Applied Research in English Education
Vol.3 No.2 e-ISSN: 2962-1607
Applied Research on English Education (AREE)

# AI-Based Writing Assessment in Second Language Context: A Bibliometric Analysis of Trends in Coherence and Cohesion

**Sitti Kamila Meutia Sani**[1*]
**Viqi Ardaniah**[2]
[1,2]Universitas Airlangga, INDONESIA

**Abstract**
*The use of AI-based writing assessment in language education is growing rapidly, offering both opportunities and challenges in evaluating second language writing (L2). While these tools are increasingly common, most still focus on surface-level accuracy, such as grammar and spelling, and often overlook discourse-level elements like coherence and cohesion. These features are essential for L2 learners to produce texts with logical flow and organization. This study investigates research trends in AI-driven assessment of L2 writing from 2020 to 2025 through a bibliometric analysis of 332 articles retrieved from databases such as Scopus, JSTOR, ScienceDirect, and Emerald Insight. The analysis followed five key steps: setting objectives, selecting and cleaning data, performance analysis, keyword co-occurrence mapping, and interpretation. VOSviewer software was used to visualize thematic patterns and identify research gaps. Findings show a significant increase in AI-related research within language learning, yet coherence and cohesion remain underrepresented. Most studies emphasize technological aspects, such as system efficiency, tool development (e.g., ChatGPT), and AI innovation, rather than deeper writing features. The keyword analysis confirms this imbalance, highlighting a disconnect between research focus and educational priorities. This study calls for closer collaboration among educators, linguists, and AI developers to ensure assessment tools align with writing pedagogy. By mapping existing research trends and revealing neglected areas, this study contributes to improving AI-based writing assessment tools to better support meaning-making, structure, and clarity in second language writing.*

**Keywords:** AI-based writing assessment; Second language writing; Coherence and cohesion; Bibliometric analysis; Research trends

## 1. INTRODUCTION

Writing assessment is important for language teaching and learning, as it provides a means for measuring student progress, diagnosing problems, and

[1*]Corresponding author, email: sittikamilasani@gmail.com

evaluating instructional effectiveness (Hyland, 2003). An effective assessment must go beyond surface-level features such as grammar and spelling to incorporate higher-level aspects like coherence, cohesion, organization, and content development (Hyland, 2003; Zhang & Umeanowai, 2025). In second language (L2) contexts, writing assessment requires attention not only to linguistic correctness but also to the effective organization and development of ideas according to academic and rhetorical standards (Hyland, 2003). Coherence and cohesion are crucial to writing quality, particularly in L2 education where clear organization and logical flow are essential (Zhang & Umeanowai, 2025). Despite their importance, current AI-based writing assessment technologies often focus on surface-level features such as grammar and spelling, while deeper discourse aspects like coherence and cohesion remain unexplored (Lubis et al., 2024).

AI-based writing assessment can be understood as the application of artificial intelligence technologies, such as Automated Writing Evaluation (AWE), to assess and provide feedback on students' written texts (Wang & Zhu, 2025). These systems offer several advantages, including immediate feedback, reduced grading time, and individualized learning support (Alhusaiyan, 2025; Wang & Zhu, 2025). However, most AI-based writing assessment tools remain focused on surface-level features such as grammar, vocabulary, and sentence structure, while deeper discourse-level features, particularly coherence and cohesion, are often overlooked (Alhusaiyan, 2025; Lubis et al., 2024; Wang & Zhu, 2025). This is a critical issue because coherence and cohesion are essential components of writing quality, especially in second language (L2) learning contexts, where clear organization and logical flow of ideas are vital for effective communication. Despite their importance, these features have not been central themes in AI-based writing assessment research. To better understand how research in this area has evolved and to identify gaps related to discourse-level writing features, this study conducts a bibliometric analysis of recent literature.

Previous studies have examined different areas of AI-based writing assessment and the use of AI in language education. One common finding is that these tools often overlook important discourse features and fail to assess deeper textual qualities such as coherence and cohesion (Jiang, 2022; Wang & Zhu, 2025; Zhang & Umeanowai, 2025). Another issue is the pedagogical limitation of AI technologies, as they still require teacher intervention to address aspects that AI systems cannot fully capture (Alhusaiyan, 2025). In addition, several studies have documented the rapid growth of AI applications in education, particularly highlighting the surge in research publications following the introduction of tools like ChatGPT and the expansion of automated feedback systems (Farhat et al., 2024; Lubis et al., 2024; Wahyuni et al., 2024). Király (2024) further explored the role of AI as a virtual language teacher, focusing on its potential for delivering real-time personalized feedback to learners.

To better understand the development of research in this field, several bibliometric analyses have been conducted. Hinojo-Lucena et al. (2019) examined the application of AI in higher education, while Jaleniauskienė et al. (2023) analyzed trends in AI use within language education. Susilowati et al. (2024) extended this research by exploring the intersection between AI, education, and pedagogy, providing a broader perspective on research development in this domain. Although research on AI in education has grown, coherence and cohesion are still rarely explored in AI-based writing assessment, especially for second language (L2) learners. There is also no clear agreement on the main scopes and trends in this area. To fill this gap, this study uses a bibliometric analysis to map recent research trends and identify key topics in AI-based writing assessment for L2 contexts.

Therefore, this study aims to explore how research on AI-based writing assessment has developed in recent years, with a specific focus on coherence and cohesion in L2 contexts. By using a bibliometric approach, this study seeks to identify major research themes, publication trends, and potential gaps in the literature. The findings are expected to provide a clearer picture of how current studies address discourse-level writing features and offer insights for improving future AI-based assessment tools.

## 2. LITERATURE REVIEW

This section reviews current studies related to the use of artificial intelligence in language education, with a particular focus on writing assessment for second language (L2) learners. It covers major findings, theoretical perspectives, and research gaps relevant to the present study.

### 2.1. AI in Language Education

The use of Artificial Intelligence (AI) in language education has grown quickly in recent years. Many studies show that AI helps improve learning, especially in second language (L2) writing. As Wang and Zhu (2025) mentioned, “AI is commonly utilized for automated feedback and assessment, supporting immediate responses to learner input while alleviating the workload of language teachers.” Tools like natural language processing (NLP) and automated writing systems are now widely used to support students.

Zhang and Umeanowai (2025) also explain that “AI’s role in natural language processing (NLP), adaptive learning systems, and personalized feedback is now indispensable in modern language education.” Learners are using tools such as ChatGPT, Grammarly, and QuillBot to improve their writing. However, even though these tools are useful, they still do not fully support important parts of writing, like organizing ideas and creating a smooth flow between sentences.

#### *2.1.1 Challenges in AI-Based Writing Assessment*

While AI tools are useful for checking grammar and spelling, many still ignore deeper aspects of writing, like coherence and cohesion. These features are important, especially for L2 learners who need to organize their ideas clearly. Studies show that most AI writing systems are still limited in this area (Jiang, 2022; Lubis et al., 2024; Wang & Zhu, 2025).

Zhang and Umeanowai (2025) note that “persistent challenges persist [...] emphasizing the need for ongoing technical advancements and the comprehensive utilization of AI’s potential by educators in assessment practices.” This means that even though AI is improving, it still does not fully meet the needs of writing assessment. Wang and Zhu (2025) also state:

> “So far, review studies in AI-assisted language learning have not well incorporated the latest publications within the past few years. This gap calls for an update for review studies to reflect the most current trends and research issues in the landscape of AI in language education.”

Even though AI can give quick feedback, it still needs help from teachers to address complex parts of writing, such as idea development and flow (Alhusaiyan,

2025). Future research should focus more on how AI tools can support deeper aspects of writing, not just grammar and vocabulary.

Another important development in language education is the rise of ChatGPT as a widely discussed tool in academic research. As Farhat et al. (2024) highlights, “In just six months since its launch, ChatGPT has generated more than 500 scholarly publications, making it one of the most rapidly studied technologies in recent academic history.” This shows how quickly AI tools have gained attention in education. The study also found strong global collaboration among researchers interested in how ChatGPT can support learning and communication in different fields, including language learning.

Significant point comes from Susilowati et al. (2024), who emphasized the need for stronger links between AI and pedagogy in education. They found that although many studies focus on the technical side of AI, fewer actually explore how it fits with teaching practices and learning values. As they noted, “approaching the integration of artificial intelligence with education must be from a strong pedagogical approach; not only does the algorithm have to be right, but also the emotions and values that are appropriate are needed” (Susilowati et al., 2024). This reminds us that AI tools in education should not only be functional but also meaningful in helping teachers and students.

Hereby, despite of AI's promising peformance in languange education, especially in supporting second language writing, current studies unveil that its use in writing assessment remains limited in adressing deeper aspects of discourse. AI tools like ChatGPT and Grammarly offer fast and accessible support, but they often overlook important elements such as coherence, cohesion, and pedagogical fit. The literature also highlights the growing interest in AI research, along with calls for better integration between technology and teaching practices. These findings underline the need for more balanced development, where AI not only improves technical accuracy but also aligns with educational values and supports meaningful writing development.

Overall, the studies reviewed suggest that AI is playing a bigger role in language education, particularly in helping students with their writing. Most research has focused on how the technology works and what it can do, but there’s also growing interest in how AI fits into actual classroom practice. As these tools keep developing and becoming more common, it's worth paying attention to how they’re being used in writing instruction and what directions future research might take.

## 3. METHODS

This study used a bibliometric approach to examine research trends in AI-based writing assessment within second language (L2) contexts, focusing on publications from 2020 to 2025. Bibliometric analysis helps map the structure and trends of research fields using quantitative techniques (Donthu et al., 2021). It offers a clear way to organize large amounts of data and find research gaps. Bibliometric mapping combines performance analysis, which looks at the productivity of authors, institutions, and countries, with science mapping, which studies relationships between documents, authors, journals, and keywords (Zupic & Čater, 2015). This study followed the five-step framework from Donthu et al. (2021), which includes setting objectives, choosing data sources, collecting and cleaning data, analyzing, and interpreting results.

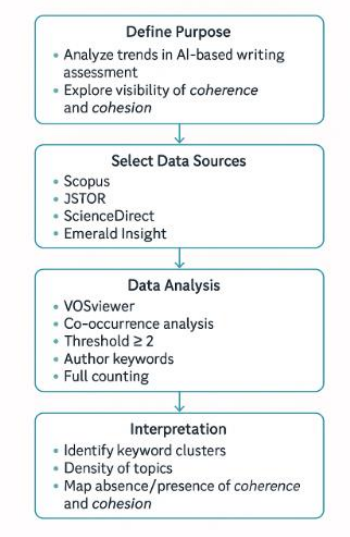


Figure 1: Research framework for bibliometric analysis. Source: Adapted from Donthu et al. (2021).

Data were collected from four academic databases: Scopus, JSTOR, ScienceDirect, and Emerald Insight, chosen for their broad coverage of research articles. The search was using keywords like “AI writing assessment,” “Second language writing assessment,” “L2 writing,”, and “Coherence and cohesion”. ” These keywords were selected because they appear frequently in recent studies on AI-based writing (Wang & Zhu, 2025; Zhang & Umeanowai, 2025). The search focused on articles published in English between 2020 and 2025 to capture the rise of AI research following the introduction of tools like ChatGPT (Farhat et al., 2024; Wahyuni et al., 2024). English was used because it is the main language of academic publishing (Donthu et al., 2021). After removing duplicates and cleaning missing fields, 332 articles were kept. Keyword co-occurrence analysis was done using VOSviewer (version 1.6.20), a tool for creating and visualizing bibliometric maps (Van Eck & Waltman, 2010). The analysis used author keywords with full counting and included keywords that appeared at least twice. The resulting network grouped keywords into clusters, showing the main research themes in the field.

## 4. RESULTS AND DISCUSSION


Figure 2: Keyword co-occurrence network generated from AI-based writing assessment literature. Source: Generated by the author using VOSviewer (version 1.6.20).

The keyword co-occurrence analysis showed that current research in AI-based writing assessment focuses on a few key areas. Keywords like “chatgpt,” “writing assessment,” and “second language writing” appeared most frequently, showing that generative AI tools, especially ChatGPT, are now central in supporting writing evaluation for L2 learners.

In addition to this focus on generative AI, the analysis also found that many studies concentrate on technical aspects, with terms like “natural language processing,” “machine learning,” and “large language models” appearing often. This indicates that research is more concerned with developing the underlying technology rather than improving how AI can assess writing quality. Although terms like “artificial intelligence” and “second language writing” were linked, the connection to writing pedagogy remains weak, suggesting limited integration between technology and education (Alhusaiyan, 2025; Wang & Zhu, 2025).

While the field appears active, with 15 links and a total link strength of 17, some critical features are notably missing. Most importantly, the keywords “coherence” and “cohesion” did not appear in the network. This is significant because these features are central to good writing, especially for L2 learners (Hyland, 2003). Their absence shows that current research has not yet given enough attention to deeper writing skills, focusing instead on surface-level language features. This finding aligns with previous studies highlighting that current AI-based writing assessment systems are effective at surface-level error detection but limited in evaluating discourse-level features such as coherence and cohesion (Jiang, 2022; Lubis et al., 2024; Wang & Zhu, 2025).

Overall, these findings suggest that although AI-based writing assessment is expanding, there is a clear gap between technology development and what is needed in writing education. Future research should focus more on including discourse-level features like coherence and cohesion to better support L2 writing assessment and align technological advances with pedagogical goals.


Figure 3: Overlay visualization of keyword co-occurrence network. Source: Generated by the author using VOSviewer (version 1.6.20).

The overlay visualization provides further insights into the temporal trends of the field. Most keywords had average publication years clustered around 2023 and early 2024, indicating that interest in AI-based writing assessment has surged recently, particularly after the public release of ChatGPT. The lack of temporal separation among keywords suggests that the research area is developing rapidly but remains in an early, condensed stage of evolution. This pattern is consistent with the characteristics of emerging research fields, where growth is initially concentrated around a few dominant topics before diversifying (Zupic & Čater, 2015; Farhat et al., 2024).

The density visualization revealed differences in the prominence of keywords based on their frequency and co-occurrence strength. Keywords such as “writing assessment,” “second language writing,” “chatgpt,” and “artificial intelligence” appeared in the densest and most illuminated regions of the map, reflecting their dominance in the current research discourse. In contrast, technical terms like “machine learning” and “large language models” were located towards the periphery, suggesting that while technical developments are foundational, they are not the central focus of most studies (Van Eck & Waltman, 2010).

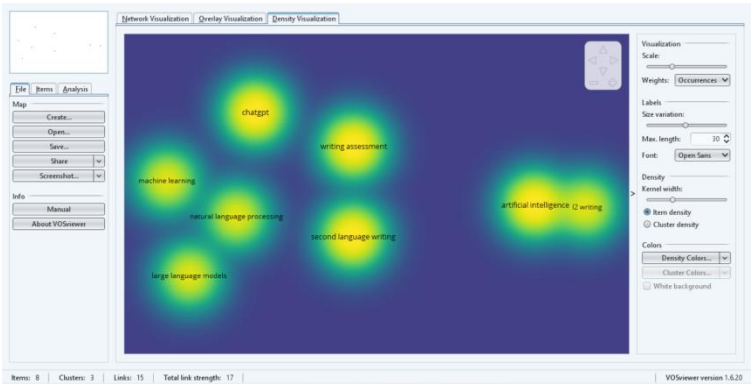


Figure 4: Density visualization of keyword co-occurrence network. Source: Generated by the author using VOSviewer (version 1.6.20)

The absence of “coherence” and “cohesion” from both the network and density visualizations highlights a key gap between technological focus and pedagogical needs. Coherence and cohesion are essential components of writing quality, particularly for L2 learners who often struggle with text organization and logical flow.

Their absence indicates that current AI-based assessment tools may prioritize surface-level language features, such as grammar and vocabulary (Alhusaiyan, 2025; Wang & Zhu, 2025), over deeper discourse-level attributes that are critical for meaningful writing assessment.

Overall, the findings point to a growing but technologically-driven research landscape in AI-based writing assessment. While advancements in AI tools are significant, the lack of attention to discourse-level writing features signals the need for future research to bridge the gap between technological capabilities and educational priorities (Wang & Zhu, 2025; Lubis et al., 2024). This gap is particularly important in L2 contexts, where the ability to produce coherent and cohesive texts is fundamental to writing competence (Hyland, 2003).

Greater collaboration between AI developers, applied linguists, and language educators will be necessary to ensure that future AI writing assessment tools address both mechanical accuracy and the deeper rhetorical and organizational aspects of writing. The difficulty of automatically evaluating coherence and cohesion may explain their absence in current AI research. Unlike grammar or vocabulary, these features require understanding the logical flow and semantic relationships across the text, which remain challenging for AI systems.

This concentration reflects the rapid expansion of research interest in AI-based writing assessment following the release of ChatGPT, but also highlights that this area remains in its early developmental stage. These findings not only highlight the current limitations of AI-based writing assessment but also provide important directions for future research and development efforts, which are discussed in the next section.

## 5. CONCLUSION

This study used a bibliometric approach to map research trends in AI-based writing assessment for second language (L2) learners. The analysis found three main clusters: one focused on writing evaluation with generative AI tools like ChatGPT, one on foundational AI technologies such as natural language processing and machine learning, and one on the broader connection between artificial intelligence and L2 writing. These results show that current research mostly centers on technological tools and system development.

However, important writing features like coherence and cohesion were missing from the keyword network and density maps. This is important because these features are central to writing quality but are not yet a major focus in AI-based writing assessment research. The overlay visualization also showed that the field is growing quickly, especially after the release of ChatGPT, but the research themes remain concentrated around technical areas.

These findings highlight a gap between what current AI research focuses on and what writing pedagogy values. Future research should pay more attention to discourse-level features like coherence and cohesion. To achieve better AI writing assessment tools, there needs to be more collaboration between AI developers, applied linguists, and language educators, ensuring that future systems support both technical accuracy and the deeper qualities that make writing effective.

## REFERENCES

Alhusaiyan, E. (2025). A systematic review of current trends in artificial intelligence in foreign language learning. *Saudi Journal of Language Studies, 5*(1), 1–16. https://doi.org/10.1108/SJLS-07-2024-0039

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021).

How To Conduct A Bibliometric Analysis: An Overview And Guidelines. Journal Of Business Research, 133, 285–296. https://doi.org/10.1016/J.Jbusres.2021.04.070

Farhat, F., Silva, E. S., Hassani, H., Madsen, D. Ø., Sohail, S. S., Himeur, Y., Alam, M. A., & Zafar, A. (2024). The scholarly footprint of ChatGPT: A bibliometric analysis of the early outbreak phase. Frontiers in Artificial Intelligence, 6, Article 1270749. https://doi.org/10.3389/frai.2023.1270749

Hinojo-Lucena, F.-J., Aznar-Díaz, I., Cáceres-Reche, M.-P., & Romero-Rodríguez, J.-M. (2019). Artificial Intelligence In Higher Education: A Bibliometric Study On Its Impact In The Scientific Literature. Education Sciences, 9(1), 51. https://doi.org/10.3390/Educsci9010051

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Jaleniauskienė, E., Lisaitė, D., & Daniusevičiūtė-Brazaitė, L. (2023). Artificial Intelligence In Language Education: A Bibliometric Analysis. Sustainable Multilingualism, 23(1), 159–194. https://doi.org/10.2478/Sm-2023-0017

Jiang, R. (2022). How Does Artificial Intelligence Empower Efl Teaching And Learning Nowadays? A Review On Artificial Intelligence In The Efl Context. Frontiers In Psychology, 13. https://doi.org/10.3389/Fpsyg.2022.1049401

Király, L. (2024). Artificial Intelligence As A Virtual Language Teacher. Magyar Nyelvőr, 148(5), 649–665. https://doi.org/10.38143/Nyr.2024.5.649

Lubis, A. H., Samsudin, D., Triarisanti, R., Jerusalem, M. I., & Hwang, Y. (2024) . A Bibliometric Mapping Analysis Of Publications On The Utilization Of Artificial Intelligence Technology In Language Learning. Journal Of Advanced Research In Applied Sciences And Engineering Technology, 38(1), 156–176. https://doi.org/10.37934/Araset.38.1.156176

Susilowati, D., Saputri, D. S. C., Ria, R. R. P., & Hidjah, K. (2024). Bibliometric Analysis Of Artificial Intelligent Studies In Education And Pedagogy. Nusantara Jurnal Pendidikan Indonesia, 4(3), 621–634. https://doi.org/10.14421/Njpi.2024.V4i3-1

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

Wahyuni, S., Putro, N. H. P. S., & Efendi, A. (2024). Trends In Artificial Intelligence-Infused English Language Learning: A Comprehensive Bibliometric And Content Review. Advanced Education, 12(25), 162–178. Https://doi.org/10.20535/2410-8286.315035

Wang, W., & Zhu, C. (2025). *A systematic review of AI in language education: Current trends and future directions*. Language Testing, 40(2), 175–200. https://doi.org/10.1177/02655322221133844

Zhang, X., & Umeanowai, K. O. (2024). Exploring The Transformative Influence Of Artificial Intelligence In EFL Context: A Comprehensive Bibliometric Analysis. Education And Information Technologies. Https://doi.org/10.1007/S10639-024-12937-Z

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629