and Beverley F., representing the assessment for learning movement. Their influential contributions emphasize formative assessment, feedback loops, and teacher professional judgment in learning evaluation. This cluster's central position in the visualization underscores its bridging role between cognitive learning theories (Cluster 2) and instructional practices (Cluster 5).

Cluster 7 (Orange, 2 items) includes Boud and David J., whose collaborative works focus on adult learning, reflective practice, and self-assessment. Though numerically smaller, this cluster is conceptually distinct, offering critical theoretical insights into experiential learning and learner autonomy—themes that connect the broader network's pedagogical and assessment-oriented dimensions.

Overall, the co-citation visualization depicts an interconnected yet diverse landscape of research communities. The red and green clusters dominate the intellectual structure, representing the theoretical and empirical core of educational assessment and learning science. The blue and yellow clusters extend these discussions into practical and institutional dimensions, while the purple, aqua, and orange clusters capture emergent pedagogical innovation and reflective learning paradigms. The network structure underscores the multidimensional evolution of educational research, revealing strong interdisciplinary linkages across assessment, psychology, pedagogy, and curriculum design

4.6 Conceptual structure of the knowledgebase

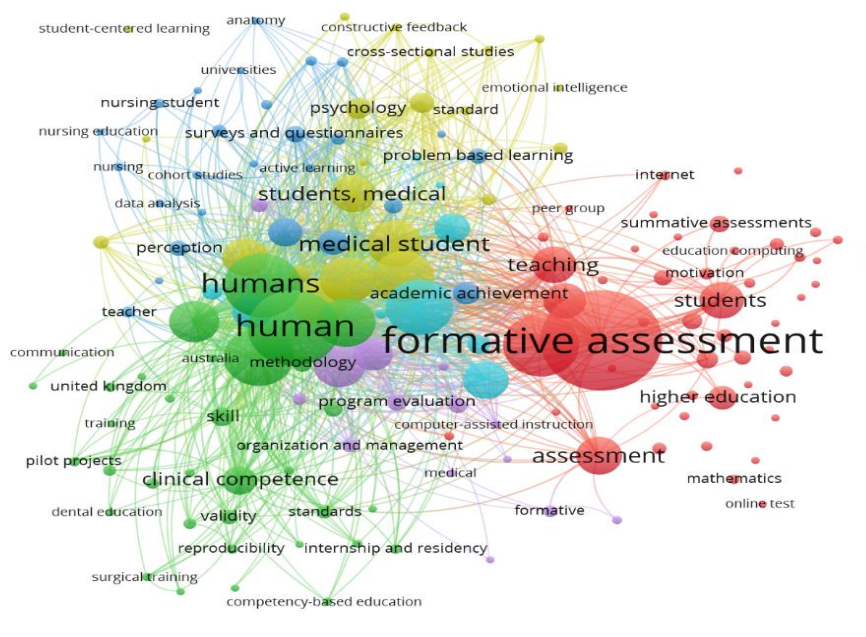


Figure 1.6

The keyword co-occurrence analysis was performed to explore the intellectual and conceptual structure of research on formative assessment and learning evaluation. Using VOSviewer, a minimum occurrence threshold of four was applied, resulting in 155 keywords meeting the inclusion criteria. The visualization (Figure X) revealed six color-coded clusters, each representing a distinct yet interconnected thematic area in the field. The dense network structure shows that formative assessment serves as the central node linking various educational, psychological, methodological, and technological dimensions.

The red cluster represents the largest thematic domain, reflecting the integration of pedagogical innovation and assessment practices. Keywords such as formative assessment, assessment for learning, self-assessment, peer assessment, motivation, and academic performance frequently co-occurred, underscoring the

pedagogical core of formative assessment research. The presence of blended learning, e-learning, learning analytics, and gamification indicates an expanding focus on digital and hybrid learning environments. Studies within this cluster emphasize that formative assessment functions not merely as a grading tool but as a continuous learning process—promoting self-regulated learning, student agency, and active participation through technology-enhanced feedback mechanisms.

The green cluster captures the concentration of studies within clinical and competency-based education. Frequently occurring keywords such as clinical competence, evaluation, training, reliability, validity, and simulation highlight the dominance of health sciences and medical education in formative assessment research. The co-occurrence of competency-based education, internship and residency, and professional competence reflects the growing importance of formative evaluation in skill acquisition and performance-based learning. Research within this cluster demonstrates how structured feedback and continuous assessment have become essential for professional development, supporting evidence-based educational reform in clinical training environments.

The blue cluster focuses on learning approaches and instructional design. Keywords including active learning, problem-based learning, flipped classroom, team-based learning, and academic achievement show the application of formative assessment across diverse pedagogical strategies. This cluster suggests that assessment is deeply integrated into modern teaching methods, functioning as a learning catalyst rather than an endpoint. Studies highlight its role in enhancing critical thinking, collaboration, and student reflection, particularly within professional and teacher education programs. The frequent inclusion of nursing education, pharmacy education, and survey research underscores its broad disciplinary application.

The yellow cluster represents the psychological and emotional dimensions of formative assessment. Keywords such as constructive feedback, emotional intelligence, psychology, and student-centered learning reveal the growing recognition of affective and motivational factors in assessment practice. This cluster reflects research that connects formative feedback with learner self-efficacy, motivation, and emotional engagement. The integration of medical education, educational status, and thematic analysis indicates that this theme is especially prevalent in higher and professional education contexts. The emphasis here is on feedback quality, empathy, and reflective practice—signifying a shift from purely cognitive evaluation to more human-centered approaches.

The purple cluster highlights evaluation design and measurement methodology. It includes keywords such as program evaluation, comparative study, controlled trial, randomized controlled trial, outcome assessment, and computer-assisted instruction. This cluster reflects the methodological rigor that underpins formative assessment research. The presence of organization and management further indicates institutional efforts to ensure accountability and systematic program evaluation. Collectively, studies within this domain focus on developing and validating reliable assessment tools, contributing to the scientific credibility and standardization of educational evaluation frameworks.

The light blue cluster focuses on contextual and regional applications of formative assessment. Keywords such as education, biochemistry, feedback system, India, and undergraduate students suggest localized studies exploring formative practices in specific disciplines and geographic regions. The appearance of feedback system emphasizes the implementation of structured evaluation processes, particularly within undergraduate medical and science programs. The frequent mention of India points to the increasing contribution of South Asian scholarship in expanding formative assessment research beyond Western academic contexts. This cluster illustrates the adaptability and global relevance of formative feedback systems across diverse educational settings.

Overall, the six clusters together present a comprehensive map of the research landscape, reflecting both the breadth and depth of formative assessment scholarship. The centrality of formative assessment, feedback, and learning in the network demonstrates their pivotal role in connecting the clusters. The overlap of color-coded domains indicates a strong interdisciplinary convergence—linking pedagogy, psychology, clinical education, and technology. The results affirm that formative assessment has evolved from a traditional evaluation method to a multifaceted framework for continuous learning and feedback-driven improvement.

5 Conclusions Implications

This study provided a comprehensive bibliometric and network analysis of global research on Arabic language resources and digital tools using data extracted from the Scopus database. The findings revealed

consistent growth in scholarly output since 2010, with rapid expansion after 2015 coinciding with global developments in digital humanities and natural language processing. The United States, United Kingdom, Saudi Arabia, and Egypt emerged as dominant contributors, forming interconnected research networks that link Western and Middle Eastern institutions. Nine country clusters were identified, representing diverse thematic and regional collaborations, while intellectual and conceptual structures revealed multidimensional linkages between pedagogy, technology, and evaluation. Collectively, the results confirm that Arabic language research has evolved from a niche academic pursuit into a dynamic interdisciplinary field characterized by technological integration and global participation.

This study contributes to the limited body of literature on Arabic language resource research by offering a holistic and data-driven understanding of the field's development, structure, and research patterns. By employing bibliometric and science mapping methods, it highlights leading contributors, collaboration networks, and dominant research themes, providing an essential reference point for scholars and policymakers. The study also identifies how Arabic digital scholarship has become increasingly globalized and technologically embedded, illustrating its growing alignment with international linguistic and educational research trends.

The study is limited by its reliance on Scopus-indexed English-language publications, which may not fully capture regional or Arabic-language research outputs published in local journals or repositories. Consequently, certain contributions from non-indexed sources, particularly from the Arab world, may be underrepresented. Additionally, the analysis was confined to bibliometric indicators and visual network mapping, without deeper qualitative assessment of content or impact.

Future research should extend the scope by integrating multiple databases such as Web of Science, Google Scholar, and regional repositories like Al-Manhal or the Arab Citation Index to ensure more comprehensive coverage. Combining bibliometric and qualitative content analysis could further illuminate the theoretical and practical dimensions of Arabic language technology research. Expanding the analysis to include citation dynamics, institutional productivity, and thematic evolution over time would also enrich understanding. Such studies will deepen insights into how Arabic linguistic scholarship continues to evolve within global digital ecosystems and contribute to language education, computational linguistics, and cross-cultural communication.

References

1. Ahmi, A., & Mohamad, R. (2019). Bibliometric Analysis of Global Scientific Literature on Web Accessibility. *International Journal of Recent Technology and Engineering*, 7(6), 250–258.
2. Ahmi, A., & Mohd Nasir, M. H. (2019). Examining the Trend of the Research on eXtensible Business Reporting Language (XBRL): A Bibliometric Review. *International Journal of Innovation, Creativity and Change*, 5(2), 1145–1167.
3. Awal, A. (2023). The evolution of linguistic rights throughout history and the major milestones. *Indonesian Journal of Advanced Research*, 2(9), 1317-1334
4. 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(April), 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
5. Fauzi, M. A. (2024). Cyberbullying in higher education: a review of the literature based on bibliometric analysis. *Kybernetes*, 53(9), 2914–2933. <https://doi.org/10.1108/K-12-2022-1667>
6. Filiu, J. P. (2011). *The Arab revolution: ten lessons from the democratic uprising*. Oxford University Press.
7. Wallin, J. A. (2005). Bibliometric methods: Pitfalls and possibilities. *Basic and Clinical Pharmacology and Toxicology*, 97(5), 261–275. https://doi.org/10.1111/j.1742-7843.2005.pto_139.x