Journal of Pedagogy and Education Science
Vol. 5, No. 01, pp. 127-143
journal.iistr.org/index.php/JPES
DOI: 10.56741/IISTR.jpes.01513

Trends in Corpus Linguistics Research: A Bibliometric Analysis (2004-2024)

[1]Varia Virdania Virdaus, [1]Gunarti Dwi Lestari*, [1]Wiwin Yulianingsih, [1]Rofik Djalal Rosyanafi, [2]Muchamad Arif, [3]Hamidah Binti Mohamad
Corresponding Author: *gunartilestari@unesa.ac.id
[1] Universitas Negeri Surabaya, Surabaya, Indonesia
[2] Universitas Narotama, Surabaya, Indonesia
[3] Universiti Tun Abdul Razak, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history
Received 22 October 2025
Revised 28 January 2026
Accepted 6 February 2026

ABSTRACT

This bibliometric study analyzes changes in corpus linguistics research from 2004 to 2024, emphasizing discourse analysis, digital literacy, and educational technology. Despite being the crucial tool for language analysis, corpus linguistics is frequently not taught in teacher preparation programs. Literacy has evolved digitally as a result of the growing prevalence of computers and the internet. Research indicates that educators must improve their technological proficiency. Using an analysis of 128 papers from the Scopus database, this study looks at the gaps in the extensive body of scientific literature on corpus linguistics. Studies indicate that scientific output has increased significantly in recent years, particularly since 2018. Papers are currently the most widely used publication format. The United States excels at assisting and cooperating with other nations. According to the investigation, Robert C. Meurant and Rodney H. Jones were the primary authors and had the highest number of citations and contributions, respectively. Traditional language methods are no longer the main focus of research; instead, an interdisciplinary framework that combines digital technology and education has taken its place. As a result, there is now a close connection between instruction and online and data resources. According to the research, literacy in corpus linguistics is now seen as a skill influenced by contemporary teaching strategies, the accessibility of data, and student engagement in the classroom.

Keywords
Bibliometric Analysis
Corpus Linguistics
Digital Literacy
Education
Research Trend

This is an open-access article under the CC–BY-SA license.

Introduction

Corpus linguistics (CL) refers to the study of language based on real-world examples of language use collected for linguistic research. It involves the computational analysis of large corpora consisting of written texts and spoken recordings. Corpus linguistics examines naturally occurring language samples using specialized software and enables researchers to interpret data both quantitatively and qualitatively. Importantly, corpus linguistic is not a theory of language but a methodological approach for analyzing linguistic data. In the context of language education, corpus linguistics is an effective tool for teachers to develop students' ability to write essays, reports, and similar texts [1].

The fundamental problem with using corpus linguistics in language education is that there is a gap between its theoretical promise as a useful tool and the practical problems it faces, such not having enough training, being hard to access, and not being used enough in the curriculum. Corpus linguistics is still not widely used and depends on individual initiative, therefore most teachers and future teachers still utilize their gut feelings instead of real evidence when teaching language [2–3].

A study also suggests potential implications for language teacher educators in effectively introducing CL to student teachers within virtual learning environments [4]. Teacher training programs still don't talk enough about using ready-made language corpora to prepare class data and make materials. They also don't put future teachers in the role of linguistic researchers who are watching developments in the foreign (English) language. A study showed that teachers and students learned a new foreign language over time through existing corpora, text retrieval and compilation, and lastly self-help concordances. As part of digital literacy, the teacher will give an outline of how to make your own corpus [5].

The technology revolution has changed the way we read and write a lot, from paper formats like newspapers and notebooks to digital formats like web sites and social networks. So, the fact that computers and the Internet so widely available has changed the way people think about literacy in a big way, moving the idea into a digital setting (i.e., "digital literacy") [6]. Digital literacy encompasses the ability to find, evaluate, and share information using digital tools. This is closely related to information literacy and requires mastery of digital technology for effective information utilization [7].

There are studies showing that teachers really do need to improve their digital literacy when they use digital technologies to teach literature, language, or any other subject [8]. There is an increasing focus on incorporating digital literacy into educational strategies to mitigate disparities and equip pupils for the digital era. Nevertheless, there is a necessity for more extensive norms and discussions regarding the ramifications of digital technology in education [9–10]. The incorporation of digital instruments in education, including e-learning platforms and corpus linguistics tools, can augment student participation and autonomy. Nevertheless, educators frequently encounter obstacles, including inadequate training and limited time to cultivate digital literacy competencies. Therefore, learning digital skills is essential [11–12].

Corpus linguists may want to look into teachers' expertise of discourse analysis in the field of Corpus Linguistics from different points of view [13]. There isn't a lot of scientific data on Corpus Linguistics from around the world. Bibliometric analysis is one of the ways that scholars study literature. Bibliometric analysis is a good way to figure out how a work has affected the growth of science [14–15]. Researchers have often utilized bibliometric barometers to look at trends. These include research fields, paper sources, publication output, language sources, nation and institution distribution, the first author's place, the number of citations, and keywords [14], [16].

This study analyzes trends in corpus linguistics spanning 2004-2024 that explore discourse analysis, digital literacy, language education, and the use of technology in education. Each research aims to develop an understanding of the role of technology in media and education and to identify patterns in digital communication that can enhance learning and discourse analysis.

Corpus linguistics can be conceptually positioned as a pedagogical approach that supports the development of digital literacy through data-driven engagement with authentic language use. By interacting with corpora, learners are required to search, evaluate, interpret, and apply linguistic data, processes that align closely with core components of digital literacy. From a learning theory perspective, corpus-based learning reflects principles of self-regulated learning (SRL), as learners actively set learning goals, monitor patterns in language data, and reflect on their findings. At the same time, corpus linguistics is consistent with constructivist and data-driven learning frameworks, in which knowledge is constructed through exploration, discovery, and interaction with real-world data rather than passive instruction. This conceptual linkage helps situate corpus linguistics not only as a research method but also as a meaningful pedagogical tool within digitally mediated learning environments.

Grounded in this conceptual framework, this study aims to analyze trends in corpus linguistics research from 2004 to 2024 using a bibliometric approach. Specifically, this study addresses the following research questions: i) What is the main information about the data,

documents, and citations of corpus linguistics in 2004-2024? ii) Which sources contain research on corpus linguistics worldwide between 2004 and 2024 and the distribution across countries? iii) How is the contribution and impact of authors in the worldwide linguistic corpus analyzed? iv) How have research trends on literacy developed in corpus-based linguistic studies in the past two decades? v) How do the corpus linguistics literature's representations of teaching, the classroom, students, and digital tools relate to literacy?

Material and Methods

This study utilized a literature and bibliometric analysis of publications to provide a valuable reference for future research [15]. The term of bibliometrics was initially introduced by Pritchard [17], with the expectation that it would be openly utilized in any research aimed at quantifying the communication writing process and would swiftly achieve recognition within the field of information science [18]. This method is well-suited for analyzing research trends, including the distribution of papers by country, author contributions, and citation counts. Researchers optimized Scopus Elsevier, the preeminent academic bibliographic database globally.

“Corpus Linguistics” was searched on the Scopus website using the titles, abstracts, and keywords from 2004 to 2024. The search strategy was: TITLE-ABS-KEY ("corpus" OR "linguistics" AND "digital literacy") AND (EXCLUDE (PUBYEAR , 2025)). The data were collected in August 1, 2025. Researcher narrowed down their search to studies published between 2004 and 2024, using the "OR" and "AND" operators and the keywords "corpus, linguistics, and digital literacy." This gave researchers 128 articles that they then looked at more closely. Also, descriptive analysis was done to make sense of the data. Researcher used Microsoft Excel to look at this data. The primary things that were looked at in the research were the types of information used, the sources of the papers, the language used, the distribution of the papers by country and organization, the contributions and effects of the authors, the authors' origins, the most cited works, and the trends in research from 2004 to 2024. The R Studio 2024.12.1.0 version [19] and VoSViewer 1.6.20 version programs were used to find patterns in corpus linguistics studies. This study employed a bibliometric method to analyze data on corpus linguistics [20].

Results and Discussions

A. Publication and Citation Trends

Fig. 1 illustrates the annual scientific output of corpus linguistics research, based on the number of articles published per year. The graph shows slow growth in the early years, with only one document published in 2004. This was followed by a sharp increase in publications starting around 2018, indicating significant growth in the field's scientific production. This

upward trend reflects several factors, including an increase in research capacity, growing interest in related topics, and easier access to publication platforms. The trend may also be influenced by the impact of global events, such as technological advances or pandemics. The study noted a particularly sharp increase in 2021 to 18 documents, followed by a slight decrease in 2022, before reaching a high of 22 documents in 2025. This consistent, upward-trending fluctuation suggests that corpus linguistics research continues to gain attention from researchers over time.

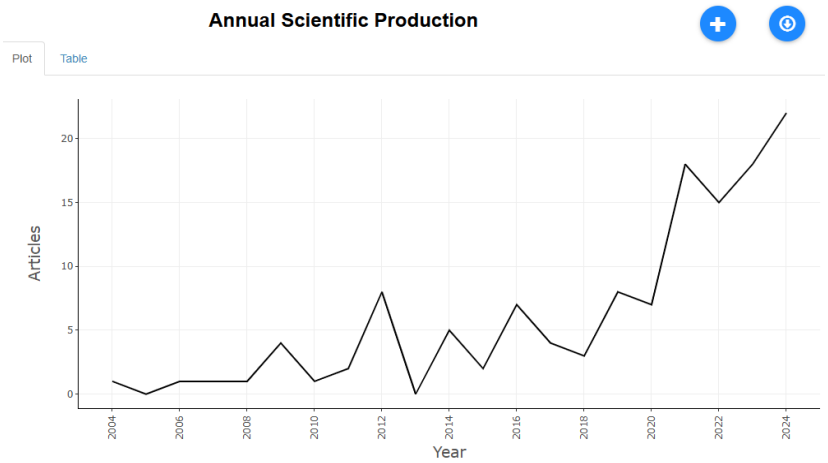


Fig. 1. Annual Scientific Production

The fluctuating annual trend in research on corpus linguistics shows that the field is gaining attention from researchers. Specifically, the number of documents published increased sharply to 18 in 2021 before a slight decrease to 15 in 2022. The trend then continued its upward trajectory, reaching 22 documents in 2025. This growth reflects an increased research capacity and greater interest in the topic over time.

The bibliometric trends identified in this study reflect more than a quantitative growth in corpus linguistics research; they indicate a qualitative shift in how language learning and literacy are conceptualized. The sharp increase in publications after 2018 coincides with the expansion of digital learning environments and greater access to linguistic data, suggesting that corpus linguistics is increasingly viewed as a pedagogical resource rather than solely a linguistic research method. The prominence of keywords related to literacy, digital tools, and classroom contexts signals a transition toward learner-centered and data-driven pedagogies, where learners actively engage with authentic language data. This shift aligns with contemporary educational paradigms that emphasize autonomy, critical inquiry, and self-regulated learning, positioning corpus-based approaches as facilitators of deeper cognitive and metacognitive engagement in language learning.

DOCUMENT TYPES	
article	69
book	14
book chapter	10
conference paper	24
conference review	2
erratum	2
letter	1
review	6

Fig. 2. Document Type

Fig. 2 shows the distribution of document types in corpus linguistics research from 2004 to 2024. Articles are the most common, with 69 publications, followed by conference papers at 24. Books rank third with 14 publications, while book chapters follow with 10. The remaining document types include reviews (6), and both conference reviews and errata, which each have two publications. Overall, the data illustrates that articles are the dominant publication format, followed by conference papers and books.

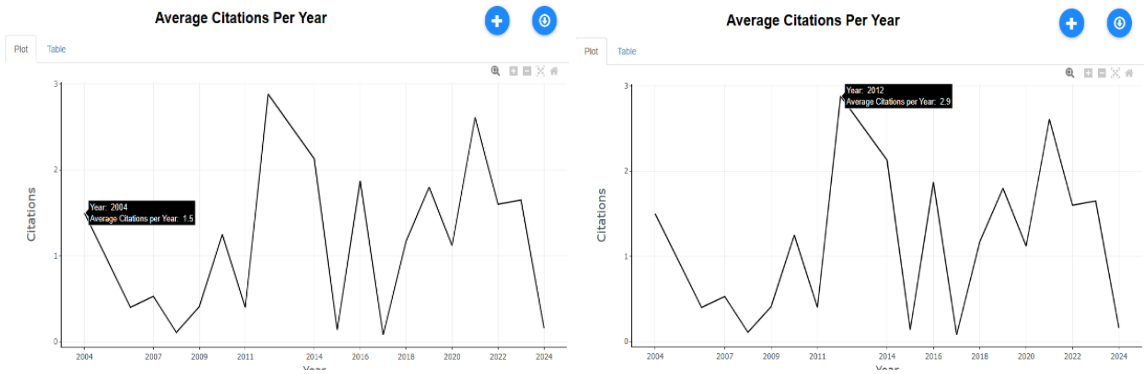


Fig. 3. Average Citation Per-Year

Fig. 3 illustrates the volatility of the average citations per year for corpus linguistics research from 2004 to 2024. The average began at 1.5 in 2004, but then reached its lowest points of 0.1 in 2008, 2015, and 2017. The highest peak was in 2012, with an average of 2.9 citations, followed by a smaller peak of 2.6 citations in 2021. These significant fluctuations in average citations suggest that the field has experienced periods of both low and high impact. This volatility may reflect the intermittent publication of highly influential papers that garner a large number of citations, interspersed with periods of less-cited research. The peaks in 2012 and 2021 indicate a greater scholarly interest and broader recognition of works published during those years, while the dips could signify a shift in research focus or a gap in seminal publications.

B. Sources and Global Distribution of Publications

Fig. 4 presents an overview of the most relevant journals in the corpus linguistics field. The top two journals, Communications in Computer and Information Science and Computers and Composition, each published four documents. Following them, the journal *LIDIL* has three

documents, while eight other sources each contribute two. This data is crucial for identifying scholarly centers of influence and evaluating the literature's focus within this research domain. The visualization was generated using the "Most Relevant Sources" menu in the Biblioshiny software, based on the number of documents published per source.

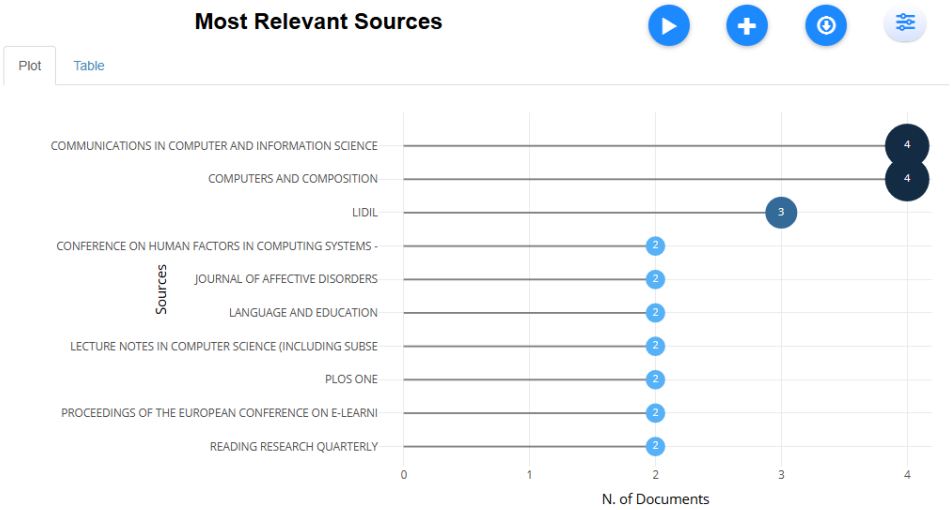


Fig. 4. Relevant Sources

Fig. 5 displays the distribution of publication contributions by country, specifically for corresponding authors in the bibliometric study. The visualization breaks down the total documents authored by each country into two types of collaboration: Single Country Publication (SCP), where all authors are from the same country; and Multiple Country Publication (MCP), which signifies international collaboration. This data is significant because it helps to identify the geographical centers of scholarship and illustrates the level of international cooperation in corpus linguistics research. This figure was generated using the Corresponding Author's Countries module in Biblioshiny, which facilitates the interpretation of both international collaboration and scientific production based on the lead author's country of origin. The visualization shows that the United States (USA) is the country with the highest contribution, having produced over 12 documents, with a notable proportion resulting from international collaboration (MCP). Other countries with substantial contributions include Canada, South Korea, and Spain. Countries like Germany, Hungary, and Belgium show lower contributions but still demonstrate involvement in scientific publications on the analyzed topics.

The data reveals that the United States (USA) is the country with the highest contribution to the field, having produced more than 12 documents. A significant proportion of these publications resulted from Multiple Country Publication (MCP), indicating a high level of international collaboration. Other countries demonstrating substantial contributions include Canada, South Korea, and Spain. Even nations with lower contribution figures, such as Germany, Hungary, and Belgium, show active involvement in scientific publications on the analyzed

topics. Overall, this visualization is significant as it not only identifies the geographical centers of scholarship but also clearly illustrates the level of international collaboration within corpus linguistics research. Furthermore, it aids in mapping global scientific networks related to this specific research domain.

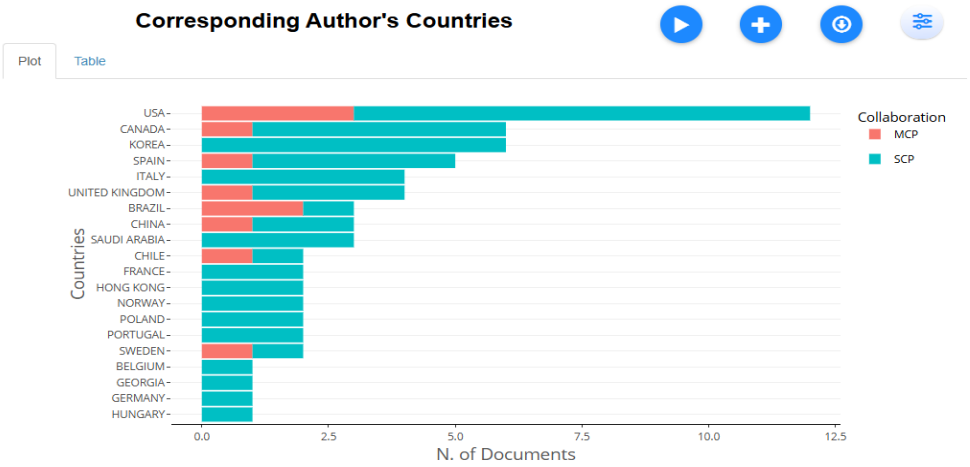


Fig. 5. Author's Countries

This visualization aids in mapping global scientific networks related to specific research topics in corpus linguistics. The figure itself was generated from the Corresponding Author's Countries module of Biblioshiny. By using a bar chart visualization approach, it effectively facilitates the interpretation of international collaboration and scientific production based on the country of origin of the lead author. This makes the level of cooperation and the geographical centers of scholarly influence immediately clear to the reader.

C. Author Contributions and Impact

Fig. 6 is a horizontal bar chart that offers a bibliometric representation of the ten most significant writers in the research domain, ranked by the volume of their document contributions. The horizontal axis displays the quantity of documents (N. of Documents), while the vertical axis lists the authors' names.

The author with the greatest number of contributions is Robert C. Meurant with a total of six documents: [21–26]. Rodney H. Jones with four documents, ranks second: [27–30]. Christoph A. Hafner submitted three documents: [29], [28], [31]. Seven additional writers each provided two documents: Alves: [32–33]; Cassany: [34–35]; De Sousa AP: [32–33]; Early : [36–37]; Gambarato: [38–39]; Gold: [40–41]; and Hautemo: [42–43]. This image indicates that the distribution of author contributions is diverse; yet, the majority of authors remain within the modest contribution range of two documents. The preeminence of Robert C. Meurant signifies the substantial impact of that author on the examined subject matter.

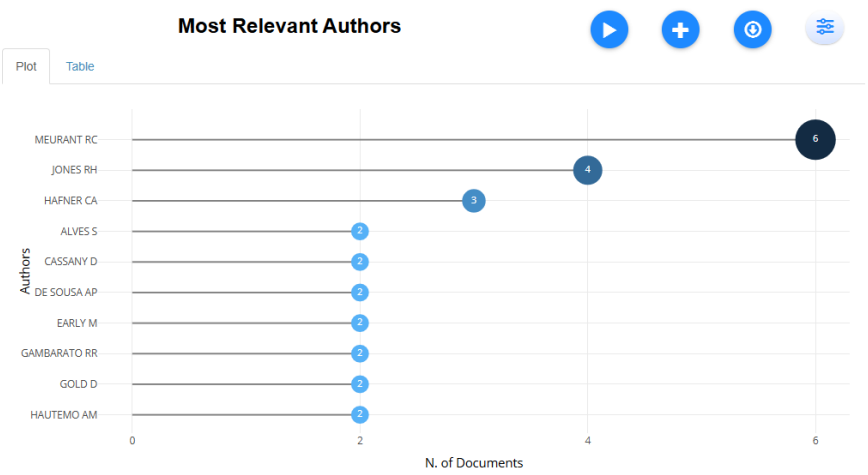


Fig. 6. Most Relevant Author

Fig. 7 below presents the local bibliometric metrics of several authors involved in a particular field of study. Some of the important indicators shown include: (a) h-index: measures the productivity and citation impact of an author in their field. (b) g-index: an alternative index that emphasizes the large contribution of highly cited articles. (c) m-index: the ratio of h-index to time (years since first publication), describing an author's annual productivity. (d) TC (Total Citations): the total number of citations received by the author. (e) NP (Number of Publications): the number of author publications in the local dataset. (f) PY_start (Publication Year Start): the first year the author engaged in publications in this field.

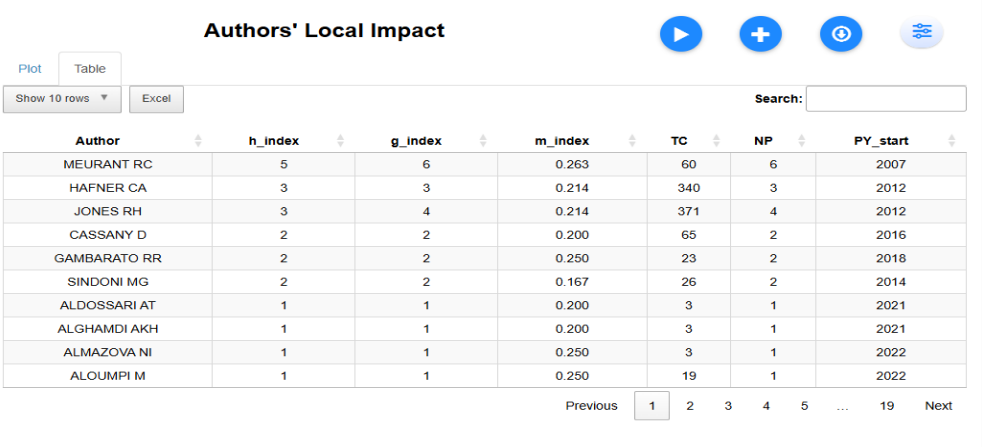


Fig. 7. Authors' Local Impact

Key findings in this study are: (1) Robert C. Meurant is the most prominent author locally, with values: h-index: 5, g-index: 6, m-index: 0.263, total citations (TC): 60, and number of publications (NP): 6. His initial publication in 2007 demonstrated his consistently significant contributions to the field [26]. (2) Rodney H. Jones, despite having a lower h-index: 3, has the highest total citations of 371, indicating that some of his publications have a huge impact despite their limited number.

Other authors such as Christoph A. Hafner, Cassany, and Gambarato also show relevant local impact, with h-index values between 2 and 3 and citations above 20. Newer authors such as Aloumri M (PY_start 2022) or Almazova [44] (PY_start 2021) show potential but are still in the early stages of contribution with low h-index values and citations.

Furthermore, we can identify Robert C. Meurant and Rodney H. Jones as thought leaders in this field based on their academic contributions. The high m-index of Robert C. Meurant indicates consistency of contributions per year. Authors with a high g-index also reflect the presence of some highly influential works that merit further review.

D. Evolution of Literacy Trends in Corpus-Based Studies

The bibliometric visualization utilizing co-occurrence analysis in VOSviewer illustrates the dynamics and trends of emerging topics in literacy study within a body of scientific literature (See Fig. 8). The term "literacy" emerges as the primary keyword in the network, indicated by its significant size and central placement inside the cluster.

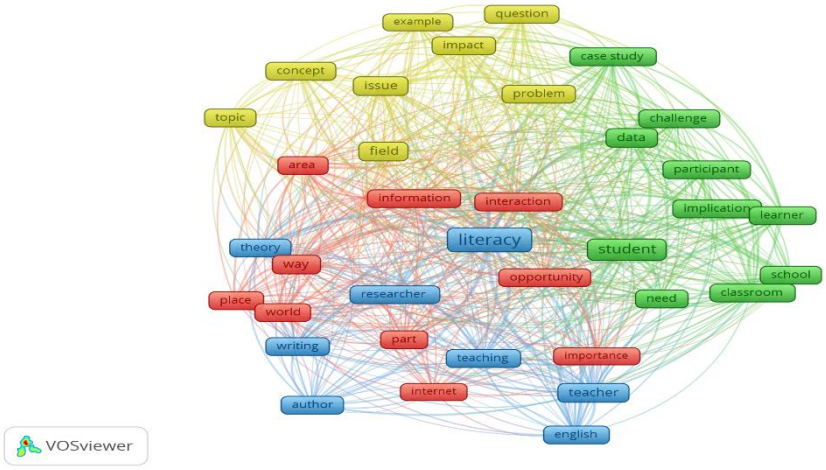


Fig. 8. Keyword Co-Occurrence Network

The data reveals that literacy is the most frequently employed term and serves as a focal point for many subtopics within this research domain [45]. Literacy research has progressed beyond linguistics, integrating perspectives from several fields, including education, which encompasses elements relating to students, classrooms, educators, learners, and institutions. Technology and data: data, internet, information, and interaction. The methodology and theoretical analysis involve the identification of the subject, issue, concept, and illustration. This trend signifies that corpus-based literacy research has evolved from traditional linguistic approaches to an interdisciplinary model, especially through the integration of digital technology and education [46–47].

The visualization, Mapping Research Trends by Cluster, illustrates the categorization of trends into distinct clusters. Central Emphasis of Cluster Azure Hue Instruction in academic literacy and language, encompassing literacy, composition, pedagogy, educators, and English

[48]. Verdant Educational dimension and student engagement (student, institution, classroom, participant, learner). Red indicates interaction, information, and learning environment (interaction, opportunity, information, internet). Yellow indicates conceptual issues and methodological ramifications (concept, impact, example, case study). The presence of these clusters indicates a shift from simply defining literacy to examining its application, pedagogy, and the influence of technology and social context. The graphic does not overtly represent time; nonetheless, the terms "literacy," "internet," "data," and "digital," in conjunction with the themes of "challenge," "interaction," and "opportunity," suggest a shift in focus towards digital literacy and online learning environments during the previous decade.

In summary, trends in literacy research within corpus-based language studies reveal an increased emphasis on digital learning and interactive contexts [49]. The notion of literacy is expanding to include social and digital competencies alongside reading and writing skills. Research trends integrate corpus-based empirical data, pedagogical approaches, and current challenges such as digital literacy, student engagement, and creative teaching strategies. Literacy encompasses reading and writing skills as well as social and digital competencies. This research trend combines corpus-based empirical data with educational methodology and current concerns, including digital literacy, student engagement, and creative teaching strategies.

E. Literacy's Intersection with Pedagogy and Digital Tools

In the academic literature that studies corpus data-based literacy, the following figure shows how the themes of education (teaching, classroom, student) and digital technology (internet, data, information) are shown.

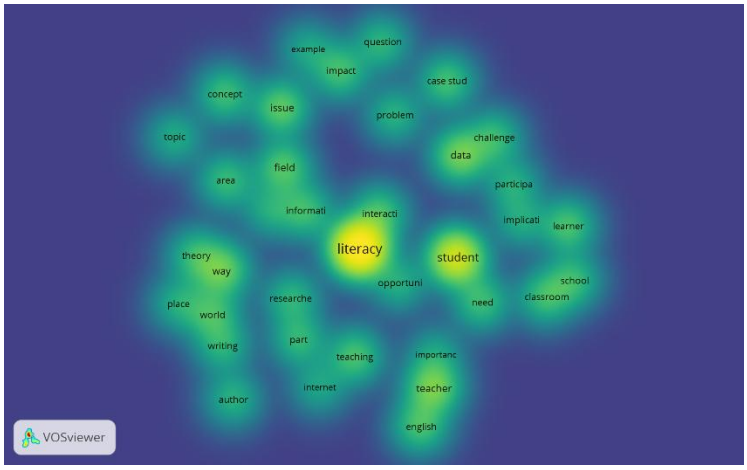


Fig. 9. Density Visualization

From Fig. 9, it can be assumed that the education keywords are: there is a strong link between these educational ideas because the words student, classroom, teacher, learner, and school are all in the same green cluster. The word teaching appears in the blue cluster and is closely associated with literacy, writing, research, and importance. This demonstrates that

pedagogical contexts frequently study literacy topics and emphasize the interaction between teachers and students in the classroom. Digital tools/data keywords are: internet, information, and data are spread across the blue and red clusters but are closely connected. What this means is that digital technologies and data play a big role in shaping modern ways of teaching literacy, which is similar to the digital corpus linguistics approach.

Literacy is the link between the education and digital worlds because of the way keywords relate to each other. There are also important links between learning to read and write and using technology, such as interaction, opportunity, and challenge. The close integration of teaching and technology (internet, data, information) signals a new direction in literacy research—digital and multimodal literacy in educational contexts [50].

Literacy in the context of corpus linguistics is no longer considered a mere reading-writing skill but as an ability shaped by classroom interaction (student, classroom, teacher), data access and analysis (internet, information, data), and modern corpus-based pedagogical approaches (teaching, research) [51], [47], [52].

Conclusion

The discipline is changing to adapt to the digital age, as demonstrated by this bibliometric analysis of corpus linguistics research from 2004 to 2024. According to the study, which examined 128 articles from Scopus, publications have been increasing, particularly since 2018. The most widely used type of document is an article. According to the data, the United States is the largest donor and makes a concerted effort to cooperate with other nations. Two distinguished authors, Robert C. Meurant and Rodney H. Jones, were found by the study. Their high h-index and cumulative citations attest to their important contributions to the field. The report demonstrates a distinct trend: a shift away from conventional language instruction techniques and toward a more interdisciplinary strategy that blends education and digital technologies. Terms like "literacy," "internet," "data," and "digital" highlight how important it is to be able to read and write online. According to the study, the definition of literacy is changing to include not only traditional reading and writing skills but also social and digital competencies. Technology and pedagogy together offer a fresh perspective on literacy research. Literacy is currently viewed as a skill that is developed through schooling, data analysis, and the use of contemporary, corpus-based teaching techniques. The results show that in order for teachers to effectively use resources like corpus linguistics, they need to improve their digital literacy.

The findings of this bibliometric analysis offer several practical and future-oriented implications. For educators, corpus linguistics can be integrated as a data-driven pedagogical tool to support digital literacy, learner autonomy, and critical language awareness in classroom practice. Teacher-training programs are encouraged to incorporate corpus-based learning and

digital literacy competencies into pre-service and in-service curricula, enabling teachers to design evidence-based instructional materials and to facilitate self-regulated learning. For future researchers, this study highlights the need for more empirical and context-specific investigations, particularly in underrepresented regions such as Indonesia, focusing on teachers' corpus literacy, classroom implementation, and learning outcomes. Further research may also explore interdisciplinary frameworks that connect corpus linguistics, educational technology, and learning theories to strengthen both theoretical and practical contributions to the field.

Implication and Recommendation

Based on the bibliometric findings, several practical and policy-oriented recommendations can be formulated for the Indonesian educational and research context. First, the steady global growth of corpus linguistics research—particularly after 2018—highlights an urgent need for Indonesian higher education institutions to integrate corpus linguistics more systematically into language education curricula. Teacher education programs in Indonesia should explicitly include corpus-based approaches as part of digital literacy and research methodology courses, enabling future teachers to move beyond intuition-based language teaching toward evidence-based pedagogical practices.

Second, the dominance of publications and collaborations from countries such as the United States, Canada, and South Korea indicates that Indonesian researchers remain underrepresented in international corpus linguistics scholarship. Therefore, Indonesian universities and research institutions are encouraged to promote international research collaboration, joint publications, and participation in global research networks related to corpus linguistics, digital literacy, and educational technology. Such initiatives would enhance the visibility and impact of Indonesian research at the international level.

Third, the strong linkage between literacy, digital technology, and pedagogy identified in the keyword co-occurrence and density analyses suggests that corpus linguistics can serve as a strategic tool for strengthening digital literacy in Indonesian classrooms. Policymakers and curriculum developers should consider corpus-based learning as part of national digital literacy programs, particularly in English language education and higher education contexts. This approach aligns well with Indonesia's ongoing digital transformation in education and the increasing emphasis on data-driven learning environments.

Finally, given the limited number of Indonesian-authored publications identified in this bibliometric dataset, future research should focus on local empirical studies examining corpus literacy among Indonesian teachers and students. Such studies would provide context-specific insights into challenges, readiness levels, and pedagogical opportunities, thereby bridging the gap between global research trends and local educational realities.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] B. F. Klimova, "Using corpus linguistics in the development of writing," *Procedia - Soc. Behav. Sci.*, vol. 141, pp. 124–128, 2014, doi: 10.1016/j.sbspro.2014.05.023.

[2] T. Cobb and A. Boulton, "Classroom applications of corpus analysis," in *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, U.K.: Cambridge University Press, 2015, ch. 26, pp. 478–497, doi: 10.1017/cbo9781139764377.027.

[3] G. R. Bennet, "An introduction to corpus linguistics," in *Using Corpora in Language Learning Classroom: Corpus Linguistics for Teachers*, pp. 1–315, 2010, doi: 10.4324/9781315843674.

[4] A. Ebrahimi and E. Faghih, "Integrating corpus linguistics into online language teacher education programs," *ReCALL*, vol. 29, no. 1, pp. 120–135, 2017, doi: 10.1017/S0958344016000070.

[5] J. Krajka, "Non-native teachers investigating new Englishes: Is data-driven teaching a part of 21 st century digital literacy?," *Aula Abierta. Revista de Investigación, Formación e Innovación en Educación (RIFIE)*, vol. 50, no. 2, e585, 2021, doi: 10.17811/rifie.50.2.2021.585-592.

[6] C. Girón-García and M. Esbrí-Blasco, "Analysing the digital world and its metaphoricity: Cybergenres and cybermetaphors in the 21st century," *Cult. Leng. y Represent.*, vol. 22, pp. 21–35, 2019, doi: 10.6035/CLR.2019.22.2.

[7] D. C. Mbandje, M. J. Loureiro, and M. Lucas, "Digital competence and information literacy: Clarifying concepts based on a literature review," *EMI - Educ. Media Int.*, vol. 60, no. 3–4, pp. 306–316, 2023, doi: 10.1080/09523987.2023.2324584.

[8] L. B. Barra, A. M. Fillola, and M. P. Ripoll, "Identidad lectora de profesores noveles y mediación literaria digital : Entrecruces entre trayectorias y competencias," *Educação e Pesquisa*, vol. 48, no. 51, e233911, 2022, doi: 10.1590/S1678-4634202248233911.

[9] I. Dussel and F. Williams, "Los imaginarios sociotécnicos de la política educativa digital en México (2012-2022)," *Profesorado*, vol. 27, no. 1, pp. 39–60, 2023, doi: 10.30827/profesorado.v27i1.26247.

[10] W. M. Tagata and F. C. Ribas, "Rethinking digital literacy practices and educational agendas in times of Covid-19 uncertainty (Repensando práticas de letramento digital e agendas educacionais em tempos de), *Rev. Bras. Linguist. Apl.*, vol. 21, no. 2, pp. 399-431, 2021, doi: 10.1590/1984-6398202117285.

[11] L. Lofthus and K. Silseth, "Students choosing digital sources: Studying students' information literacy in group work with tablets," *E-Learning and Digital Media*, 2019, doi: 10.1177/2042753019835882.

[12] S. Dhillon and N. Murray, "An investigation of EAP teachers' views and experiences of e-learning technology," *Educ. Sci.*, vol. 11, no. 2, e54, 2021, doi: 10.3390/educsci11020054.

[13] J. Romero-Barranco and P. Rodríguez-Abruñeiras, "Current trends in corpus linguistics and textual variation," *Res. Corpus Linguist.*, vol. 9, no. 2, pp. 1-13, 2021, doi: 10.32714/ricl.09.02.01.

[14] N. Suprapto *et al.*, "Research trend on TPACK through bibliometric analysis (2015-2019)," vol. 10, no. 4, pp. 1375–1385, 2021, doi: 10.11591/ijere.v10i4.22062.

[15] S. Thanuskodi, "Journal of social sciences: A bibliometric study," vol. 24, no. 2, pp. 77-80, 2010, doi: 10.1080/09718923.2010.11892847.

[16] Y. N. Deda, H. Disnawati, R. Ekawati, and N. Suprapto, "Research trend on dyscalculia by bibliometric analysis during 2017-2022," vol. 13, no. 1, pp. 69–79, 2024, doi: 10.11591/ijere.v13i1.25992.

[17] A. Pritchard, "Statistical bibliography or bibliometrics?" *Journal of Documentation*, vol. 25, no. 4, pp. 348–349, 1969.

[18] M. Parra-González, A. Segura-Robles, M. R. Vicente-Bújez, and J. López-Belmonte, (2020). "Production analysis and scientific mapping on active methodologies in Web of Science," *Int. J. Emerg. Technol. Learn. (IJET)*, vol. 15, no. 20, pp. 71-86, 2020. [Online]. Available: https://www.learntechlib.org/p/218320/.

[19] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, no. 4, pp. 959-975, 2017, doi: 10.1016/j.joi.2017.08.007.

[20] N. Roslim, M. H. T. Abdullah, N. F. M. Roshdan, Y. J. Ng, and S. A. R. Kalajahi, "Learner corpus research: A bibliometric analysis," *Int. Res. Educ.*, vol. 11, no. 2, pp. 76-93, 2023, doi:

10.5296/ire.v11i2.21396.
[21] R. C. Meurant, "The iPad and EFL digital literacy," in *Th. Kim, S. K. Pal, W. I. Grosky, N. Pissinou, T. K. Shih, and Ślęzak, D. (eds) Signal Processing and Multimedia MulGraB SIP 2010, Comm. Comput. Inf. Sci.* vol. 123, pp. 224–234, 2010, doi: 10.1007/978-3-642-17641-8_27.
[22] R. C. Meurant, "Developing critical L2 digital literacy through the use of computer-based internet-hosted learning management systems such as Moodle," in *Int. Conf. Multimedia, Comput. Graph., and Broad.*, 2009, pp. 76–83. doi: 10.1007/978-3-642-10512-8_9.
[23] R. C. Meurant, "The significance of second language digital literacy why English-language digital literacy skills should be fostered in Korea," in *ICCIT 2009 - 4th Int. Conf. Comput. Sci. Convergence Inf. Technol.*, 2009, pp. 369-374, doi: 10.1109/ICCIT.2009.192.
[24] R. C. Meurant, "Computer-based internet-hosted assessment of L2 literacy: Computerizing and administering of the Oxford Quick Placement Test in ExamView and Moodle," in *Int. Conf. Multimedia, Comput. Graph., and Broad.*, 2009, pp. 84-91, doi: 10.1007/978-3-642-10512-8_10.
[25] R. C. Meurant, "Mixed success on us.cyworld: Reflections on Korean EFL college student efforts to establish an L2 presence on an American social networking site," in *2008 Int. Conf. Convergence Hybrid Inf. Technol. (ICHIT)*, 2008, pp. 440–445. doi: 10.1109/ICHIT.2008.204.
[26] R. C. Meurant, "L2 digital literacy: Korean EFL students use their cell phone videocams to make an L2 English video guide to their college campus," in *Proc. the 2007 Int. Conf. Intelligent Pervasive Comput., (IPC)*, 2007, pp. 169–173. doi: 10.1109/IPC.2007.17.
[27] R. H. Jones, "Is my mobile phone listening to me? Conspiratorial thinking, digital literacies, and everyday encounters with surveillance," *Conspiracy Theo. Disc. John Benjamins Pub. Co.*, pp. 49–69, 2022, doi: 10.1075/dapsac.98.03jon.
[28] R. H. Jones, C. A. Hafner, *Understanding Digital Literacies: A Practical Introduction (2nd ed.)*. London, U.K.: Routledge, 2021, doi: 10.4324/9781003177647.
[29] R. H. Jones, C. A. Hafner, *Understanding Digital Literacies: A Practical Introduction (1st ed.)*. London, U.K.: Routledge, 2012, doi: 10.4324/9780203095317.
[30] R. H. Jones, "The text is reading you: Teaching language in the age of the algorithm," vol. 62, e100750, 2021, doi: 10.1016/j.linged.2019.100750.
[31] C. A. Hafner, "Digital discourses research and methods," in *A. Phakiti, P. De Costa, L. Plonsky, S. Starfield (eds) The Palgrave Handbook of Applied Linguistics Research Methodology*. London, U.K.: Palgrave Macmillan, 2018, pp. 375–390. doi: 10.1057/978-1-137-59900-1_18.
[32] S. Alves and A. P. de Sousa, "Cyberjournalism: The influence of design on public participation," in *N. Martins and D. Brandão (eds) Advances in Design and Digital Communication III*, *Int. Cof. Des. Dig. Commun.*. Cham, Switzerland: Springer Nature, 2023, pp. 479–494. doi: 10.1007/978-3-031-20364-0_42.
[33] S. Alves and A. P. de Sousa, "The impact of visual design on public participation. Case study in cyberjournalism," in *N. Martins, D. Brandão, A. Fernandes-Marcos (eds) Perspectives on Design and Digital Communication*, Cham, Switzerland: Springer Nature, 2024, pp. 33–52. doi: 10.1007/978-3-031-41770-2_3.
[34] L.-T. Zhang and D. Cassany, "The 'danmu' phenomenon and media participation: Intercultural understanding and language learning through 'The Ministry of Time,'" *Media Educ. Res. J.*, vol. 27, no. 58, pp. 19–29, 2019, doi: 10.3916/C58-2019-02.
[35] D. Cassany, "Recursos lingüísticos en línea: Contextos, prácticas y retos (Online linguistic resources: Contexts, practices and challenges," vol. 49, Suppl. 1, pp. 7–29, 2016, doi: 10.4067/S0718-09342016000400002.
[36] M. Kendrick, M. Early, A. Michalovich, and M. Mangat, "Digital storytelling with youth from refugee backgrounds: Possibilities for language and digital literacy learning," *TESOL Q.*, vol. 56, no. 3, pp. 961–984, 2022, doi: 10.1002/tesq.3146.
[37] M. E. Kendrick and M. Early, "Addressing the language and literacy needs and challenges of students with refugee experiences: Integrated supports," *Educ. Sci.*, vol. 14, no. 12, 2024, doi: 10.3390/educsci14121354.
[38] F. C. Komesu, R. R. Gambarato, and L. E. Tenani, "Acta Scientiarum ' I will not become an Internet meme ': visual-verbal textualization process in the study of the power and resistance in Brazil," vol. 40, no. 2, pp. 1–11, 2018, doi: 10.4025/actascilangcult.v40i2.43714.
[39] R. R. Gambarato and F. Komesu, "What are you laughing at? Former Brazilian President Dilma Rousseff's internet memes across spreadable media contexts," *J. Creat. Commun.*, vol. 13, no. 2, pp. 85–103, 2018, doi: 10.1177/0973258618761405.
[40] D. Gold and H. J.W, "Writing instruction in U.S. colleges and schools: The twentieth century and the new millennium," in *A Short History of Writing Instruction: From Ancient Greece to The Modern*

United States, U.S.: Taylor and Francis, 2020, pp. 272–316.
[41] M. Camargo and M. C. Woods, "Writing instruction in school and college english: The twentieth century and the new millennium," in *A Short History of Writing Instruction: From Ancient Greece to Contemporary America*, London, U.K.: Routledge, 2012, pp. 232–272. doi: 10.4324/9780203134368.
[42] A. M. Hautemo and T. Mushaandja, "The intersection between African oral literature and technology," in *2024 IST-Africa Conference, IST-Africa 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.23919/IST-Africa63983.2024.10569318.
[43] A. M. Hautemo, "Blogging as a Web 2.0 communication tool in an English and applied linguistics tertiary environment," in *2023 IST-Africa Conference, IST-Africa 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.23919/IST-Africa60249.2023.10187861.
[44] N. I. Almazova, A. V. Rubtsova, N. B. . Smolskaia, and A. A. Andreeva, "Digital creative projects in the formation of digital competence of teachers of English as a foreign language," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland, 2022. doi: 10.1007/978-3-030-89708-6_61.
[45] C. Yan, H. Wang, X. Luo, "Knowledge mapping of data literacy: A bibliometric study using visual analysis," *J. Librariansh. Inf. Sci.*, 2024, doi: 10.1177/09610006241285512.
[46] M. Chitez and R. Rogobete, "Beyond conventional digital approaches in the humanities: Corpus linguistics for language related university disciplines," *Innov. Lang. Learn. 13th Ed. Int. Conf.,* 2020. [Online]. Available: https://conference.pixel-online.net/library_scheda.php?id_abs=4853.
[47] P. H. Karlsen and M. Monsen, "Corpus literacy and applications in Norwegian upper secondary schools: Teacher and learner perspectives," vol. 19, no. 1, pp. 118–148, 2020, doi: 10.35360/njes.500.
[48] Ursula Wingate, "Academic literacy across the curriculum: Towards a collaborative instructional approach," *Lang. Teach.*, vol. 51, pp. 349–364, 2018, doi: 10.1017/S0261444816000264.
[49] K. Mills, "A review of the 'Digital Turn' in the new literacy studies," *Rev. Educ. Res.*, vol. 80, no. 2, pp. 246–271, 2010, doi: https://doi.org/10.3102/0034654310364401.
[50] R. A. Mora, "Challenges and opportunities for literacy and technology in ELT teacher education," *2nd Int. Colloquia Res. Innov. Foreign Lang. Educ. Universidad de La Salle Bogotá*, 2011. [Online]. Available: https://eric.ed.gov/?id=ED547648.
[51] T. M. G. Roberto, "Corpus linguistics in language teaching," *Revista Foco Interdiscip. Stud.,* vol. 16, no. 7, pp. 1–6, 2023, doi: 10.54751/revistafoco.v16n7-089.
[52] Q. Ma, M. M. Chiu, S. Lin, and N. B. Mendoza, "Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study," *ReCALL,* vol. 35, no. 1, pp. 19–39, 2023, doi: 10.1017/S0958344022000180.

Authors



Varia Virdania Virdaus is a doctoral candidate in Education Sciences at Universitas Negeri Surabaya, Indonesia. Her research interests include early childhood education, digital literacy, educational technology, and linguistics. She actively explores innovative pedagogical approaches integrating technology to enhance language development and learning outcomes. (email: 24011916010@mhs.unesa.ac.id).



Gunarti Dwi Lestari is a Professor and Coordinator of the Ph.D. Program in Education Sciences at Universitas Negeri Surabaya, Indonesia. Her expertise lies in non-formal education and deep learning pedagogy. She contributes extensively to academic leadership, curriculum innovation, and transformative educational practices in higher education. (email: gunartilestari@unesa.ac.id).



Wiwin Yulianingsih earned her doctoral degree from Universitas Negeri Malang, Indonesia. She is a lecturer at the Faculty of Education, Universitas Negeri Surabaya. Her scholarly work centers on non-formal education, community empowerment, and lifelong learning development through inclusive and sustainable educational programs. (email: wiwinyulianingsih@unesa.ac.id).



Rofik Djalal Rosyanafi received his doctoral degree from Universitas Negeri Malang, Indonesia. He serves as a lecturer at Universitas Negeri Surabaya. His research focuses on the development of non-formal education curricula, emphasizing innovation, community-based learning, and sustainable educational program design. (email: rofikrosyanafi@unesa.ac.id).



Muchamad Arif is affiliated with Narotama University, Surabaya, Indonesia. He frequently serves as a journal editor and trainer in English language programs. His research interests focus on AI-based learning media, educational innovation, and digital technology integration to enhance teaching and learning effectiveness. (email: muchamad.arif@narotama.ac.id).



Hamidah Binti Mohamad is Program Leader at Tun Razak Graduate School, Universiti Tun Abdul Razak, Malaysia. She holds a Ph.D. in Education from IIUM. Her specialization includes curriculum and instruction, English language and literature, and communication studies, with extensive experience in academic leadership. (email: hamidah_mohamad@unirazak.edu.my).