Mapping the Landscape of AI-Assisted L2 Writing Assessment: A Bibliometric and Trend-Forecasting Study (2021–2025)

Nurgül Bekdemir [a] *, Kadir Kesgin [b]

[a] Ordu University, Ordu, 52200, Türkiye
[b] Bandırma Onyedi Eylül University, Balıkesir, 10200, Türkiye

Abstract

Language assessment has been undergoing a significant transformation with the rise of Artificial Intelligence (AI), particularly in writing, which is widely recognized as the most complex of the four language skills due to its multidimensional nature. Recent studies highlight how AI tools—such as ChatGPT—support feedback generation, streamline editing, and reduce teachers' assessment workload. Despite the increasing volume of publications in AI-based writing assessment, there remains a lack of bibliometric studies that map the scholarly landscape of this emerging field. This study addresses that gap by conducting a bibliometric and predictive analysis of AI-based L2 writing assessment research published between 2021 and 2025. Data were retrieved from ScienceDirect, Wiley Online Library, SpringerLink, and SAGE Journals and analyzed using the BiBLoX platform for trend forecasting, topic modeling (LDA), and co-authorship network mapping. Citation predictions were modeled using machine learning algorithms, including Random Forest Regression. The results reveal evolving thematic focuses, leading contributors, and publication trends, offering a data-driven overview of the field and highlighting directions for future research.

Keywords: Artificial intelligence; L2 assessment; writing; bibliometric analysis; genAI

1. Introduction

Language assessment plays a critical role in language education, particularly in ensuring reliability and validity. Among the four language skills, writing is widely recognized as the most complex, especially in a second language (L2) context, due to its multidimensional nature. It involves evaluating various components such as accuracy, cohesion, content, organization, and language use (Williams & Beam, 2019). This

---

* Corresponding author name. ORCID ID.: https://orcid.org/ 0000-0002-9107-2653
  E-mail address: nurgulbekdemir@odu.edu.tr

complexity poses challenges for both teachers and learners; however, advancements in educational technology have made writing assessment more systematic and transparent.

Recent developments in Artificial Intelligence (AI), particularly through Large Language Models (LLMs), have significantly influenced language assessment practices. AI tools are now being utilized for providing automated feedback, evaluating written productions based on multiple criteria, and reducing teachers' assessment workload. These tools also offer learners opportunities to revise their work and receive instant feedback, fostering deeper learning. Studies have shown that such tools support both language acquisition and assessment processes (Li, 2024; Poole, 2024; Wang, 2024). Moreover, experimental research has begun to examine the reliability and validity of AI-assisted L2 writing assessment (Chen et al., 2025; Poole, 2024; Song & Tang, 2025).

Despite the growing body of research in AI-based writing assessment, there is a notable lack of bibliometric studies that systematically map this emerging field. Existing literature lacks a data-driven framework that reveals publication trends, citation networks, and key contributors. Bibliometric analyses have the potential to uncover research patterns, influential scholars, and thematic gaps in the literature. To address this need, the present study conducts a bibliometric analysis of AI-based L2 writing assessment research published between 2021 and 2025, aiming to provide a comprehensive overview of the current scholarly landscape.

Language assessment is a critical issue in language teaching procedure, especially in a reliable and valid way. Assessment of writing, which is one of the productive skills, is considered being a complex skill in a second language as there are a number of elements to be assessed such as accuracy, cohesion, content, organization and language use (Williams & Beam, 2019). Thus, it is a challenging process for both teachers and students, however, the opportunities that technology has provided lead writing assessment to be more systematic and transparent.

With the recent improvements in artificial intelligence (AI), language assessment has begun being transformed. AI tools based on Large Language Models (LLM) are used in various ways, i.e., automated feedback on written productions, assessment of students' written productions in terms of the required competences and minimizing teachers' workload in assessment. Moreover, students have opportunities to edit their productions and learn from these tools through instant feedback. Thus, the recent studies indicate that AI provides benefits regarding language learning and assessment (Li, 2024; Poole, 2024; Wang, 2024). In addition to the studies investigating how students benefited from these tools (Guo, 2024; Wang, 2024), the experimental studies aiming to reveal the reliability and validity of L2 assessment with these AI tools have been taking attention (Chen et al., 2025; Poole, 2024; Song & Tang, 2025).

Given the current literature on AI-based writing assessment, there is a gap in the literature that investigates the current issue from the perspective of bibliometric

analysis. It is indicated that there is a lack of data-driven framework that comprises the general picture of AI-based writing assessment, development trend and citation network. Otherwise, bibliometric analysis studies reveal not only academic trends, prolific authors and quality publications but also the focus of the literature and the gaps in the literature. Aiming to fill the gap in the literature, the current study explores the studies on AI-based writing assessment through bibliometric analysis and maps the current status of the literature.

2. Literature Review

AI technologies that have transformed the language learning process have also assumed significant roles in language assessment (Chen et al., 2020; Hwang et al., 2020; Xu & Ouyang, 2022). Recent review and meta-analysis studies emphasize the growing need for research into AI-based language assessment (Chen et al., 2024). In particular, Chen et al. (2024) found that AI tools have a moderately positive effect on writing skills. However, most existing AI tools are rule-based systems tested only within short-term, classroom-specific contexts, limiting the generalizability of their results. Lin et al. (2022) argue that pre-trained models do not necessarily guarantee effective performance without consideration of learner-specific variables such as language proficiency, competencies, and instructional context.

One prominent approach in AI-driven assessment is Automated Writing Evaluation (AWE), which offers fast and objective analysis of student writing. These tools typically assess surface-level features such as word count, syntactic structure, and language accuracy (Cheng et al., 2023). However, deeper aspects of writing—such as textual cohesion, argumentative structure, and meta-discourse markers (MDMs) that reflect writer stance and rhetorical organization—are often overlooked (Hyland, 2005). As a result, while AWE tools may effectively evaluate form, they fall short in assessing the communicative intent and content quality of writing (Chan et al., 2024).

Another major contribution of AI to writing assessment lies in its capacity to deliver instant feedback, which has been shown to positively affect learners' writing performance, self-regulation, and attitudes (Nazari et al., 2021). For example, Osawa (2024) demonstrated that Notion AI helped students manage their writing process with improved structure and awareness. Similarly, Guo et al. (2024) found that AI-supported peer feedback significantly enhanced students' feedback literacy and revision practices. In addition to peer interaction, AI itself can act as a collaborative agent in writing. Li (2024) noted that ChatGPT played a supportive role across all stages of the writing process, while Wang (2024) reported that learners used it to balance motivational and emotional challenges.

Nonetheless, the limitations of current AI systems persist—particularly their inability to reliably assess discourse-level competencies. Chan et al. (2024) argued that elements

like metadiscourse should be integrated into assessment frameworks, while Won et al. (2025) found that GPT-2 failed to replicate L2 students' linguistic patterns, raising concerns about fairness and validity in assessment.

Beyond assessment, the integration of AI into writing instruction also supports learners' metacognitive development. Hartwell and Aull (2023) observed that AI-driven tools not only assist in editing and revision but also promote greater awareness of writing as a recursive and interactive process. Wang (2024) and Su et al. (2023) highlighted how tools like ChatGPT guide learners in generating ideas, organizing content, and improving discourse coherence—aligning with the iterative nature of writing emphasized in genre-based frameworks (Bhatia, 2014).

However, the literature reflects a disproportionate focus on English as the target language, with Less Commonly Taught Languages (LCTLs) remaining underrepresented. This linguistic imbalance stems from the dominance of English in technological development and the lack of multilingual training data in AI models. Won et al. (2025) stressed the need to train Large Language Models (LLMs) on diverse learner data to ensure equitable assessment across linguistic and cultural contexts.

While automated feedback and AI-supported writing environments offer innovative pathways for assessment, the field still lacks longitudinal research, cross-contextual comparisons, and studies targeting disadvantaged learners (Chen et al., 2024; Winke & Koné, 2025). A more holistic and context-sensitive design of AI-based assessment tools is needed—one that integrates discourse-level skills, supports teacher and learner AI literacy, and aligns with pedagogical goals (Godwin-Jones, 2024; Guo et al., 2024; Chan et al., 2024).

AI tools that have been transforming the language learning process also have significant roles in language assessment (Chen et al., 2020; Hwang et al., 2020; Xu & Ouyang, 2022). Review and meta-analysis studies demonstrated the requirement of research in AI-based language assessment (Chen et al., 2024). Chen et al. (2024) revealed AI tools have a moderately positive effect, especially on writing skills. Most AI tools are designed as systems with rule-based algorithms and only tested as classroom-based assessments in a limited time; however, it may not be inferred that pre-trained systems have to result in good performance (Lin et al., 2022). A more effective assessment process could be achieved with the help of factors such as students' proficiency level, competencies, and teaching context. Also, automated writing evaluation (AWE) systems, AI tools designed for reliable, fast, and objective assessment, evaluate learners' written productions through several superficial elements like word count, sentence structure, and language accuracy (Cheng et al., 2023). However, some significant elements in writing, such as cohesion in the text, argumentation, and meta discourse markers (MDM)—i.e., organization of discourse and writer's perspective—are neglected (Hyland, 2005). Consequently, it could be said that the form of the production could be successfully

assessed; however, the content and communicative function in writing require development (Chan et al., 2024).

The other contribution of AI to language assessment is the transformation of the feedback process. AI tools that provide instant feedback positively influence the writing competence, self-regulation, and attitudes of students (Nazari et al., 2021). It was exemplified in Osawa's study (2024) that students could manage their writing process with Notion AI at a higher and structured level of awareness. Additionally, the quality of students' comment and feedback literacy could be increased by AI-supported peer feedback (Guo et al., 2024). In addition to the peer feedback, cooperation between learners and AI during written production comes forward with supportive methods. Li (2024) shows that ChatGPT has a supportive role in all the writing process, while Wang (2024) supported that students try to balance motivation and learning anxiety. It was also revealed that the current AI systems were not sufficient to assess the discourse competence, and structural elements like meta-discourse markers should be considered (Chan et al., 2024). In Won et al.' s study (2025), it was revealed that GPT-2 cannot simulate the language patterns of L2 students, and it creates a risk for fairness in assessment.

The integration of AI tools into the writing process was about not only evaluation and assessment of the production but also improving learners' awareness towards writing skills and enhancing their metacognitive skills during the process (Hartwell & Aull, 2023). To exemplify, AI-tools like ChatGPT were conducted during different phases of the writing process, which led the learners' writing behaviors to transform (Wang, 2024). They provided learners with the information on essay type, suggestions for editing and developing discourse awareness (Su et al., 2023), which corresponds to the repetitive and interactive nature of the writing process (Bhatia, 2014; Wang, 2024).

As for linguistic diversity and cultural context, the studies on Less Commonly Taught Languages (LCTL) were limited. These languages were underrepresented in the literature and English is the subject in technology-based studies due to being lingua franca. It depends on the fact that AI tools cannot reflect multilingual and cultural differences. Similarly, LLMs like GPT should be trained with the data of the multilingual learners in order to have fair results (Won et al., 2025).

Automated assessment, instant feedback, and learning during the writing process are the opportunities provided to create new language learning and assessment environments. There are some studies focusing on short-term effects; however, there are limited studies with long-term effects, cross-contextual comparisons, and process-based assessment (Chen et al., 2024). Also, there is a lack of studies on disadvantaged students and effective models in LCTL contexts (Winke & Koné, 2025). It could be suggested that studies investigate a more holistic and contextual design of AI-based language assessment tools, the increase in AI literacy of learners and teachers and the

development of higher-level language skills like discourse (Godwin-Jones, 2024; Guo et al., 2024; Chan et al., 2024).

3. Method

This study adopts a predictive bibliometric analysis to explore the research landscape of AI-based writing assessment in L2 education by utilizing the BiBLoX platform, which provides real-time data integration, trend prediction, and scientific mapping. As a next-generation bibliometric tool, BiBLoX enables dynamic data retrieval from multiple databases, including Web of Science, Scopus, and TRDizin, and processes the data through automated pipelines for analysis without requiring manual updates (Kesgin & Zeren Özer, 2025).

The data collection process was initiated by defining a set of relevant keywords such as "AI writing assessment," "ChatGPT," "generative AI," "automated feedback," and "L2 writing." The search was limited to peer-reviewed journal articles published between 2021 and 2025, and only English-language publications were included. All retrieved entries were automatically filtered for duplicates using both DOI matching and fuzzy similarity measures (Levenshtein distance ≥ 90%), ensuring the consistency and accuracy of the dataset.

For citation forecasting, early citation signals were extracted and operationalized as the number of citations received within the first six months after publication, serving as a critical indicator of initial impact. These values, along with other variables such as journal impact factor, author h-index, number of co-authors, and keyword presence in the title, were used as input features in predictive modeling. Three machine learning models—Linear Regression, Support Vector Regression (with RBF kernel), and Random Forest Regression—were applied using the scikit-learn framework. The model evaluation was conducted through Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2). Among the models, Random Forest yielded the most robust predictions (RMSE = 4.16; R^2 = 0.73), indicating its effectiveness in modeling short-term citation behavior.

To explore thematic trends, Latent Dirichlet Allocation (LDA) was used for topic modeling, where the optimal number of topics was determined through coherence score analysis. After pre-processing the abstracts via tokenization, lemmatization, and stopword removal, the coherence metric (c_v) suggested that seven topics provided the best balance between interpretability and semantic distinction (coherence score = 0.498). Topic distributions were then visualized over time to demonstrate thematic evolution in the field.

In the analysis of collaborative patterns, a co-authorship network was generated using NetworkX. Authors were treated as nodes and joint publications as undirected edges. To

ensure relevance and reduce noise in the network, only those author pairs with at least two shared publications were included in the graph. Network centrality and modularity measures were computed to identify influential contributors and collaborative clusters in the dataset.

All visualizations—including publication trends, citation trajectories, keyword co-occurrence maps, and topic evolution charts—were generated through BiBLoX's interactive dashboard using Python-based backends and JavaScript libraries (Plotly, D3.js). These interactive outputs enabled a nuanced understanding of both structural and temporal dynamics within the AI-based writing assessment literature.

This methodological framework, grounded in predictive bibliometrics, offers a comprehensive and replicable approach for analyzing scholarly trends, while also identifying emerging themes and research gaps within a rapidly developing interdisciplinary field.

4. Results

4.1. Publication and Citation Trends

The results for annual publication output in a five-year period reveals significant trends in research productivity in L2 assessment through AI. The results demonstrate the annual publication count from 2021 to 2025, resulting in a clear upward trend in scholarly publications over the observed years (Figure 1).
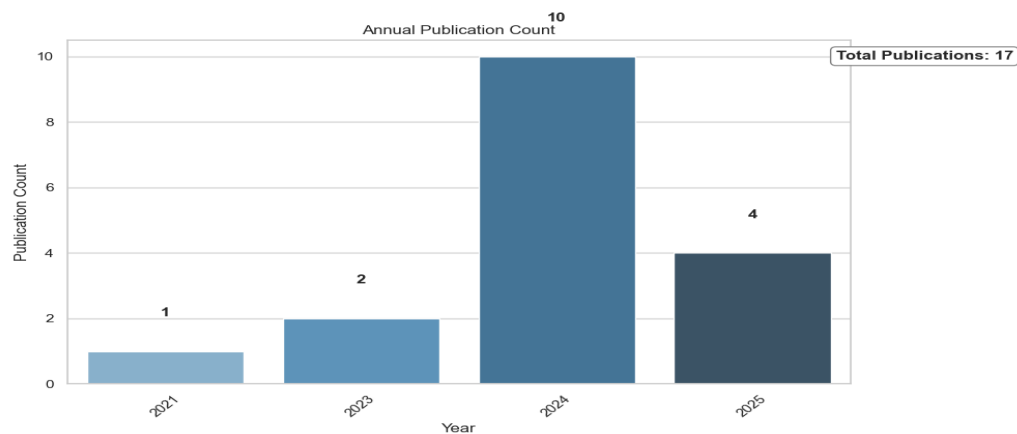

Figure 1. Annual Publication Count

The number of publications was only 1 in 2021, however, it increased slightly to 2 in 2023, before reaching a peak of 10 publications in 2024. Although there was a subsequent decrease to 4 publications in 2025, a significant growth in annual publication could be seen. In total, the number of publications equals to 17 in the five-year period, and 2024

was the year with the most publications. This pattern may demonstrate that L2 assessment through AI started to gain scholars' attention as AI has been increasingly used.

The annual citation trends reveal a fluctuating pattern in scholarly impact over the five-year period (Figure 2; Figure 3).
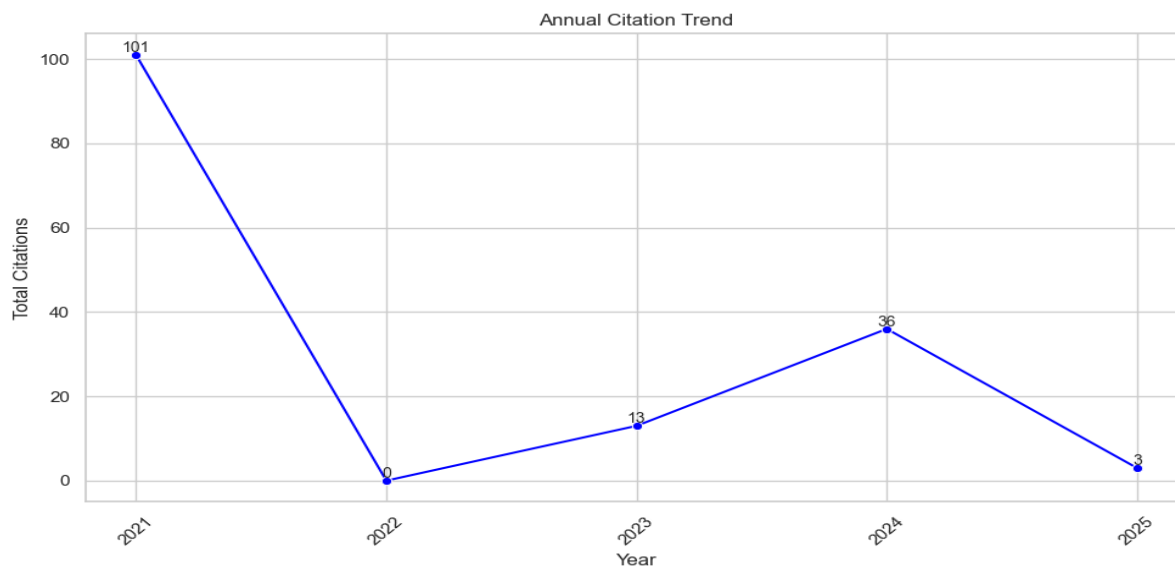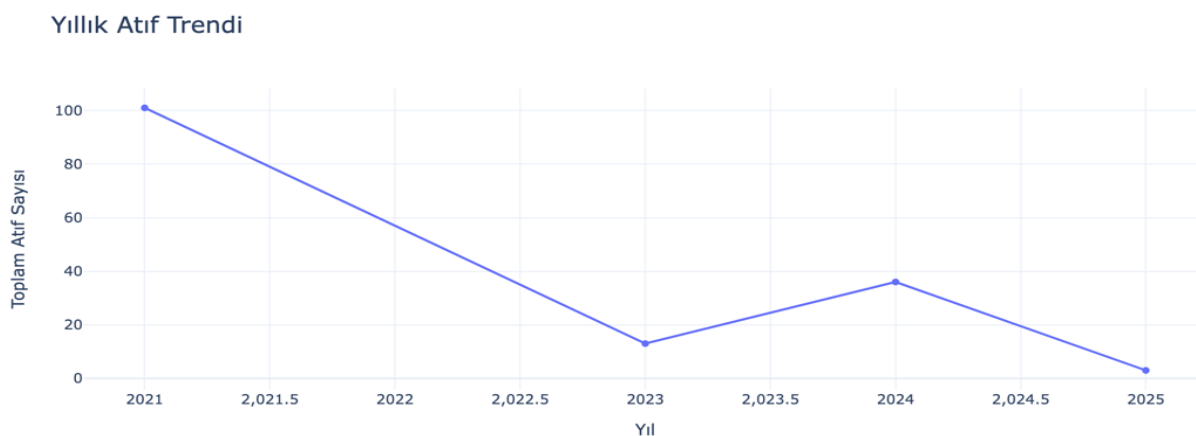

Figure 2. Annual Citation Trend

Figure 3. Annual Citation Trend (6-Month Intervals)

Figure 2 shows annual citation trends in L2 assessment through AI while Figure 3 demonstrates it with 6-month intervals during the whole period, in which the changes in the annual citation trend were almost equal. The highest number of citations was revealed in 2021 (n=101), demonstrating that the publications had a high impact due to the new research field. On the other hand, a dramatic decrease in the citations (n=0)

occurred in 2022, followed by a slight increase in 2023 (n=13). The upward trend was maintained in 2024, with the second highest citation count during this period (n=36). However, it demonstrated a sharp decrease in 2025, with a dramatically low number (n=3). The fluctuating trend indicates that 2021 was the time that AI-based language assessment, which was a new field for scholars, had scholarly interest even though the fluctuation in the rest of the period may depend on research relevance of visibility in academic networks.

4.2. Predictions for Future Trends

Based on the results of the 5-year trend in publications and citation counts, the prediction for the future year publications is demonstrated in Figure 4 and the predictions for the future year citation is shown in Figure 4.
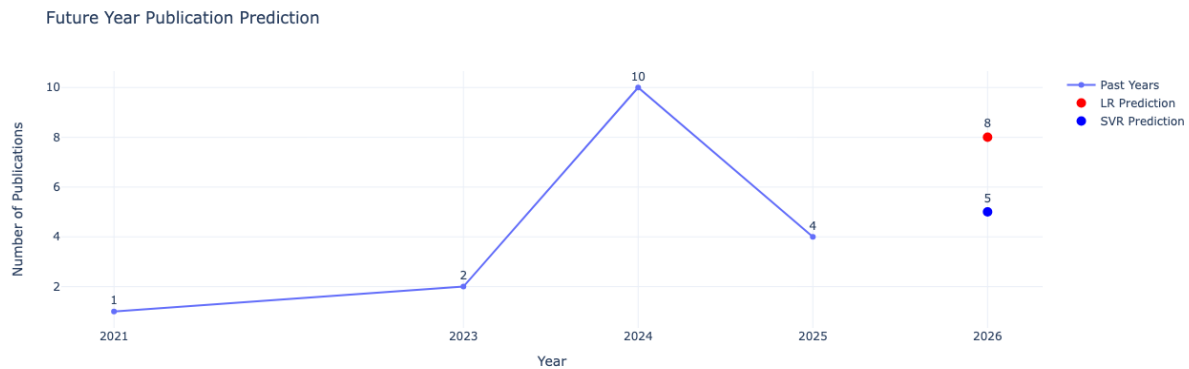

Figure 4. Future-year Publication Prediction

As it was given in Figure 1, the changes in the number of annual publications have also been demonstrated in Figure 4 as well as the predicted number of the publications in 2026. The results of linear regression (LR) model forecasted 8 publications in 2026; however, employing a more cautious approach, the Support Vector Regression (SVR) model predicted 5 publications for the next year. The reason for the difference in both predictions could be the fluctuations in the 2021-2025 period. Moreover, it could be inferred that the SVR model potentially provides a realistic estimate as the fluctuation in the results could be modelled in a better way.
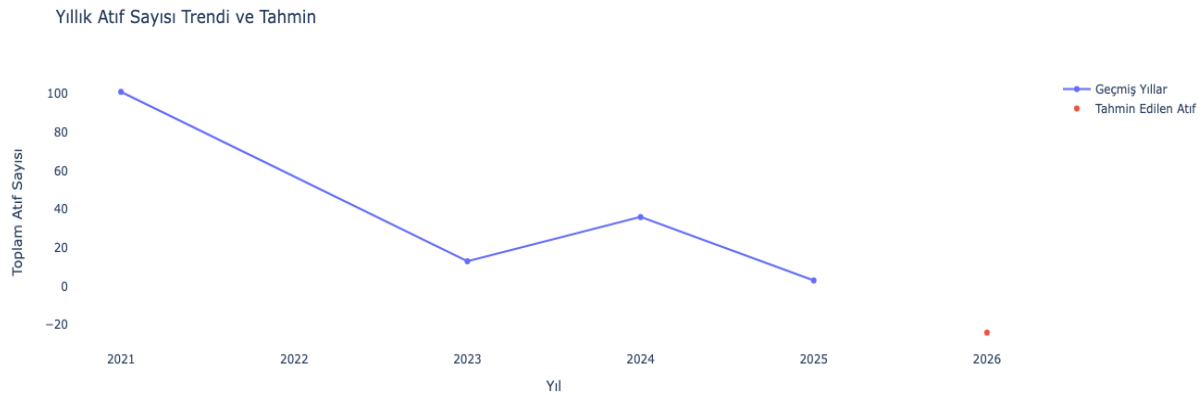
Figure   5.   Future   Year   Citation   Prediction

As it was given in Figure 2, the changes in the number of citation counts have also been demonstrated in Figure 5 in addition to the predicted number of the citations in 2026. The citation count was estimated to be -24.1 in 2026; however, not being a valid prediction, it reveals the model adopted may not be appropriate for the data. Overall, despite the periodic increase in the number of publications, the fluctuation and expected decrease in the citations suggests the lack of contribution to the academic impact. It indicates reevaluation of both quality of the publication and impact on the field.

4.3. Citation and Publication Relationships

The relationship between the publication year and the citation count was investigated and visualized below. Figure 6 demonstrates the trend with a regression line in scatter plot whereas Figure 7 shows the relationship with a bubble chart.
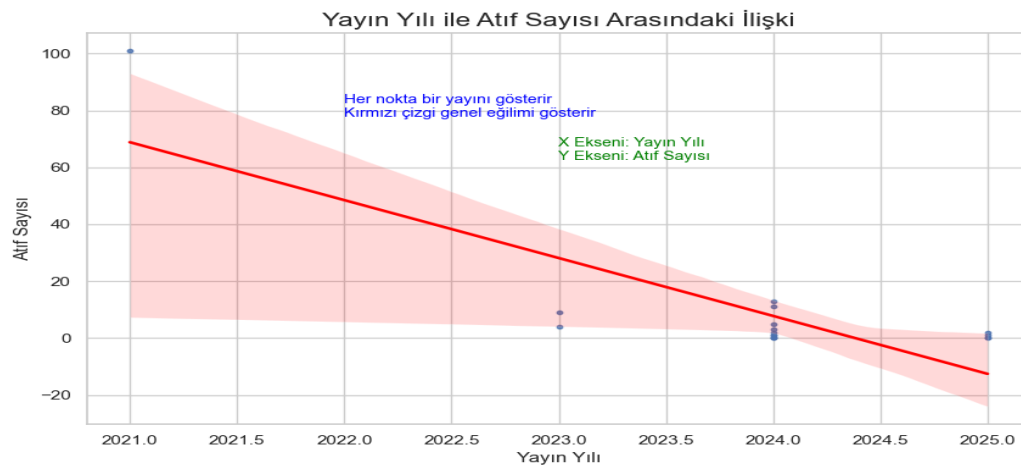
Figure 6. Correlation Between Publication Year and Citation Count

In the scatter plot, the representation of each publication is shown in each point, and the regression line demonstrates the overall trend (Figure 6). The trend line indicates a negative correlation between the year of publication and the citation count as there is decrease in the citation count over the years. It could also be said that the publications in the older years had more citation counts while the recent ones had less citation counts. It could be based on both the effect of duration, which is called citation lag, and the changes in the quality and visibility of publications.
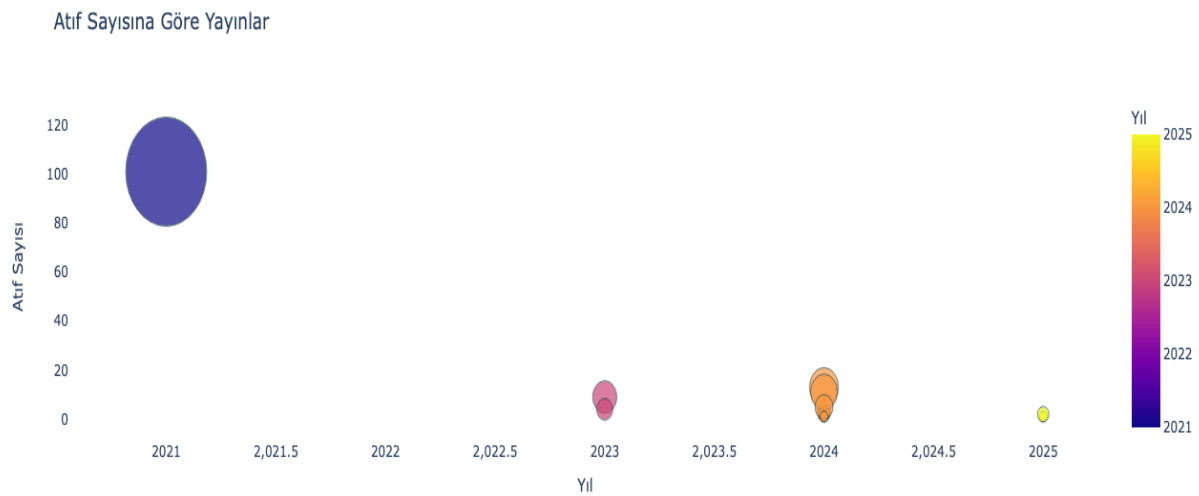

Figure 7. Publications Based on Citation Count

Figure 7 shows the citation count for each publication with the bubbles' size. The size of the bubbles represents the citation count, and its colors show the publication year. A publication in 2021 was cited the most (n=101); however, the publications in the other years were represented with smaller bubbles indicating lower citation counts. The difference in the citation counts could depend on the limited time for academic visibility.

Based on these two graphs, it could be clearly seen that there is a decrease in the citation impact over the years. A publication in 2021 comes forward with its citation count; however, the publications in the other years had less impact. It may reveal the requirements for the strategies to increase the impact and visibility of the publications.

4.4. Geographic and Institutional Insights

The countries where the scholars published their research were also investigated. Figure 8 demonstrates the density of publication on the world map and the countries

were colored based on the number of publications. The yellow-colored country, the USA, was the one that most publications were conducted; however, the dark colored places on the map, such as China, Hong Kong and Macao, were the ones where less studies were published.

Most Published Countries
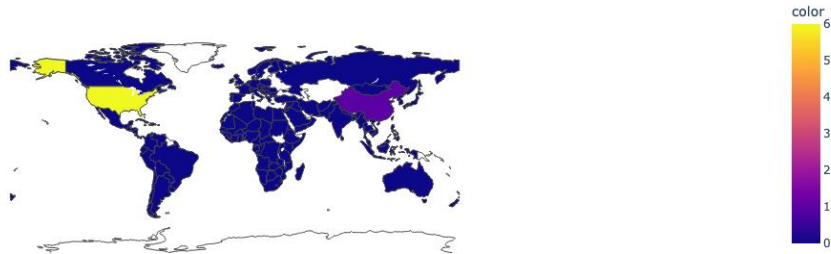
Figure   8.   Map   of   Most   Published   Countries

Figure 9 illustrates the number of publications based on the countries. The USA contributes to the literature with the highest number of publications (n=6). Each of the other countries, which are Hong Kong, China and Macao had only one publication. The results indicate that most contribution to the literature was made by the USA, but the Asian countries still had limited contribution. It could be suggested to increase international cooperation in publications for diversity.

Figure 9. Most Published Countries

4.5. Keyword and Journal Analysis

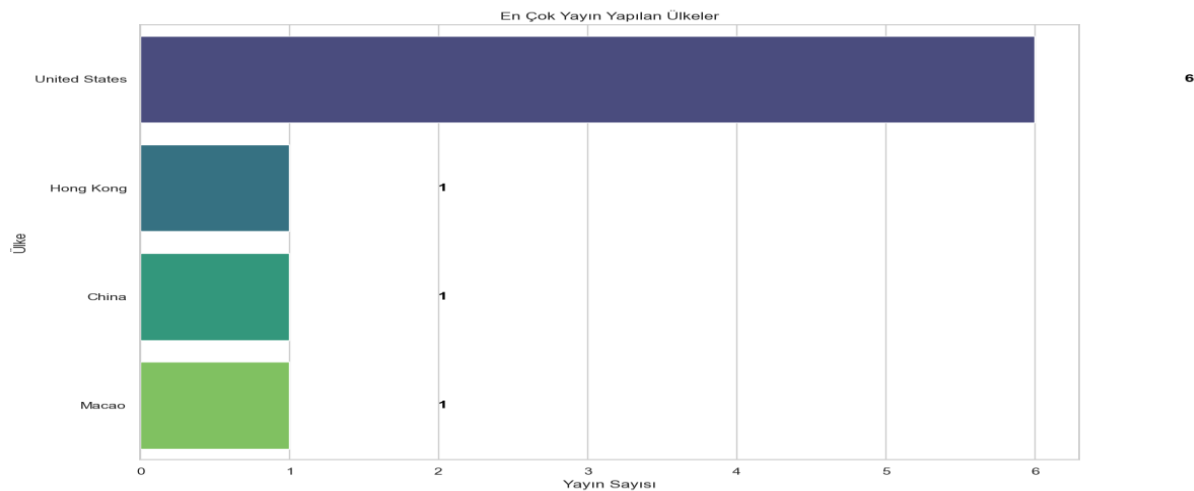Most frequent keywords, their correlation with the citation count, and the top journals were analyzed in order to reveal the thematic focus. Figure 10 shows the keywords that were most frequently used in the publications. The keyword "artificial intelligence" is the most utilized one (n=7), followed by "ChatGPT" (n=5), "second language writing" (n=4) and "writing assessment" (n=4). The terms "AI" and "L2 writing" were also recurring terms to be used (n=3). The other terms, "generative artificial intelligence", "automated essay scoring", "generative AI", and "language learning" were equally utilized as keywords (n=2).

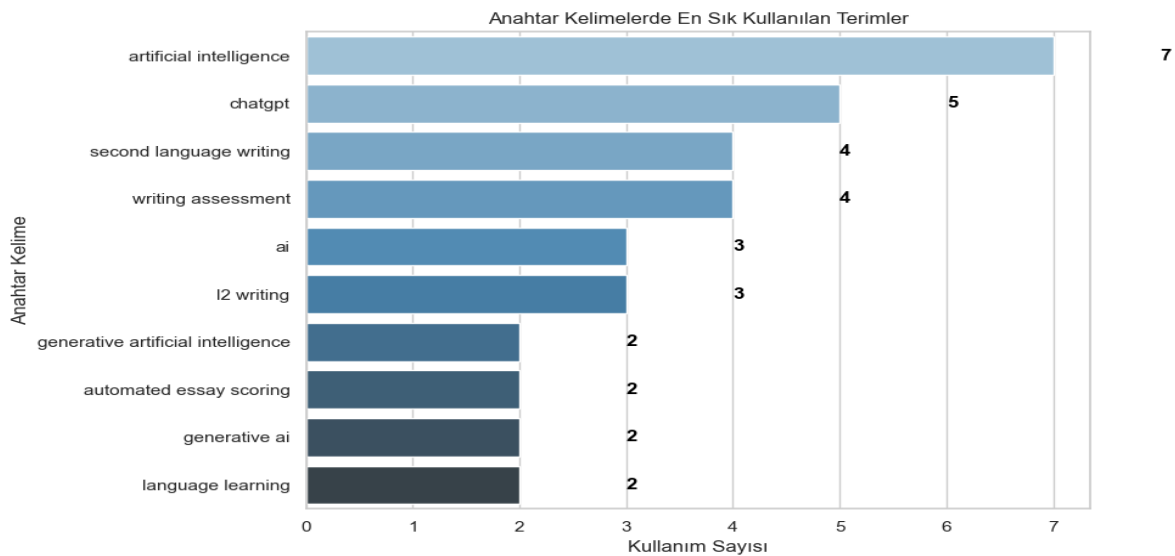Figure   10.   Most   Used   Keywords

Moreover, the word cloud below is the visual demonstration of the keywords (Figure 11). Words such as writing, artificial intelligence, language, assessment, learning, and ChatGPT emerged in big fonts referring to highly frequent and relevant ones with the publications. Also, the other terms like automated, feedback, and generative support the main theme.

Figure 11. Word Cloud of the Keywords

As for the relationship between the number of keywords in the titles and the citation counts, Figure 12 demonstrates a slight positive correlation between them. Despite not having a strong correlation, the rising regression line may illustrate the reason for more citations when the keywords were used in the title as the publications would be more visible and searchable through specific keywords in the databases.

Figure 12. Correlation Between Keyword Count and Citation Count

The journals that published the studies were also investigated (Figure 13). Assessing Writing and Language Learning and Technology were the journals that had the highest number of publications (n=2). All the other journals, which were *Language Learning and Technology, Journal of Technology & Chinese Language Teaching, Language Learning and Teaching, Higher Education Advances, Technology, Knowledge and Learning, British Journal of Educational Technology, Heliyon,* and *RELC Journal,* published one article. It could be seen from the distribution that the publications were hosted in journals on writing assessment, educational technology and language acquisition.

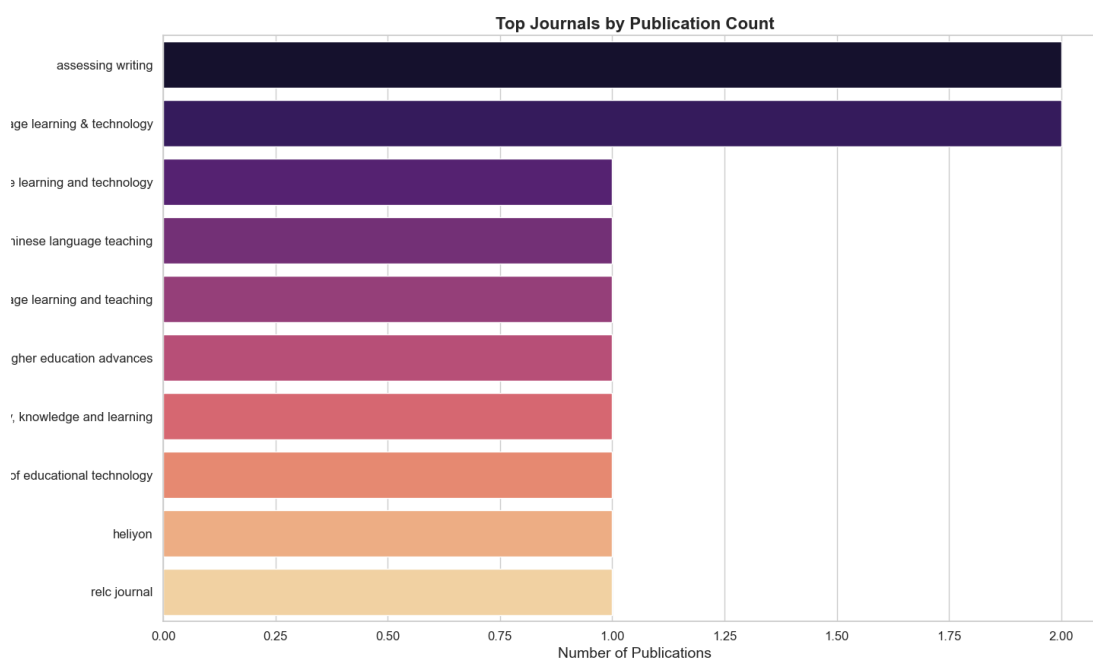**Top Journals by Publication Count**

Figure 13. Top Journals by Publication Count

The analysis of keywords and journals shows that publications mostly focused on the effect of AI on L2 writing and writing assessment. In spite of the diversity in keywords, AI and educational terms were highly adopted in the publications. Additionally, there was a slight increase in the effect of using more keywords in the titles on the citation counts. Consequently, the fact that the publications were made in the journals whose specializations were certain indicates that the publications were situated in language education and educational technology.

4.6. Author and Article Metrics

Author and article metrics were also investigated. Figure 14 shows the top 10 authors who contributed to the literature based on the publication count.
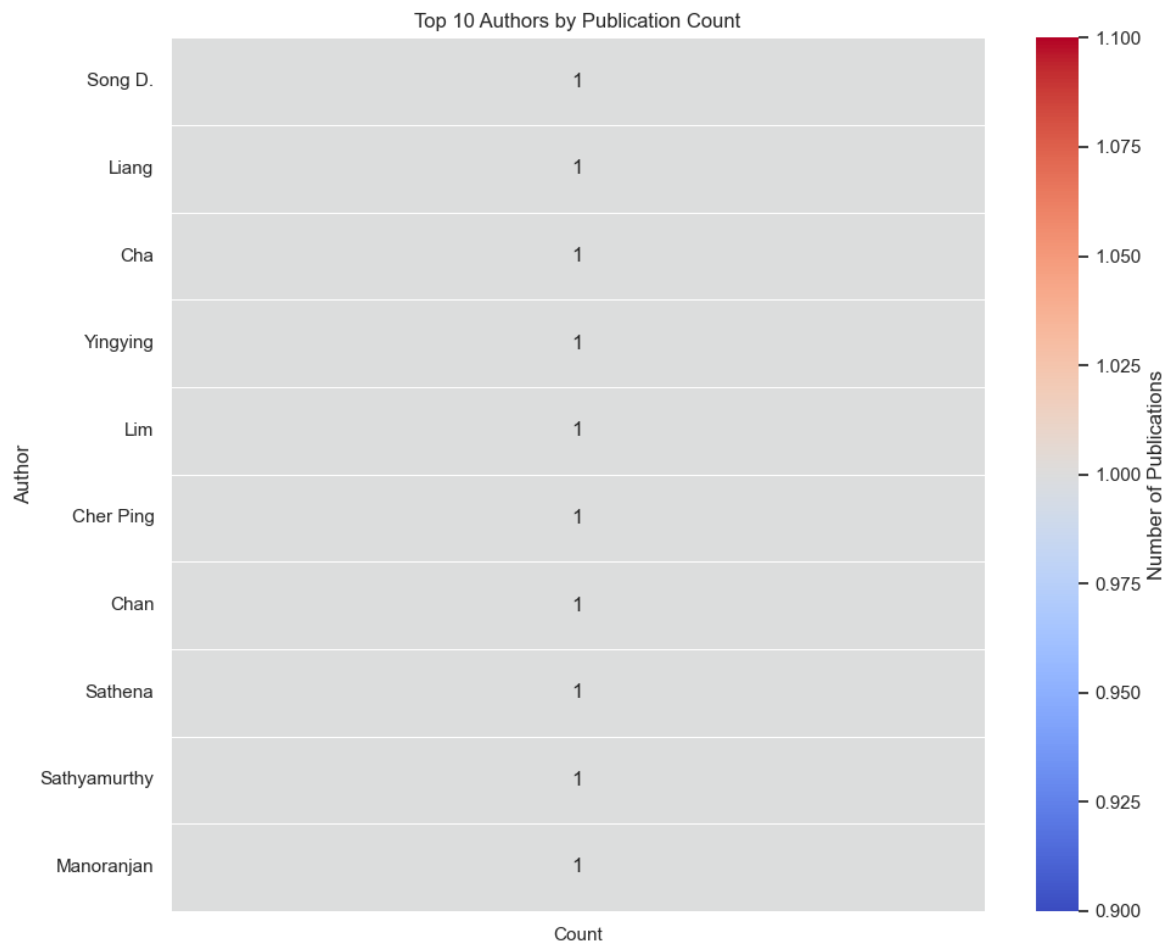

Figure 14. Top 10 Authors by Publication Count

It could be seen that the authors in the chart made an equal contribution with their publication (n=1). It shows that there is no dominance of any authors, which leads to diverse authorships in the field. It could also be understood that the AI-based L2 assessment has attracted a number of scholars' attention instead of only a few prolific ones, implying that the field has been improving.

Table 1 presents the most cited publication in the field of AI-based L2 assessment.

Table 1. Most cited publication in AI-based L2 assessment

Title	Author(s)	Citation Count
Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial	Nazari, N., Shabbir, M. S., & Setiawan, R	101
Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers	Wang C.	13
Editorial Introduction – AI, corpora, and future directions for writing assessment	Hartwell, K., & Aull, L	9
Distributed agency in second language learning and teaching through generative AI	Godwin-Jones, R.	5
Integrating Automated Written Corrective Feedback into E-Portfolios for second language Writing: Notion and Notion AI	Osawa, K.	4
Can ChatGPT Reliably and Accurately Apply a Rubric to L2 Writing Assessments? The Devil is in the Prompt(s)	Poole F.J. & Coss, M.D.	3
Leveraging ChatGPT for Second Language Writing Feedback and Assessment	Li M.	2
A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness	Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P.	2
Integrating Metadiscourse Analysis with Transformer-Based Models for Enhancing Construct Representation and Discourse Competence Assessment in L2 Writing: A Systemic Multidisciplinary Approach	Chan, S., Sathyamurthy, M., Inoue, C., Bax, M., Jones, J., & Oyekan, J.	1

Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial, written by Nazari et al. (2021) and published in *Heliyon,* was the publication that was cited most in the data set with 101 citation count. In this study, Nazari et al. (2021) investigated the effect of AI-powered writing tools on ESL postgraduate students' academic writing skill. With the participation of 120 students, the study adopted an experimental research design. The results of the study indicated that AI tools both improve writing skills and positively affect students' emotional and cognitive engagement with academic writing.

The second study, which had the most citation count (n=13), was *Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers,* written by Wang (2024*)* and published in *Technology, Knowledge and Learning.* It was aimed to explore the way students, who were both native and non-native English speakers, engage with ChatGPT in academic writing. The

results of the qualitative study show that ChatGPT was used in order to brainstorm, outline ideas, revise, and edit, as well as its benefits, such as a fast writing process, lower cognitive load, immediate feedback, and new learning opportunities. Despite the benefits, some concerns occurred, e.g., inaccurate and superficial information, plagiarism, and academic dishonesty.

All the studies have common qualities despite different contexts. As is known, they were on the assessment of writing and academic writing skills based on AI tools. Most of them focus on EFL context and assess the students' experiences, performances, and perceptions during the AI-based writing process (Nazari et al., 2021; Li, 2024; Osawa, 2024; Poole & Coss, 2024; Wang, 2024). Even though there were methodological differences, which creates a richness in research methods and data, the common ground was to investigate the educational impact of AI tools on L2 writing skill and assessment.

Besides the top cited publications, the current study also revealed 10 recently published articles on AI-based writing assessment with the citation count (Table 2.)

Table 2. Recently Published Articles

Title	Author(s)	Publication Date	Citation Count
Exploring AI-assistance in L2 Chinese writing with standardized assessment tasks	Song, D., & Tang, A. F.	2025-01-01	1
A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness	Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P.	2025-01-01	2
Advancing Language Assessment with GPT: Is It Nonnative-Language Friendly?	Won, D. O., Shin, Y. K., Kim, H. J., & Yoo, I. W.	2025-01-01	0
Why we need more research on technology applications in less-commonly-taught-language (LCTL) programs	Winke, P., & Koné, K.	2025-01-01	0
Can ChatGPT Reliably and Accurately Apply a Rubric to L2 Writing Assessments? The Devil is in the Prompt(s)	Poole F.J. & Coss, M.D.	2024-06-01	3
Leveraging ChatGPT for Second Language Writing Feedback and Assessment	Li M.	2024-01-01	2
Bridling, Taming and Riding the AI Beast	Despain J.A.	2024-01-01	0
Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers	Wang C.	2024-01-01	13
Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays	Guo, K., Zhang, E. D., Li, D., & Yu, S.	2024-01-01	1

It shows that publications on AI-based writing have increased, especially in recent years. Most of the publications in the data set focus on L2 writing and how large language models like ChatGPT could be used to give feedback. For instance, Wang (2024) emphasized the perceptions of students towards AI usage during the writing process with

a relatively high citation count (n=13). Similarly, Poole and Coss (2024) and Li (2024) had a publication on the reliability of ChatGPT for L2 writing assessment and quality of feedback. Notably, the fact that recent publications such as Song (2025) and Chen et al. (2025) have just been cited indicates the scholarly interest and contemporary field. Consequently, publications that adopted both qualitative and quantitative research designs reveal that AI tools like ChatGPT provide potential opportunities to improve L2 writing skills and create some ethical, technical, and pedagogical issues.

5.   Discussion and Conclusion

This study examined the current landscape of AI-based writing assessment in second language (L2) education by conducting a bibliometric and predictive analysis of scholarly publications from 2021 to 2025. The results reveal emerging research trends, core thematic focuses, and the distribution of scholarly contributions across journals, authors, and countries. Together, these findings provide insights into the developmental trajectory of this rapidly evolving field.

The keyword and journal analyses indicate that AI-related tools—particularly ChatGPT—are becoming central to research on L2 writing assessment. Terms such as artificial intelligence, writing assessment, feedback, and automated essay scoring appeared frequently, reflecting a growing emphasis on the integration of genAI tools in writing instruction and evaluation (Li, 2024; Wang, 2024; Cheng et al., 2023; Chen et al., 2024; Guo et al., 2024; Nazari et al., 2021). These findings are further supported by word cloud visualizations that underscore the dominance of these themes. The prevalence of such keywords demonstrates researchers' increasing attention to the pedagogical and assessment capabilities of large language models in writing contexts.

A mild positive correlation was observed between the number of keywords used in article titles and their citation counts, suggesting that strategic keyword selection may contribute to improved academic visibility. While this supports Kızılöz's (2020) argument that keyword diversity can influence citation performance, the weak strength of this correlation also implies the involvement of additional factors, including research methodology, novelty of findings, and journal prestige. Thus, while titles play a role in discoverability, they are only one of several variables that shape a publication's academic impact.

The analysis of publication venues revealed a concentration of studies in journals with a focus on writing assessment and educational technology, such as Assessing Writing and Language Learning and Technology. This suggests that research in this area sits at the intersection of language education and AI-enhanced learning environments (Yang & Wang, 2025). At the same time, the diversity of publication outlets indicates that the

topic appeals to a wide array of academic disciplines, confirming its interdisciplinary potential.

Authorship and geographical data point to a field that is expanding, but not yet consolidated. The fact that each author contributed only one publication indicates a lack of dominant figures and reflects an inclusive scholarly environment. However, this also signals the absence of sustained collaborative networks or core research groups. Most studies originated from the United States, aligning with findings by Won et al. (2025), who noted the prevalence of English-language and US-based studies in AI-assisted writing assessment. This geographic imbalance may be attributed to disparities in digital infrastructure, language dominance, and academic funding.

Overall, the findings suggest that AI-based writing assessment in L2 education remains in an early, exploratory stage of development. Although central research themes have begun to emerge, the absence of established academic clusters or sustained international collaborations highlights the need for more coherent and longitudinal research efforts. Future studies should focus on fostering interdisciplinary partnerships, developing datasets that capture diverse linguistic and cultural contexts, and designing methodologies that can evaluate long-term learning outcomes and discourse-level writing skills. As the field matures, ensuring equity in AI development and assessment— particularly for less commonly taught languages and underrepresented regions—will be essential for inclusive progress.

The current study reveals the thematic focuses of AI-based writing assessment, publication trends, and leading contributors. First, keyword analysis demonstrated that the focus of research was on the themes of *"artificial intelligence," "ChatGPT," "writing assessment"* and *"second language writing"*. Word cloud visualization also supports the findings of the bar chart with the leading themes, i.e., *"Artificial Intelligence", "writing", "assessment", "ChatGPT"* and *"feedback"* which indicates that there has been a tendency to integrate genAI tools into the field of language teaching and assessment (Li, 2024; Wang, 2024). Moreover, keywords like *"automated essay scoring," "feedback,"* and *"L2 writing"* show the increasing significance of adopting technological tools in writing assessment (Cheng et al., 2023; Chen et al., 2024; Guo et al., 2024; Nazari et al., 2021).

Second, a positive correlation between keyword and citation counts was observed, indicating that the usage of more keywords leads to an increase in academic visibility. It was supported by Kızılöz (2020) that the diversity in the keywords in the titles could increase the citation counts. However, not having a strong correlation shows that getting cited depends not only on the diversity of keywords in the titles of publications but also on other factors such as the content of the study, the method, and the journals.

Next, based on the journal analysis, most studies were published in the journals whose scope was writing assessment and educational technology, such as *"Assessing Writing"* and *"Language Learning and Technology"*. This result demonstrates that the focus of

research is on language assessment, foreign language education, and technology-assisted learning (Yang & Wang, 2025). Additionally, the diversity in the journals in which the studies were published indicates the interdisciplinary structure of the issue. Moreover, the authorship analysis reveals that a wide range of scholars made contributions to the literature and there is not any dominance of specific authors. These contributions mostly arose from the USA, which could be supported by the study of Won et al. (2025) indicating that English is the language that most studies on AI-based writing assessment are conducted. Also, it could be understood that the studies are conducted in the country where technological improvement and infrastructure in academic publishing are improved. In addition, the fact that every author had one publication in the dataset indicates that the field is still improving and scholarly contribution emerges from various centers. This diversity and the imbalance in the publication places show that scientific cooperation and academic networks have not been established; however, they could be developed with the increasing improvement in the field.

Consequently, AI-based writing assessment is revealed to be a recent and improving field of study in the exploration phase. The keywords are determined even if a centralized scientific community has not yet occurred. It is suggested that interdisciplinary research could be conducted with the longitudinal research methodology and more systematic datasets.

Declaration of Conflicting Interests and Ethics

"The authors declare no conflict of interest."

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research,* 3, 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Bhatia, V. (2014). *Worlds of written discourse: A genre-based view*. London: Continuum.

Callon, M., Law, J., & Rip, A. (1983). Co-word analysis: A new method for mapping the dynamics of science. *Scientometrics, 22*(1), 155–205. https://doi.org/10.1007/BF02002100

Chan, S., Sathyamurthy, M., Inoue, C., Bax, M., Jones, J., & Oyekan, J. (2024). Integrating Metadiscourse Analysis with Transformer-Based Models for Enhancing Construct Representation and Discourse Competence Assessment in L2 Writing: A Systemic Multidisciplinary Approach. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special Issue), 318-347. https://doi.org/10.21031/epod.1531269

Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P. (2024). A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness. *Journal of Computer Assisted Learning. 41*(1). DOI: 10.1111/jcal.13064

Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society, 26*(1), 112–131.

Despain, J. (2024, June). Bridling, Taming and Riding the AI Beast. In *10th International Conference on Higher Education Advances (HEAd'24)*. Editorial Universitat Politècnica de València.

Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–31. https://hdl.handle.net/10125/73570

Guo, X., Zhang, E.D., Li, D. & Yu, S. (2024). Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays. *British Journal of Educational Technology*. DOI: 10.1111/bjet.13540

Hartwell, K., & Aull, L. (2023). Editorial Introduction – AI, corpora, and future directions for writing assessment. *Assessing Writing, 57. https://doi.org/10.1016/j.asw.2023.100769*

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Bloomsbury Publishing.

Kesgin, K., & Zeren Özer, D. (2023). Redefining Bibliometric Intelligence: BiBLoX for Real-Time Science Trend Mapping and Machine Learning-Driven Forecasting. Authorea Preprints. https://doi.org/10.22541/au.174106668.81840735

Kızılöz, H. E. (2020). Bilimsel Makalelerin Atıf Sayısı Tahmini. Avrupa Bilim ve Teknoloji Dergisi, (Özel Sayı), 370-375.

Li, M. (2024). Leveraging ChatGPT for second language writing feedback and assessment. *International Journal of Computer-Assisted Language Learning and Teaching, 14*(1), 1-11. 10.4018/IJCALLT.360382

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open, 3*, 111–132. https://doi.org/10.48550/arXiv.2106.04554

Mazloumian, A. (2012). Predicting scholars' scientific impact. *Science, 338*(6110), 936–939. https://doi.org/10.1126/science.1227823

Nazari, N., Shabbir, M.S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial. *Heliyon, 7* (5). https://doi.org/10.1016/j.heliyon.2021.e07014

Newman, M. E. J. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences (PNAS), 98(2), 404–409. https://doi.org/10.1073/pnas.98.2.404

Osawa, K. (2024). Integrating automated written corrective feedback into e-portfolios for second language writing: using Notion AI. *RELC Journal, 55(3) 881–887. DOI: 10.1177/00336882231198913*

Poole, F. J., & Coss, M. D. (2024). Can ChatGPT Reliably and Accurately Apply a Rubric to L2 Writing Assessments? The Devil is in the Prompt (s). *Journal of Technology & Chinese Language Teaching*, *15*(1), 1-24.

Song, D., & Tang, A. F. (2025). Exploring AI-assistance in L2 Chinese writing with standardized assessment tasks. *Language Learning and Technology, 29*(2), 161-189. 10943501.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Tahamtan, I., & Bornmann, L. (2022). A systematic review of citation prediction studies. *Journal of Informetrics, 16*(4), 101317. https://doi.org/10.1016/j.joi.2022.101317

Wang, C. (2024). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning, 70*, https://doi.org/10.1007/s10758-024-09744-3

Williams, C., & Beam, S. (2019). Technology and writing: Review of research. *Computers & Education, 128*, 227–242. https://doi.org/10.1016/j.compedu.2018.09.024

Winke, P., & Koné, A. (2025). Why we need more research on technology applications in less-commonly- taught-language (LCTL) programs. *Language Learning and Technology. 29*(2), 1-10.

Won, D.O., Shin, Y. K., Kim, H.J., & Yoo, I. W. H. (2025). Advancing language assessment with GPT: Is it nonnative-language friendly? *Language Assessment Quarterly, 22*(1), 1–28. https://doi.org/10.1080/15434303.2024.2444349

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9*(1), 59. https://doi.org/10.1186/s40594-022-00377-5

Yan, E., & Ding, Y. (2011, October). Citation count prediction: Learning to estimate future citations for literature. Proceedings of the American Society for Information Science and Technology (ASIST), 48(1), 1–10. https://doi.org/10.1002/meet.2011.14504801134

Yang, Z., & Wang, P. (2025). Current status and research trend of English language assessment: a bibliometric analysis. *Language Testing in Asia*, *15*(1), 11. https://doi.org/10.1186/s40468-024-00317-w